

# Exact Distribution of the $F$ -statistic under Heteroskedasticity of Unknown Form for Improved Inference

Jianghao Chu\*   Tae-Hwy Lee<sup>†</sup>   Aman Ullah<sup>‡</sup>   Haifeng Xu<sup>§¶</sup>

October 26, 2020

## Abstract

The exact finite sample distribution of the  $F$  statistic using the heteroskedasticity-consistent (HC) covariance matrix estimators of the regression parameter estimators is unknown. In this paper, we derive the exact finite sample distribution of the  $F$  ( $= t^2$ ) statistic for a single linear restriction on the regression parameters. We show that the  $F$  statistic can be expressed as a ratio of quadratic forms, and therefore its exact cumulative distribution under the null hypothesis can be derived from the result of Imhof (1961). A numerical calculation is carried out for the exact distribution of the  $F$  statistic using various HC covariance matrix estimators, and the rejection probability under the null hypothesis (size) based on the exact distribution is examined. The results show the exact finite sample distribution is remarkably reliable, while, in comparison, the use of the  $F$ -table leads to a serious over-rejection when the sample is not large or leveraged/unbalanced. An empirical application highlights that the use of the exact distribution of the  $F$  statistic will increase the accuracy of inference in empirical research.

*Key Words:* Heteroskedasticity, Finite sample theory, Imhof distribution

*JEL Classifications:* C1, C2

---

\*Department of Economics, University of California, Riverside. Email: jchu018@ucr.edu.

<sup>†</sup>Department of Economics, University of California, Riverside. Email: tae.lee@ucr.edu.

<sup>‡</sup>Department of Economics, University of California, Riverside. Email: aman.ullah@ucr.edu.

<sup>§</sup>Corresponding author. Department of Statistics, School of Economics and Wang Yanan Institute for Studies in Economics (WISE), Xiamen University. Email: xhf1984@hotmail.co.jp.

<sup>¶</sup>The authors are grateful for the useful comments from the editor and three anonymous referees.

# 1 Introduction

The presence of heteroskedasticity of unknown form in the disturbances of linear regression models has been a topic of major concern for inference in economic models. Under heteroskedasticity, the ordinary least squares (OLS) estimators of the regression parameters are well known to be consistent. However, the usual estimators for the covariance matrix of the regression parameters are inconsistent and/or biased. Numerous work has been done to formulate an unbiased and consistent estimate for the covariance matrix of the regression parameters. Earlier contributions include [Eicker \(1963\)](#), [Huber \(1967\)](#) and [Hinkley \(1977\)](#).<sup>1</sup> [White \(1980\)](#), in a seminal paper, introduces the well-known “White Standard Errors” which is able to estimate the covariance matrix of the regression parameters consistently under an unknown form of heteroskedasticity of the disturbances.

Not long after [White \(1980\)](#) established the asymptotic consistency of the heteroskedasticity-consistent (HC) covariance estimator, it was quickly recognized that inferences based on the “White Standard Errors” over-reject the null hypothesis when the sample is not large. [MacKinnon and White \(1985\)](#) provide simulation evidence and, linking the problem to the reduction in error variance brought by least squares fitting, further propose variants of the “White Standard Errors” to achieve better finite sample properties through adjusting for degrees of freedom and leverages and by using the jackknife. These corrections, however, are far from satisfactory, as shown in simulations by [Angrist and Pishke \(2009\)](#), who note that the covariance estimates are not merely biased but also have higher volatilities than default OLS estimates, which contributes to their high rejection rates. Extensions on inferences using variants of the HC covariance estimators include [McCaffrey and Bell \(2002\)](#), [Cribari-Neto \(2004\)](#), [MacKinnon \(2013\)](#), [Hausman and Palmer \(2012\)](#) and [Imbens and Kolesár \(2016\)](#).

Recent studies have pointed out that the over-rejection problem remains problematic in considerably large samples when the data are leveraged/unbalanced. Based on simulation studies [Young \(2016\)](#) examines the distortion of the  $t$ -statistic due to interactions of the hypothesis and the data. [Chesher and Austin \(1991\)](#) use the Imhof method to calculate exact finite-sample distributions of  $t$ -statistics distinguished by the variance estimators, investigate how these  $t$ -distributions differ with the first-order asymptotic normal approximation and the second-order Edgeworth approximations, show that their tests based on the Jackknife standard error estimator generally outperform in terms of the test size being close to the

---

<sup>1</sup>A related literature to the heteroskedasticity-consistent (HC) covariance matrix estimation of the regression coefficients is on the estimation of the heteroskedastic error variances themselves. [Rao \(1970\)](#) proposes the MINimum Norm Quadratic Unbiased Estimation (MINQUE) of the heteroskedastic error variances. [Horn et al. \(1975\)](#) proposes an “almost unbiased” estimator as an estimator of heteroskedastic error variance, which [MacKinnon and White \(1985\)](#) uses to construct HC2 as appears below in Equation (7). The variance estimator of [Horn et al. \(1975\)](#) is almost unbiased as its asymptotic bias is of order  $n^{-1}$  when errors are heteroskedastic, and it is unbiased when errors are homoskedastic.

nominal size, and compare the behavior of tests under homoskedasticity and under changes in the regression designs. Hansen (2017) derives the exact distribution of a  $t$ -type statistic, and through simulations show that using the “White Standard Errors” has significant size distortion when the data is highly leveraged. He also proposes corrections for the  $t$ -type test statistic under i.i.d. normal disturbances. However, the proposed methods perform poorly when the disturbances are heteroskedastic. The basic idea behind these papers are that the nominal sample size may be “misleading” as the regression results may depend heavily on only a part of the observations rather than equally on all the data, which will lead to a higher volatility of the results and the over-rejection of the test statistic.

In view of these issues, in contrast, this paper derives the exact finite sample distribution of an  $F$ -type ( $t^2$ -type) test statistic for testing regression parameters under the heteroskedasticity of unknown form. This test statistic is valid for a single linear restriction on the regression parameter including the test for the zero restriction on each coefficient. It is shown that this proposed  $F$ -type test statistic can be expressed as a ratio of quadratic forms, and therefore its exact cumulative distribution under the null hypothesis can be easily written, and straightforwardly implemented from the result of Imhof (1961) on the distribution of a quadratic form.<sup>2</sup> Further, we carry out numerical calculation of the proposed test statistic using Imhof (1961) to present the cumulative distribution functions and probability of rejections under various HC covariance estimators of regression estimators. The accuracies are confirmed from their corresponding simulation-based results. It is shown that the exact distribution of the  $F$ -type test statistic based on various HC covariance estimators provides accurate testing results.<sup>3</sup>

We conduct Monte Carlo simulations to examine the behavior of the test’s size of the  $F$  statistic under the exact Imhof distribution and compare them with values in the  $F$ -table. It is found that the empirical size behavior of the  $F$  statistic using the exact Imhof distribution is generally very good except when the sample size is very small (e.g.,  $n = 15$ ). Using the exact Imhof distribution is always much better than using the  $F$ -table. When the Imhof distribution is used, the HC estimator of MacKinnon and White (1985) exhibits only slightly better size behavior in comparison with other HC estimators in White (1980), Hinkley (1977), and Cribari-Neto (2004) as all of these HC estimators are somewhat similar when the exact Imhof distribution is employed. It indicates that the selection between HC estimators appears irrelevant for empirical researches if the exact Imhof distribution is used. To the contrary, when the  $F$ -table is used, the HC estimator of MacKinnon and White (1985) is clearly much better than the HC estimators of White (1980), Hinkley (1977), and

---

<sup>2</sup>For applications of Imhof (1961) in a separate context of studying distributional properties of estimators, see Bao et al. (2017), Ullah (2004), Nakamura and Nakamura (1998), and Farebrother (1985), among others.

<sup>3</sup>The accurate test size of using the exact Imhof distribution would also produce more accurate confidence intervals which can be computed from inverting the test statistics using the Imhof distribution.

slightly better than the HC estimator of [Cribari-Neto \(2004\)](#). The robustness of our results, under a non-normal distribution of errors and when the estimated parameters are used in the  $F$ -distribution, are also analyzed.

In an empirical application of the exact Imhof distribution, it is shown that the  $p$ -values of the  $F$  statistic from the exact distribution are similar to the wild bootstrap  $p$ -values, while the  $p$ -values from the  $F$ -table are different from the wild bootstrap  $p$ -values. This empirical application highlights that the exact  $p$ -values obtained from the exact Imhof distribution of the  $F$  statistic does increase the accuracy of inference in empirical research.

The remainder of this paper is organized as follows. In Section 2, the model and the  $F$  test statistic are discussed, various HC covariance matrix estimators are presented, and the need of the exact distribution for the statistic is discussed to set the goal of this paper. In Section 3 we derive the exact distribution for the statistic, and compare it for the different HC estimators by a numerical calculation. In Section 4, Monte Carlo simulation is conducted to examine the size behavior of the  $F$  statistic. In Section 5, we present an empirical application. Section 6 concludes.

## 2 Test Statistic and HC Covariance Matrix Estimators

In this paper we consider the linear model with independent, heteroskedastic and normally distributed errors

$$y_i = \beta_0 + \sum_{k=1}^p \beta_k x_{ik} + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma_i^2), \quad i = 1, \dots, n. \quad (1)$$

Let the error vector be denoted as  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)'$  with its variance-covariance matrix  $\text{Var}(\varepsilon) \equiv \mathbf{\Omega} = \text{diag}\{\sigma_i^2\}$ . Let  $\mathbf{X}$  be the nonstochastic matrix with its  $i$ th row being  $\mathbf{x}'_i = (1, x_{i1}, \dots, x_{ip})$ . The regression parameter vector  $\beta = (\beta_0, \beta_1, \dots, \beta_p)'$  is estimated by the ordinary least squares (OLS) estimator

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \quad (2)$$

which has the variance-covariance matrix

$$\text{Var}(\hat{\beta}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \equiv \mathbf{\Sigma}. \quad (3)$$

Let  $e_i$  be the  $i$ th element of the  $n \times 1$  residual vector  $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\beta}$ , and denote  $h_{ij} \equiv \mathbf{x}'_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_j$ .

We are interested in testing for the null hypothesis with a *single* linear restriction on the

parameter vector,  $H_0 : \mathbf{r}'\beta = r_0$ , where  $\mathbf{r}'$  is a  $1 \times (p + 1)$  non-random vector and  $r_0$  is a scalar constant. We consider the  $F$  statistic of the form, under  $H_0$ ,

$$F = t^2 = \left( \frac{\mathbf{r}'\hat{\beta} - r_0}{\sqrt{\mathbf{r}'\hat{\Sigma}\mathbf{r}}} \right)^2 = \frac{\varepsilon'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{r}\mathbf{r}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon}{\mathbf{r}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\hat{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{r}}, \quad (4)$$

where  $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\hat{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \equiv \hat{\Sigma}$  is a consistent estimator of  $\Sigma$ .

The following approaches to estimate  $\Sigma$  have been proposed in the literature. Each approach has a particular form of  $\hat{\Omega}$ , which leads to a different HC estimator for  $\hat{\Sigma}$ .

- HC0 is the original formulation used in White standard errors. [White \(1980\)](#) shows that, when using

$$\hat{\Omega} = \text{diag} \{e_i^2\}, \quad (5)$$

$\mathbf{X}'\hat{\Omega}\mathbf{X}/n = \sum_{i=1}^n e_i^2 \mathbf{x}_i \mathbf{x}_i' / n$  is a consistent estimator of  $\mathbf{X}'\Omega\mathbf{X}/n$ .

- HC1 ([Hinkley, 1977](#)) adjusts for degrees of freedom in HC0 and sets

$$\hat{\Omega} = \text{diag} \left\{ \frac{e_i^2}{1 - p/n} \right\}. \quad (6)$$

- HC2 ([MacKinnon and White, 1985](#)) adjusts for the leverage values  $h_{ii}$ ,

$$\hat{\Omega} = \text{diag} \left\{ \frac{e_i^2}{1 - h_{ii}} \right\}. \quad (7)$$

Note that, in the special case of a ‘balanced’ data for which  $h_{ii} = p/n$  for all  $i$ , HC2 becomes HC1 of [Hinkley \(1977\)](#). [MacKinnon and White \(1985\)](#) construct HC2 based on the results of [Horn et al. \(1975\)](#). Wu (1986, Equation 5.2 and Lemma 3) shows that, if  $h_{ij} = 0$  then  $E(e_i^2) = (1 - h_{ii})\sigma_i^2$ , and more generally  $E(e_i^2) = (1 - h_{ii})\sigma_i^2 + O(n^{-1})$ . Further, Wu (1986, Theorem 5) shows that  $\hat{\Sigma}$  using HC2,  $\hat{\Sigma} = (\mathbf{X}'\mathbf{X})^{-1} \{ \sum_{i=1}^n e_i^2 \mathbf{x}_i \mathbf{x}_i' / (1 - h_{ii}) \} (\mathbf{X}'\mathbf{X})^{-1}$  has similar properties, i.e., if  $h_{ij} = 0$  then  $E(\hat{\Sigma}) = \Sigma$ , and more generally  $E(\hat{\Sigma}) = \Sigma + O(n^{-1})$ .

- HC3 is introduced by [MacKinnon and White \(1985\)](#). HC3 is the ‘jackknife’ estimator of the covariance matrix of  $\hat{\beta}$  based on Efron (1982, Equation 3.13). [MacKinnon and White \(1985\)](#) shows that the jackknife estimator of  $\Sigma$  can be expressed as  $\hat{\Sigma} =$

$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{\Omega}}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$  with

$$\hat{\mathbf{\Omega}} = \text{diag} \left\{ \frac{e_i^2}{(1 - h_{ii})^2} \right\}.^4 \quad (8)$$

- HC4 (Cribari-Neto, 2004) adjusts the residuals by a leverage factor that increases with the leverage

$$\hat{\mathbf{\Omega}} = \text{diag} \left\{ \frac{e_i^2}{(1 - h_{ii})^{\delta_i}} \right\}, \quad (9)$$

where  $\delta_i = \min \left\{ 4, \frac{nh_{ii}}{p} \right\}$  increases with the leverage values  $h_{ii}$ .

The above five HC estimators can be expressed in a unified form by setting  $\hat{\mathbf{\Omega}} = \text{diag} \{g_i e_i^2\}$ , with  $g_i = 1$  for HC0,  $g_i = n/(n - p)$  for HC1,  $g_i = 1/(1 - h_{ii})$  for HC2,  $g_i = 1/(1 - h_{ii})^2$  for HC3, and  $g_i = 1/(1 - h_{ii})^{\delta_i}$  for HC4.

In applied work, inference using the  $F$  statistic with each of these HC estimators is usually based on its asymptotic distribution, because its exact finite sample distribution is not available. However, when the asymptotic distribution is used to assess the significance of an  $F$ -value, it is well known that the  $F$  statistic using each of these HC estimators suffers from substantial size distortion even when  $n$  is fairly large, especially when the data are leveraged/unbalanced. See MacKinnon and White (1985), Hausman and Palmer (2012), Long and Ervin (2000), MacKinnon (2013), among many others.

In view of this problem together with the fact that no exact finite sample distribution theory for the  $F$  statistic using HCs is available, our goal in this paper is to derive one, as presented in the next section.

### 3 Exact Distribution of the $F$ Statistic

In this section we show that the  $F$  statistic can be expressed as a ratio of quadratic forms, and therefore its exact cumulative distribution under the null hypothesis can be easily obtained from the result of Imhof (1961) on the distribution of a quadratic form.

#### 3.1 Exact Distribution of the $F$ Statistic Using HC

We now derive the exact distribution of the  $F$  test statistic shown in Equation (4) using each of HC0, HC1, HC2, HC3 and HC4. Denote  $d_i = \mathbf{r}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i$  and  $\mathbf{D} = \text{diag}(d_i^2)$ . The

---

<sup>4</sup>The original HC3 in MacKinnon and White (1985, Equation 12) is slightly different. See also Hausman and Palmer (2012, p. 233).

denominator of Equation (4) can be rewritten as  $\mathbf{e}'\mathbf{D}\mathbf{G}\mathbf{e}$ , where  $\mathbf{G} = \text{diag}(g_i)$ . Denoting the OLS residual vector  $\mathbf{e} = \mathbf{M}\boldsymbol{\varepsilon}$  with  $\mathbf{M} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ , Equation (4) can then be rewritten as a ratio of quadratic forms:

$$F = t^2 = \frac{\boldsymbol{\varepsilon}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{r}\mathbf{r}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}}{\boldsymbol{\varepsilon}'\mathbf{M}\mathbf{D}\mathbf{G}\mathbf{M}\boldsymbol{\varepsilon}} = \frac{\boldsymbol{\varepsilon}'\mathbf{A}\boldsymbol{\varepsilon}}{\boldsymbol{\varepsilon}'\mathbf{B}\boldsymbol{\varepsilon}}, \quad (10)$$

where  $\mathbf{A} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{r}\mathbf{r}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$  and  $\mathbf{B} = \mathbf{M}\mathbf{D}\mathbf{G}\mathbf{M}$ . Thus, the cumulative distribution function (CDF) is

$$\Pr(F \leq c) = \Pr(t^2 \leq c) = \Pr\left(\frac{\boldsymbol{\varepsilon}'\mathbf{A}\boldsymbol{\varepsilon}}{\boldsymbol{\varepsilon}'\mathbf{B}\boldsymbol{\varepsilon}} \leq c\right) = \Pr(\boldsymbol{\varepsilon}'(\mathbf{A} - c\mathbf{B})\boldsymbol{\varepsilon} \leq 0) = \Pr(\boldsymbol{\varepsilon}'\mathbf{N}\boldsymbol{\varepsilon} \leq 0), \quad (11)$$

where  $\mathbf{N} = \mathbf{A} - c\mathbf{B}$ .

**Remark 1:** Under heteroskedasticity, the  $F$  statistic is the ratio of two quadratic forms that may not be independent. Independence requires that  $\mathbf{A}\boldsymbol{\Omega}\mathbf{B} = 0$ , which however is not guaranteed since  $\boldsymbol{\Omega}$  can be any positive-definite diagonal matrix. This violates the independence assumption of the Snedecor's  $F$  distribution and raises questions for inference using the Snedecor's  $F$  distribution.

Following Imhof (1961) and assuming normal errors  $\varepsilon_i \sim N(0, \sigma_i^2)$  with heteroskedasticity, Equation (11) can be written as

$$\Pr(F \leq c) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{\sin \theta(u)}{u\rho(u)} du, \quad (12)$$

where

$$\theta(u) = \frac{1}{2} \sum_{j=1}^J [\nu_j \tan^{-1}(\lambda_j u)], \quad (13)$$

$$\rho(u) = \prod_{j=1}^J (1 + \lambda_j^2 u^2)^{\frac{1}{4}\nu_j}, \quad (14)$$

$\lambda_j$  is the non-zero eigenvalues of  $\mathbf{N}\boldsymbol{\Omega}$ ,  $J$  is the number of non-zero  $\lambda_j$ , and  $\nu_j$  is the degrees of freedom of  $\varepsilon_j^2$ . Specifically, in this situation, all of  $\nu_j = 1$  because  $\varepsilon_j^2$  follows the chi-squared distribution with the degree of freedom 1.<sup>5</sup> Then, substituting Equations (13) and (14) into

---

<sup>5</sup>The notation  $h_r, r = 1, \dots, m$ , used in Equations (1.1) and (3.2) of Imhof (1961) is  $\nu_j, j = 1, \dots, J$  in this paper. The values of  $\delta_r$  and  $x$  in these Imhof's equations are both zero in this paper.

Equation (12), we obtain the Imhof distribution function of the  $F$  statistic:

$$\Pr(F \leq c) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{\sin\left(\frac{1}{2} \sum_{j=1}^J [\tan^{-1}(\lambda_j u)]\right)}{u \prod_{j=1}^J (1 + \lambda_j^2 u^2)^{\frac{1}{4}}} du \equiv \text{Imhof}(\hat{\mathbf{\Omega}}, \mathbf{N}\mathbf{\Omega}, c). \quad (15)$$

The first argument  $\hat{\mathbf{\Omega}}$  of the function,  $\text{Imhof}(\hat{\mathbf{\Omega}}, \mathbf{N}\mathbf{\Omega}, c)$ , is to indicate that the  $F$  statistic in Equation (4) and Equation (10) uses an HC estimator  $\hat{\mathbf{\Omega}}$  discussed in Section 2. The second argument  $\mathbf{N}\mathbf{\Omega}$  of the function indicates that the eigenvalues  $\lambda_j$  in Equation (15) are those of  $\mathbf{N}\mathbf{\Omega}$ .

**Remark 2:** Note that, while the first argument  $\hat{\mathbf{\Omega}}$  of the function,  $\text{Imhof}(\hat{\mathbf{\Omega}}, \mathbf{N}\mathbf{\Omega}, c)$ , is estimated, the second argument  $\mathbf{N}\mathbf{\Omega}$  includes the unknown true  $\mathbf{\Omega}$ . Hence, the CDF,  $\text{Imhof}(\hat{\mathbf{\Omega}}, \mathbf{N}\mathbf{\Omega}, c)$ , will be referred to as the ‘‘Infeasible’’ Imhof distribution. On the other hand, when the estimated  $\mathbf{N}\hat{\mathbf{\Omega}}$  is used to compute the eigenvalues  $\lambda_j$  in Equation (15), we will refer to  $\text{Imhof}(\hat{\mathbf{\Omega}}, \mathbf{N}\hat{\mathbf{\Omega}}, c)$  as the ‘‘Feasible’’ Imhof distribution.

### 3.2 Exact Distribution of the $F$ Statistic When $\mathbf{\Omega}$ Is Known

To compare the exact distribution of the  $F$  statistic when different estimators  $\hat{\mathbf{\Omega}}$  are used, we also derive the oracle exact distribution of the  $F$  statistic using the true  $\mathbf{\Omega}$ . Hence, we call this oracle exact CDF as the ‘‘Benchmark’’, as it plays a benchmark role in the comparison. The Benchmark is computed as follows.

Assuming that  $\mathbf{\Omega}$  is known, Equation (4) can be rewritten as follows:

$$F^* = \frac{\varepsilon' \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{r} \mathbf{r}' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \varepsilon}{\mathbf{r}' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{\Omega} \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{r}} = \frac{\varepsilon' \mathbf{A} \varepsilon}{b^*} = \varepsilon' \left( \frac{\mathbf{A}}{b^*} \right) \varepsilon. \quad (16)$$

The  $F$  statistic using the true covariance matrix  $\mathbf{\Omega}$  is denoted as  $F^*$ . The CDF for  $F^*$  is

$$\Pr(F^* \leq c) = \Pr\left(\varepsilon' \left( \frac{\mathbf{A}}{b^*} \right) \varepsilon \leq c\right) = \Pr(\varepsilon' \mathbf{N}^* \varepsilon \leq c), \quad (17)$$

where  $\mathbf{N}^* \equiv \mathbf{A}/b^*$ , with  $\mathbf{A}$  being the same as before and  $b^* \equiv \mathbf{r}' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{\Omega} \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{r}$  being a scalar.

Similarly, using [Imhof \(1961\)](#), Equation (17) can be written as

$$\Pr(F^* \leq c) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{\sin \theta(u)}{u \rho(u)} du, \quad (18)$$

where



$$\theta(u) = \frac{1}{2} \sum_{j=1}^J [\nu_j \tan^{-1}(\lambda_j^* u)] - \frac{1}{2} cu, \quad (19)$$

$$\rho(u) = \prod_{j=1}^J (1 + \lambda_j^{*2} u^2)^{\frac{1}{4} \nu_j}, \quad (20)$$

$\lambda_j^*$  is the non-zero eigenvalues of  $\mathbf{L} \equiv \mathbf{N}^* \boldsymbol{\Omega}$ , and  $\nu_j = 1$  because  $\varepsilon_j^2$  follows the chi-squared distribution with 1 degree of freedom.

Then, substituting Equations (19) and (20) into Equation (18), we obtain the CDF of  $F^*$ , which we refer to as Benchmark,

$$\Pr(F^* \leq c) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{\sin\left(\frac{1}{2} \sum_{j=1}^J [\tan^{-1}(\lambda_j^* u)] - \frac{1}{2} cu\right)}{u \prod_{j=1}^J (1 + \lambda_j^{*2} u^2)^{\frac{1}{4}}} du \equiv \text{Benchmark}(\boldsymbol{\Omega}, \mathbf{L}, c). \quad (21)$$

The first argument  $\boldsymbol{\Omega}$  of the function,  $\text{Benchmark}(\boldsymbol{\Omega}, \mathbf{L}, c)$ , is to indicate the  $F^*$  in Equation (16) uses the true  $\boldsymbol{\Omega}$ . The second argument  $\mathbf{L}$  of the function indicates that the non-zero eigenvalues  $\lambda_j^*$  in Equation (21) are those of  $\mathbf{L} \equiv \mathbf{N}^* \boldsymbol{\Omega}$ .

**Remark 3:** Only one eigenvalue of  $\mathbf{L}$  is one and all the others are zero.<sup>6</sup> Hence the eigenvalues of  $\mathbf{L}$  do not depend on  $\boldsymbol{\Omega}$  or the sample size  $n$ . The Benchmark CDF,  $\Pr(F^* \leq c)$ , remains unchanged even when the sample size  $n$  or the error variance-covariance matrix  $\boldsymbol{\Omega} = \text{diag}\{\sigma_i^2\}$  changes. Furthermore, as pointed out by the referees, the benchmark distribution is equivalent to the  $\chi^2$  distribution with 1 degree of freedom.<sup>7</sup> This makes the Benchmark distribution a perfect benchmarking criterion to be compared with, justifying its name. These properties of the Benchmark distribution enable us to evaluate the exact Imhof distributions of the  $F$  statistic using different HC estimators, by comparing them with respect to the Benchmark distribution. That is what we do next.

<sup>6</sup>Note that  $\mathbf{L}$  and  $\mathbf{H} \equiv \boldsymbol{\Omega}^{1/2} \mathbf{L} \boldsymbol{\Omega}^{-1/2} = \boldsymbol{\Omega}^{1/2} (\mathbf{N}^*) \boldsymbol{\Omega}^{1/2}$  have the same characteristic equations and the same eigenvalues, because  $|\mathbf{L} - \lambda \mathbf{I}| = 0$  and  $|\boldsymbol{\Omega}^{1/2}| |\mathbf{L} - \lambda \mathbf{I}| |\boldsymbol{\Omega}^{-1/2}| = 0$ . As  $b^*$  is a scalar,  $\mathbf{H} = \boldsymbol{\Omega}^{1/2} \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{r} \mathbf{r}' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Omega}^{1/2} / b^*$  is symmetric and idempotent. Hence, the eigenvalues of  $\mathbf{H}$  and  $\mathbf{L}$  are either 1's or 0's. Further,  $\text{rank}(\mathbf{H}) = 1$  because  $\mathbf{H} = \xi \xi' / b^*$  with  $\xi \equiv \boldsymbol{\Omega}^{1/2} \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{r}$  being an  $n \times 1$  vector. Therefore, only one of the eigenvalues of  $\mathbf{H}$  or  $\mathbf{L}$ , is 1 and all the other eigenvalues are 0. Then, (19) and (20) are simplified as  $\theta(u) = \frac{1}{2} \tan^{-1}(u) - \frac{1}{2} cu$  and  $\rho(u) = (1 + u^2)^{\frac{1}{4}}$ .

<sup>7</sup>For  $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Omega})$ ,  $\mathbf{r}'(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y} - \mathbf{r}' \boldsymbol{\beta} \sim N(0, b^*)$ , where  $b^* = \mathbf{r}'(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Omega} \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{r}$ . Therefore,

$$\left( \frac{\mathbf{r}'(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y} - \mathbf{r}' \boldsymbol{\beta}}{\sqrt{b^*}} \right)^2 \sim \chi_1^2.$$

### 3.3 Comparing the Exact Distribution $\Pr(F \leq c)$ of $F$ Statistic Using a Different HC w.r.t. the Benchmark Distribution $\Pr(F^* \leq c)$

In this subsection, in order to assess the effect of using the HC covariance matrix estimators, we compare the Infeasible Imhof distribution  $\Pr(F \leq c)$  in Equation (15) with the Benchmark CDF  $\Pr(F^* \leq c)$  in Equation (21). To compare them under highly leveraged data, we adopt the same design as in Hansen (2017), based on the following heteroskedastic dummy variable regression

$$y_i = \beta_0 + \sum_{k=1}^3 \beta_k x_{ik} + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma_i^2), \quad i = 1, \dots, n. \quad (22)$$

Each dummy variable  $x_{ik}$  equals 1 only for 3 observations (i.e.  $\sum_{i=1}^n x_{ik} = 3$  for  $k = 1, 2, 3$ ). We are interested only in the coefficient  $\beta_1$ . In simulation, we use  $n \in \{15, 30, 50, 100, 150, 500\}$ ,  $c \in (0, 20]$ , and

$$\sigma^2(x_1) = I(x_1 = 1) + aI(x_1 = 0), \quad (23)$$

with  $I(\cdot)$  being the indicator function. We consider two cases with  $a = 0.5$  (Case 1) and  $a = 2$  (Case 2).

The CDF plots of  $\Pr(F \leq c)$  using different HCs and the Benchmark CDF  $\Pr(F^* \leq c)$  are presented in Figure 1 (for Case 1) and Figure 2 (for Case 2). Since the results for  $n > 50$  are similar, we show only  $n = 15$  and  $n = 50$  in each figure. Note that, in these cases with heteroskedasticity, the  $F$  statistic in Equation (4) does not follow the Snedecor  $F$ -distribution, and using the  $F$ -table is incorrect. Therefore we use the Benchmark as an evaluation criterion. We compare the (vertical) distance between the CDF plots of  $\Pr(F \leq c)$  and the Benchmark CDF  $\Pr(F^* \leq c)$ , to find which HC yields the best exact distribution closest to the Benchmark. While the distance can be formally measured by the Kolmogorov-Smirnov statistic or the Cramer-von Mises statistic, Figures 1 and 2 are clear enough to tell the differences between the CDF plots.

[Insert Figures 1 and 2 Here]

As we have noted in Remark 2 after presenting Equation (21), the Benchmark distribution remains unchanged even when the sample size changes from  $n = 15$  in the upper panel to  $n = 50$  in the lower panel of each of Figures 1 and 2, or when the error variance-covariance matrix  $\mathbf{\Omega}$  changes from Case 1 to Case 2 in these figures. This invariance property makes the Benchmark CDF as a benchmark to be compared with the Infeasible Imhof CDF plots using different HC estimators. Figures 1 and 2 show that, among the Infeasible Imhof CDFs

from using HC0, HC1, HC2 and HC4, the one from using HC3 is the closest to Benchmark, indicating that HC3 is the most precise among these HC estimators.

## 4 Monte Carlo Simulation

In this section, we examine the empirical size behavior of the  $F$  statistic using the HC estimators in finite samples by Monte Carlo simulation. MacKinnon and White (1985), and Cribari-Neto and Lima (2010) have studied the behavior of the HC estimators through Monte Carlo simulations. In a recent paper, MacKinnon (2013) provided an extensive discussion on the setup of Monte Carlo simulations for studying the performance of HC estimators. We consider the following simple linear model, that is similar to the one in MacKinnon (2013), for the data generating process (DGP):

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma_i^2), \quad \sigma_i^2 = \exp(x_i^2), \quad (24)$$

where  $\beta = (\beta_0, \beta_1)' = (0, 0)'$  and  $x_i \sim N(0, 1)$ . To examine the size behavior of the  $F$  statistic, the null hypothesis  $H_0 : \mathbf{r}'\beta = r_0$  with  $\mathbf{r}' = (0 \ 1)$  and  $r_0 = 0$  is considered.

### 4.1 Size of the $F$ Statistic Using the Infeasible Imhof Distribution

In this subsection, we employ the true value of variance-covariance matrix ( $\mathbf{\Omega}$ ) to calculate the Infeasible Imhof distribution in Equation (15). The Monte Carlo average values of the Rejection Rate of the  $F$  statistic under the null hypothesis, i.e., the size of the test, are computed from the following steps:

1. Randomly generate  $x_i$  from  $N(0, 1)$ .
2. Randomly generate  $\varepsilon_i$  from  $N(0, \sigma_i^2)$  where  $\sigma_i^2 = \exp(x_i^2)$  and calculate  $y_i$  following (24) under the null hypothesis that  $\mathbf{r}'\beta = 0$ .
3. Calculate the  $F$  statistic using each of HC0, HC1, HC2, HC3 and HC4.
4. Find the Imhof critical value  $c_\alpha^{\text{Imhof}}$  from Equation (15), satisfying

$$\text{Imhof}(\hat{\mathbf{\Omega}}, \mathbf{N}\mathbf{\Omega}, c_\alpha^{\text{Imhof}}) = \alpha. \quad (25)$$

Also find the critical value from the  $F$ -table and denote it as  $c_\alpha^{\text{F}}$ .

5. Repeat Steps (2)-(4) for 1000 times and calculate rejection rates

$$\text{Rejection Rate}^{\text{Imhof}} = \frac{1}{1000} \sum I(F > c_{\alpha}^{\text{Imhof}}), \quad (26)$$

$$\text{Rejection Rate}^{\text{F}} = \frac{1}{1000} \sum I(F > c_{\alpha}^{\text{F}}). \quad (27)$$

Note that Steps (2)-(4) are conditional on the data matrix  $\mathbf{X}$  generated from Step (1).

6. Repeat Steps (1)-(5) 100 times (for different dataset of  $\mathbf{X}$ ), and calculate the average value of Rejection Rates over the 100 replications. We then present them in Table 1 for  $\alpha \in \{0.10, 0.05, 0.01\}$ .

Also, presented in Table 1 are the Monte Carlo average values of Rejection Rate using the Benchmark cumulative distribution shown in Equation (21), i.e., using the Benchmark critical value,  $c_{\alpha}^{\text{Benchmark}}$ , satisfying

$$\text{Benchmark}(\mathbf{\Omega}, \mathbf{N}^* \mathbf{\Omega}, c_{\alpha}^{\text{Benchmark}}) = \alpha. \quad (28)$$

Table 1 shows the size performance of the  $F$  test using the Infeasible Imhof critical values and the Benchmark critical values. All the size values using the Imhof critical values are generally very close to the nominal significance levels. This indicates that the Infeasible Imhof critical values are pretty precise. They are much better than the size values computed using the critical values from the  $F$ -table.

[Insert Table 1 Here]

## 4.2 Size of the $F$ Statistic Using the Feasible Imhof Distribution

Table 1 considers the Infeasible Imhof distribution which employs the unknown true variance-covariance matrix  $\mathbf{\Omega}$  of the regression error term. As the true value of  $\mathbf{\Omega}$  is unknown in practice, we obtain the Feasible Imhof distribution by replacing  $\mathbf{\Omega}$  with  $\hat{\mathbf{\Omega}}$  in Equation (25), i.e., using the Imhof critical values,  $c_{\alpha}^{\text{Imhof}}$ , satisfying

$$\text{Imhof}(\hat{\mathbf{\Omega}}, \mathbf{N} \hat{\mathbf{\Omega}}, c_{\alpha}^{\text{Imhof}}) = \alpha. \quad (29)$$

The size results of using the Feasible Imhof distribution are presented in the left panel of Table 2 under the heading of ‘‘Using Feasible Imhof Distribution’’. Note that there are two more panels in Table 1 and Table 2, which are labeled as ‘‘Using  $F$ -table’’ and ‘‘Using Benchmark’’. Although the values of these two panels in Table 1 are identical to those in Table 2, we repeat reporting them just for ease of comparison.

[Insert Table 2 Here]

From Table 2, we make the following observations:

1. The empirical size behavior of the  $F$  statistic using the Feasible Imhof distribution (Table 2) is only slightly worse than the size using the Infeasible Imhof distribution (Table 1). The size is generally very good except when the sample size is very small ( $n = 15$ ).
2. Using the Feasible Imhof distribution is almost always much better than using the  $F$ -table.
3. When the Feasible Imhof distribution is used, HC3 exhibits only slightly better size behavior in comparison with HC0, HC1, HC2, and HC4. All HC estimators are somewhat similar when the exact Imhof distribution is employed. It indicates that the selection between HC estimators appears irrelevant for empirical researches if the exact Imhof distribution is used.
4. When the  $F$ -table is used, HC3 is clearly much better than HC0, HC1, HC2, and is slightly better than HC4.
5. When the test statistic's distribution is misspecified as a Snedecor  $F$  distribution (i.e., using the  $F$  table), it appears that the empirical size decreases monotonically towards zero below the nominal size as the sample size  $n$  increases. This is much less so when the Imhof distribution is used (in both Tables 1 and 2).

### **4.3 Size of the $F$ Statistic Using the Feasible Imhof Distribution (with Skewed-normal Errors)**

To test the robustness of our method, in this subsection, we consider the regression error term following a skewed-normal distribution to investigate the size performance of the  $F$  statistic using its Feasible Imhof distribution. Tables 3 and 4, respectively, show the size of the  $F$  statistic using the Feasible Imhof distribution with skewed-normal errors for mild skewness value of .5 and more severe skewness of .9

[Insert Tables 3 and 4 Here]

From Tables 3 and 4, we find the following.

1. Comparing Tables 2 and 3, the empirical size behavior of the  $F$  statistic using the Feasible Imhof distribution under the skewed-normal error distribution with a mild skewness is similar with its size behavior under the symmetric normal error distribution. However, the empirical size behavior of the  $F$  statistic using the  $F$  table under skewed-normal error distribution with a mild skewness is obviously worse, especially when the sample size is small  $n = 15$ , than its size using the  $F$  table under the symmetric normal distribution.
2. Comparing Tables 3 and 4, in most cases, the empirical size behavior of the  $F$  statistic using the Feasible Imhof distribution under skewed-normal errors with severe skewness is only slightly worse than its size using the Feasible Imhof distribution under the skewed-normal error terms with mild skewness.
3. In general, using the Feasible Imhof distribution is almost always much better than using the  $F$ -table even when the error terms follow a skew-normal distribution.

#### 4.4 Size of the $F$ Statistic Using the Feasible Imhof Distribution (with Non-normal $t(5)$ Errors)

In the previous subsections, we have assumed that the regression error term follows a normal distribution and skewed normal distributions. In practical situations, the error terms may have fatter tails than a normal distribution. We now consider that the error term follows the Student- $t$  distribution with 5 degrees of freedom to investigate the size performance of the  $F$  statistic using its Feasible Imhof distribution.

[Insert Table 5 Here]

From Table 5, we find the following. Although the derivation of the exact Imhof distribution in Section 3 is under the assumption that the regression errors follow the normal distribution, the results with a non-normal distribution,  $t(5)$ , do not change the conclusions that we have obtained with the normal distribution in Tables 1 and 2. Importantly, all the observations we have made from Table 2, as summarized in subsection 4.2, remain valid under the  $t(5)$  distribution with fat tails. To reiterate, it means that using the Imhof distribution is almost always much better than using the  $F$ -table. If the Imhof distribution is used, all the HC estimators are somewhat similar.

## 5 Empirical Application

How do we use the exact Imhof distribution of the  $F$  statistic in empirical applications? How does it work relative to using the  $F$ -table or relative to the bootstrap methods? To answer these questions, we take the housing price data from a popular econometrics textbook by Jeffrey Wooldridge. The data called HPRICE1 in [Wooldridge \(2016\)](#) was collected from the real estate pages of the Boston Globe newspaper during 1990. It includes home sales data in the Boston area with the following information. For home  $i$ ,  $y_i$  denotes the natural logarithm of the house price,  $x_{i1}$  denotes the natural logarithm of the assessment value,  $x_{i2}$  the number of bedrooms,  $x_{i3}$  the natural logarithm of the lot size in square feet,  $x_{i4}$  denotes the natural logarithm of the house size in square feet, and  $x_{i5}$  is a dummy variable with  $x_{i5} = 1$  if the home is of colonial style and 0 otherwise. The data is highly leveraged and heteroskedastic. The sample size is  $n = 88$ .

The  $p$ -values of the  $F$  statistic to test for the null hypothesis of a single linear restriction on the coefficient vector  $\beta$ , are computed in Table 6, where six different single linear restrictions are considered. In the table, we report the four different  $p$ -values of the  $F$  statistic to test for each restriction. The  $p$ -values are computed in the following three different methods.

The first method of computing the  $p$ -value of an  $F$ -value is from using the Feasible Imhof distribution function in Equation (15) with the  $c$  value being replaced by the  $F$ -value. The “Imhof  $p$ -value” (which is the exact  $p$ -value) of an  $F$  value is computed from

$$1 - \text{Imhof}(\hat{\Omega}, \mathbf{N}\hat{\Omega}, F). \quad (30)$$

The R code for computing the Imhof  $p$ -value will be available from the authors’ websites.

The second method of computing the  $p$ -value of an  $F$ -value is from using the  $F$ -table. Although this is not correct, we present it here to show how wrong it is. We can tell this by comparing the  $F$ -table  $p$ -value and the wild bootstrap  $p$ -value.<sup>8</sup>

The third method of computing the  $p$ -value of an  $F$ -value is from using the wild bootstrap, which takes following steps. The model in Equation (1) is estimated by the ordinary least squares method,  $y_i = \hat{\beta}_0 + \sum_{k=1}^5 \hat{\beta}_k x_{ik} + e_i$ , where  $\hat{\beta}_k$ ’s are the OLS estimates and  $e_i$  is the estimated residual. Let  $e_i^{(b)} = f(e_i)v_i^{(b)}$ , where  $f(e_i) = e_i/(1 - h_{ii})$ . The choice for  $v_i^{(b)}$  is based on either Mammen’s two-point distribution suggested by [MacKinnon \(2013\)](#) or the Rademacher distribution proposed by [Davidson and Flachaire \(2008\)](#). Once the wild

---

<sup>8</sup>Unlike in simulation where we have the Benchmark in (21) which can be used to evaluate the performance of the Imhof distribution, we do not have such a benchmark criterion in empirical applications. Thus we compare the feasible Imhof  $p$ -values with the wild bootstrap  $p$ -values. Many simulation studies suggested that the wild bootstrap gives good performance (e.g., [Flachaire, 2005](#)), while some recent studies, e.g., [Djogbenou et al. \(2019\)](#), suggested that alternative variants of the wild bootstrap may perform quite differently. Here we use the wild bootstrap just to verify that the Imhof procedure works.

bootstrapped residuals  $e_i^{(b)}$  are obtained, the wild bootstrap data  $y_i^{(b)}$  can be generated from  $y_i^{(b)} = \hat{\beta}_0 + \sum_{k=1}^5 \hat{\beta}_k x_{ik} + e_i^{(b)}$ , conditional on the observed data of  $x_{ik}$ . Then, the model is estimated for the bootstrap samples and the associated values of the test statistic  $F^{(b)}$ , for  $b = 1, \dots, B$ , are calculated. Repeat these steps  $B$  times in order to estimate the wild bootstrap  $p$ -value, which is  $\frac{1}{B} \sum_{b=1}^B I(F^{(b)} > F)$ .

[Insert Table 6 Here]

The four different  $p$ -values of the  $F$  statistic for each of the six different single linear restrictions are presented in Table 6, where the wild bootstrap  $p$ -values are computed with  $B = 10,000$ . Table 6 presents the  $p$ -values of the  $F$  statistic using each of the five HC estimators (HC0, HC1, ..., HC4). In order to evaluate the Imhof  $p$ -values, we need a benchmark to compare them with. However, for empirical applications, we can not compute the Benchmark distribution in Equation (21). As we know the wild bootstrap “works” in the present setup of linear models (see Hausman and Palmer (2012) for some simulation results), we evaluate the  $p$ -values by examining how close the Imhof  $p$ -values from Equation (30) and the  $F$ -table  $p$ -values are to the wild bootstrap  $p$ -values. From Table 6, we find the following.

1. Imhof  $p$ -values are very close to the wild bootstrap  $p$ -values.
2. In most of the cases, Imhof  $p$ -values with HC3 are closer to the wild bootstrap  $p$ -values than Imhof  $p$ -values with the other HC estimators.
3. The distances between  $F$ -table  $p$ -values and Wild bootstrap  $p$ -values are larger than the distances between Imhof  $p$ -values and Wild bootstrap  $p$ -values.
4.  $F$ -table  $p$ -values with HC3 are generally closer to the wild bootstrap  $p$ -values than  $F$ -table  $p$ -values with the other HC estimators.
5. Although Imhof  $p$ -values are as good as the wild bootstrap  $p$ -values, the computation time of wild bootstrap is far longer than the Imhof procedure. For example, the computation of the Imhof procedure takes less than a second on a Windows notebook computer, while the computation of the wild bootstrap takes 45 seconds for each  $p$ -value.
6. In a few cases in Table 6, we may observe that the  $F$ -table  $p$ -values with HC3 are very close to the Imhof  $p$ -values, e.g.,  $\beta_3$ ,  $\beta_4$  and  $\beta_5$ . We speculate that these are for coefficients of variables that may be more balanced and less-leveraged. However, we also note that this is largely with HC3. Even for these three cases, the  $F$ -table  $p$ -values and the Imhof  $p$ -values are quite different when HC0, HC1, HC2, and HC4 are used.



This empirical application highlights that the  $p$ -values obtained from the Feasible Imhof distribution of the  $F$  statistic does increase the accuracy of inference in empirical research as the results demonstrate that the use of the Feasible Imhof distribution delivers the  $p$ -values similar to the wild bootstrap  $p$ -values.

## 6 Conclusions

When the regression errors are homoskedastic, the exact finite sample distribution of the  $F$  statistic is the Snedecor  $F$ -distribution. When the regression errors are heteroskedastic, the exact finite sample distribution of the  $F$  statistic using an HC covariance matrix estimator is unknown. The common practice in applied research is either to use the Snedecor  $F$ -distribution (which is wrong) or to use the asymptotic distribution. It is well known that the asymptotic inference based on White's (1980) HC standard errors (known as HC0) tends to over-reject when the sample is not large or leveraged/unbalanced. Various corrections (HC1, HC2, HC3, HC4, and more) to White's HC0 have been proposed to improve the finite sample behavior. However their over-rejection remains problematic even in fairly large samples.

In this paper, we have derived the exact finite sample distribution of the  $F$  statistic using the HC covariance matrix estimators (HC0, HC1, HC2, HC3 and HC4) when the regression errors are heteroskedastic and normal. It is shown that the  $F$  statistic under the null hypothesis using various HC covariance matrix estimators can be expressed as a ratio of quadratic forms, and therefore its exact cumulative distribution under the null hypothesis can be easily written from the result of Imhof (1961). A numerical calculation is carried out to present the exact (infeasible) Imhof distribution of the  $F$  statistic using different HCs.

We examine the Imhof procedure when the estimated parameters are used in the  $F$ -distribution, which we call the Feasible Imhof distribution. Simulation results show the Feasible Imhof distribution is remarkably reliable, and in comparison, the critical values from the  $F$ -table lead to a serious over-rejection problem when the sample is not large or leveraged/unbalanced. The  $F$ -type test statistic based on various HC covariance estimators can provide more accurate testing results in terms of the test size behavior when the Feasible Imhof distribution is used.

We have also examined robustness of our Imhof based  $F$ -distribution under non-normal distributions of errors. Given that the Imhof procedure assumes normal errors, it is important to examine how it performs under some severe departures from the normality such as the  $t_5$  distribution and departures from symmetry.

Finally, we present an empirical application which demonstrates that the use of the Feasible Imhof distribution can improve the accuracy of inference in empirical research as

evidenced by the similarity of the Feasible Imhof  $p$ -values and the wild bootstrap  $p$ -values.

## References

- Angrist, J. and J. S. Pischke (2009). *Mostly Harmless Econometrics*. Princeton University Press.
- Bao, Y., A. Ullah, and Y. Wang (2017). Distribution of the mean reversion estimator in the Ornstein-Uhlenbeck process. *Econometric Reviews* 36(6-9), 1039–1056.
- Chesher, A. and G. Austin (1991). The finite-sample distributions of heteroskedasticity robust wald statistics. *Journal of Econometric* 47, 153–173.
- Cribari-Neto, F. (2004). Asymptotic Inference under Heteroskedasticity of Unknown Form. *Computational Statistics and Data Analysis* 45(2), 215–233.
- Cribari-Neto, F. and M. D. G. A. Lima (2010). Sequences of Bias-adjusted Covariance Matrix Estimators under Heteroskedasticity of Unknown Form. *Annals of the Institute of Statistical Mathematics* 62(6), 1053–1082.
- Davidson, R. and E. Flachaire (2008). The Wild Bootstrap, Tamed at Last. *Journal of Econometrics* 146(1), 162–169.
- Djogbenou, A. A., J. G. MacKinnon, and M. O. Nielsen (2019). Asymptotic theory and wild bootstrap inference with clustered errors. *Journal of Econometrics* 212(2), 393 – 412.
- Eicker, F. (1963). Asymptotic Normality and Consistency of the Least Squares Estimators for Families of Linear Regressions. *The Annals of Mathematical Statistics* 34(2), 447–456.
- Farebrother, R. W. (1985). Eigenvalue-free methods for computing the distribution of a quadratic form in normal variables. *Statistische Hefte* 26(1), 287.
- Flachaire, E. (2005). Bootstrapping heteroskedastic regression models: wild bootstrap vs. pairs bootstrap. *Computational Statistics and Data Analysis* 49(2), 361–376.
- Hansen, B. E. (2017). The Exact Distribution of the t-ratio with Robust and Clustered Standard Errors. *University of Wisconsin*.
- Hausman, J. and C. Palmer (2012). Heteroskedasticity-robust Inference in Finite Samples. *Economics Letters* 116(2), 232–235.
- Hinkley, D. V. (1977). Jackknifing in Unbalanced Situations. *Technometrics* 19(3), 285–292.
- Horn, S. D., R. A. Horn, and D. B. Duncan (1975). Estimating Heteroscedastic Variances in Linear Models. *Journal of the American Statistical Association* 70(350), 380–385.

- Huber, P. (1967). The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions. *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability 1*, 221–233.
- Imbens, G. W. and M. Kolesár (2016). Robust Standard Errors in Small Samples: Some Practical Advice. *Review of Economics and Statistics 98*(4), 701–712.
- Imhof, J. P. (1961). Computing the Distribution of Quadratic Forms in Normal Variables. *Biometrika 48*(3 and 4), 419–426.
- Long, J. S. and L. H. Ervin (2000). Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model. *The American Statistician 54*(3), 217–224.
- MacKinnon, J. G. (2013). Thirty Years of Heteroskedasticity-Robust Inference. In *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis, Essays in Honor of Halbert L. White Jr*, pp. 437–461. Edited by X. Chen and N.R. Swanson, Springer.
- MacKinnon, J. G. and H. White (1985). Some Heteroskedasticity-consistent Covariance Matrix Estimators with Improved Finite Sample Properties. *Journal of Econometrics 29*(3), 305–325.
- McCaffrey, D. F. and R. M. Bell (2002). Bias Reduction in Standard Errors for Linear and Generalized Linear Models with Multi-Stage Samples. *Survey Methodology 28*(2), 169–181.
- Nakamura, A. and M. Nakamura (1998). Model Specification and Endogeneity. *Journal of Econometrics 83*(1-2), 213–237.
- Rao, C. R. (1970). Estimation of Heteroscedastic Variances in Linear Models. *Journal of the American Statistical Association 65*(329), 161–172.
- Ullah, A. (2004). *Finite Sample Econometrics*. Oxford University Press.
- White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica 48*(4), 817–838.
- Wooldridge, J. M. (2016). *Introductory Econometrics: A Modern Approach*. Cengage Learning.
- Young, A. (2016). Improved, Nearly Exact, Statistical Inference with Robust and Clustered Covariance Matrices Using Effective Degrees of Freedom Corrections, London School of Economics, Technical Report.

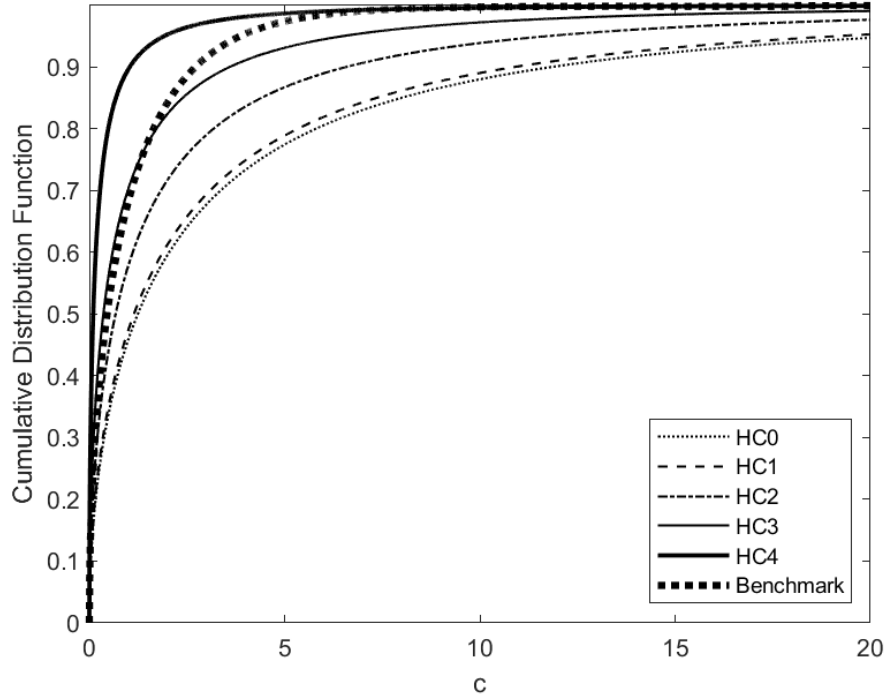
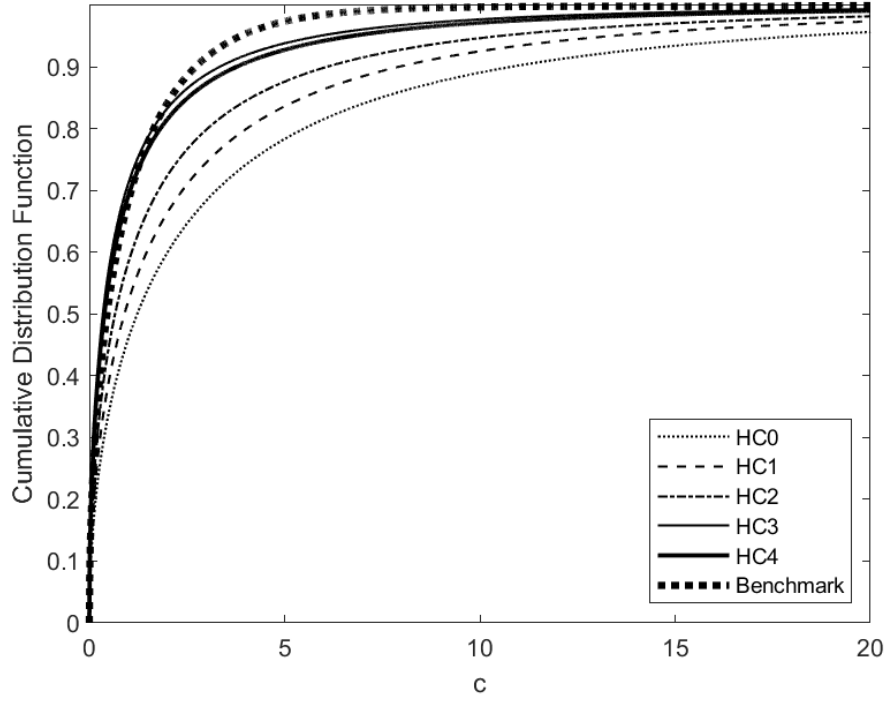


Figure 1: CDF for sample size  $n = 15$  (upper figure),  $n = 50$  (lower figure)  
 Case 1 with  $a = 0.5$  in Equation (23)

Notes: The exact Infeasible Imhof CDFs of  $F$  using different HCs are calculated using  $\text{Imhof}(\hat{\Omega}, \mathbf{N}\Omega, c)$  in Equation (15), and Benchmark is calculated using  $\text{Benchmark}(\Omega, \mathbf{L}, c)$  in Equation (21).

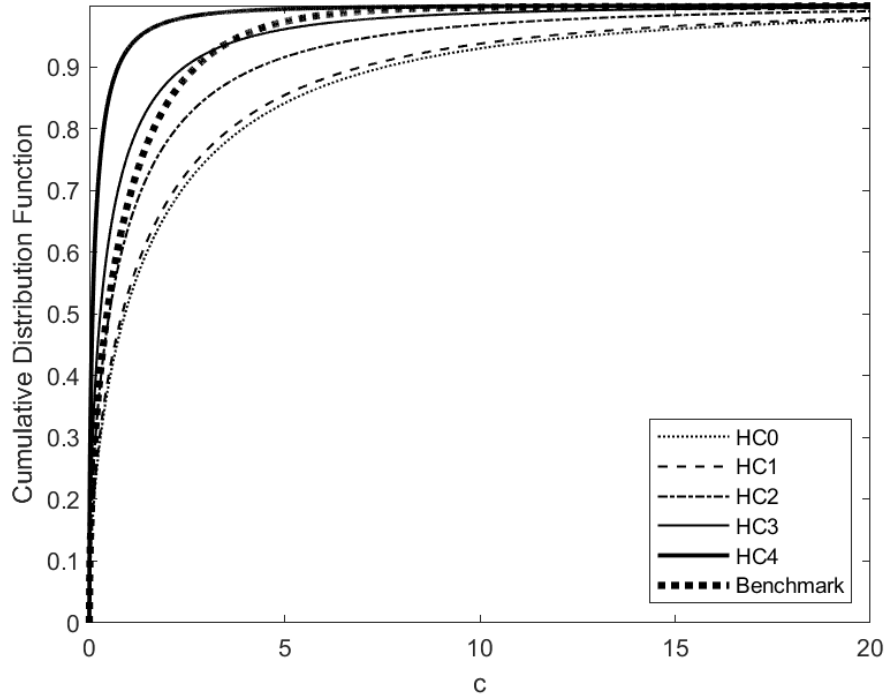
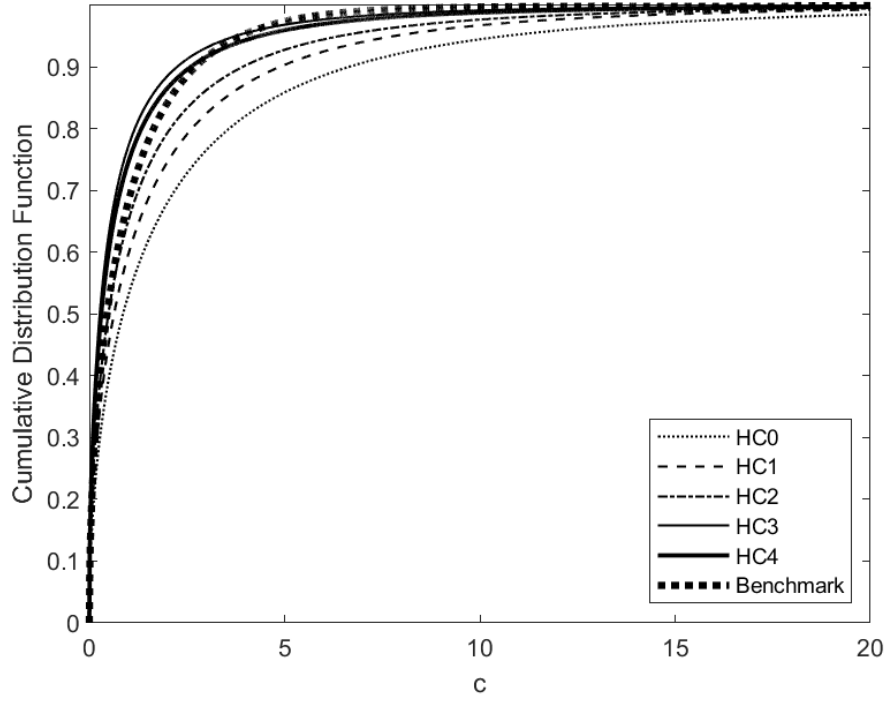


Figure 2: CDF for sample size  $n = 15$  (upper figure),  $n = 50$  (lower figure)  
Case 2 with  $a = 2$  in Equation (23)

Notes: The exact Infeasible Imhof CDFs of  $F$  using different HCs are calculated using  $\text{Imhof}(\hat{\Omega}, \mathbf{N}\Omega, c)$  in Equation (15), and Benchmark is calculated using  $\text{Benchmark}(\Omega, \mathbf{L}, c)$  in Equation (21).

Table 1: Size of the  $F$  Statistic Using the Infeasible Imhof Distribution

Heteroskedastic Normal Error Terms ( $\sigma^2(x) = \exp(x^2)$ )															
$n$	Using Infeasible Imhof Distribution				Using $F$ -table				Using Benchmark						
	HC0	HC1	HC2	HC3	HC4	HC0	HC1	HC2	HC3	HC4	HC0	HC1	HC2	HC3	HC4
$\alpha = 0.10$															
15	0.10010	0.10010	0.10061	0.09973	0.09912	0.27489	0.21714	0.15372	0.09780	0.08551					0.10091
30	0.10128	0.10128	0.10154	0.10157	0.10034	0.17675	0.15911	0.12943	0.09467	0.07539					
50	0.10081	0.10081	0.10092	0.10109	0.10106	0.13744	0.12867	0.11080	0.08780	0.06905					
100	0.09995	0.09995	0.09985	0.09982	0.10016	0.11201	0.10734	0.09641	0.08293	0.06862					
200	0.09835	0.09835	0.09845	0.09832	0.09833	0.08628	0.08426	0.07856	0.07084	0.06001					
$\alpha = 0.05$															
15	0.05007	0.05001	0.05015	0.05026	0.05011	0.15582	0.12671	0.09152	0.05574	0.05058					0.04994
30	0.04939	0.04939	0.04930	0.04942	0.04940	0.09795	0.08684	0.07131	0.05037	0.04152					
50	0.04924	0.04924	0.04926	0.04916	0.04903	0.06928	0.06320	0.05172	0.03896	0.03056					
100	0.04972	0.04972	0.04973	0.04953	0.04912	0.04828	0.04597	0.04072	0.03410	0.02701					
200	0.04880	0.04880	0.04873	0.04869	0.04872	0.03376	0.03293	0.03005	0.02640	0.02232					
$\alpha = 0.01$															
15	0.01452	0.01269	0.01092	0.01009	0.01021	0.05890	0.04517	0.03183	0.01734	0.01552					0.00987
30	0.01012	0.01012	0.01001	0.01019	0.01011	0.02737	0.02289	0.01780	0.01155	0.00951					
50	0.01001	0.01001	0.00998	0.00993	0.01012	0.01552	0.01374	0.01106	0.00784	0.00612					
100	0.01025	0.01025	0.01034	0.01027	0.01039	0.00776	0.00715	0.00626	0.00507	0.00403					
200	0.00989	0.00989	0.00989	0.00988	0.00978	0.00504	0.00488	0.00437	0.00378	0.00314					

Notes: As noted in Remark 1, the Infeasible Imhof distribution of the  $F$  statistic is  $\text{Imhof}(\hat{\Omega}, \mathbf{N}\hat{\Omega}, c)$  obtained from the cumulative distribution function in Equation (15). The Benchmark distribution is computed from the cumulative distribution function in Equation (21). In Tables 2-5, the Feasible Imhof distribution,  $\text{Imhof}(\hat{\Omega}, \mathbf{N}\hat{\Omega}, c)$  obtained with  $\mathbf{N}\hat{\Omega}$  estimated using the HC estimators in Section 2.

Table 2: Size of the  $F$  Statistic Using the Feasible Imhof Distribution

Heteroskedastic Normal Error Terms ( $\sigma^2(x) = \exp(x_i^2)$ )														
Using Feasible Imhof Distribution					Using $F$ -table					Using Benchmark				
$n$	HC0	HC1	HC2	HC3	HC4	HC0	HC1	HC2	HC3	HC4				
$\alpha = 0.10$														
15	0.13664	0.13664	0.12538	0.11759	0.11156	0.27489	0.21714	0.15372	0.09780	0.08551				0.10091
30	0.10953	0.10953	0.10479	0.10137	0.09639	0.17675	0.15911	0.12943	0.09467	0.07539				
50	0.10028	0.10028	0.09773	0.09528	0.09144	0.13744	0.12867	0.11080	0.08780	0.06905				
100	0.09818	0.09818	0.09711	0.09631	0.09431	0.11201	0.10734	0.09641	0.08293	0.06862				
200	0.09928	0.09928	0.09870	0.09832	0.09748	0.08628	0.08426	0.07856	0.07084	0.06001				
$\alpha = 0.05$														
15	0.08943	0.08941	0.08086	0.07419	0.06749	0.15582	0.12671	0.09152	0.05574	0.05058				0.04994
30	0.05799	0.05799	0.05439	0.05182	0.04739	0.09795	0.08684	0.07131	0.05037	0.04152				
50	0.04904	0.04904	0.04739	0.04591	0.04316	0.06928	0.06320	0.05172	0.03896	0.03056				
100	0.04015	0.04015	0.03954	0.03881	0.03738	0.04828	0.04597	0.04072	0.03410	0.02701				
200	0.04464	0.04464	0.04434	0.04420	0.04332	0.03376	0.03293	0.03005	0.02640	0.02232				
$\alpha = 0.01$														
15	0.03966	0.03946	0.03465	0.03095	0.02669	0.05890	0.04517	0.03183	0.01734	0.01552				0.00987
30	0.01673	0.01673	0.01543	0.01429	0.01261	0.02737	0.02289	0.01780	0.01155	0.00951				
50	0.01070	0.01070	0.01002	0.00951	0.00874	0.01552	0.01374	0.01106	0.00784	0.00612				
100	0.00867	0.00867	0.00840	0.00812	0.00761	0.00776	0.00715	0.00626	0.00507	0.00403				
200	0.00760	0.00760	0.00738	0.00722	0.00702	0.00504	0.00488	0.00437	0.00378	0.00314				

Notes: See Table 1.



Table 3: Size of the  $F$  Statistic Using the Feasible Imhof Distribution with Skewed-normal Errors

Heteroskedastic Skewed-Normal Error Terms (Skewness=0.5) ( $\sigma^2(x) = \exp(x^2)$ )												
Using Feasible Imhof Distribution						Using $F$ -table						
$n$	HC0	HC1	HC2	HC3	HC4	HC0	HC1	HC2	HC3	HC4	HC0	HC4
$\alpha = 0.10$												
15	0.13751	0.13751	0.12652	0.11916	0.11354	0.25963	0.22257	0.16130	0.10586	0.09728		
30	0.10731	0.10731	0.10248	0.09867	0.09397	0.17323	0.15587	0.12747	0.09330	0.07460		
50	0.10129	0.10129	0.09863	0.09654	0.09266	0.14477	0.13507	0.11313	0.08807	0.06827		
100	0.09008	0.09008	0.08915	0.08839	0.08615	0.10160	0.09715	0.08733	0.07468	0.05927		
200	0.09609	0.09609	0.09558	0.09504	0.09405	0.09228	0.09013	0.08391	0.07625	0.06501		
$\alpha = 0.05$												
15	0.09072	0.09072	0.08054	0.07298	0.06594	0.17055	0.13766	0.10107	0.06030	0.05122		
30	0.06366	0.06366	0.05958	0.05662	0.05212	0.10391	0.09203	0.07321	0.05137	0.04134		
50	0.04893	0.04893	0.04721	0.04577	0.04292	0.07164	0.06550	0.05306	0.03979	0.03027		
100	0.04838	0.04838	0.04764	0.04695	0.04534	0.05214	0.04952	0.04349	0.03675	0.02914		
200	0.04348	0.04348	0.04319	0.04288	0.04205	0.03457	0.03357	0.03064	0.02695	0.02206		
$\alpha = 0.01$												
15	0.03526	0.03475	0.03017	0.02669	0.02370	0.05682	0.04381	0.03108	0.01705	0.01546		
30	0.01765	0.01765	0.01602	0.01487	0.01330	0.02669	0.02247	0.01742	0.01117	0.00951		
50	0.01084	0.01084	0.01029	0.00994	0.00896	0.01540	0.01346	0.01086	0.00770	0.00617		
100	0.00904	0.00904	0.00874	0.00849	0.00802	0.00913	0.00834	0.00729	0.00554	0.00433		
200	0.00733	0.00733	0.00723	0.00717	0.00696	0.00510	0.00492	0.00435	0.00388	0.00306		

Notes: See Table 1.

Table 4: Size of the  $F$  Statistic Using the Feasible Imhof Distribution with Skewed-normal Errors

Heteroskedastic Skewed-Normal Error Terms (Skewness=0.9) ( $\sigma^2(x) = \exp(x^2)$ )												
Using Feasible Imhof Distribution						Using $F$ -table						
$n$	HC0	HC1	HC2	HC3	HC4	HC0	HC1	HC2	HC3	HC4		
$\alpha = 0.10$												
15	0.13003	0.13003	0.11870	0.11122	0.10429	0.28766	0.24038	0.15908	0.10085	0.08672		
30	0.11493	0.11493	0.10971	0.10575	0.10082	0.19048	0.16849	0.13585	0.10009	0.07799		
50	0.10327	0.10327	0.10069	0.09862	0.09538	0.14414	0.13436	0.11584	0.09271	0.07352		
100	0.09581	0.09581	0.09480	0.09366	0.09155	0.10837	0.10414	0.09332	0.07984	0.06274		
200	0.09486	0.09486	0.09437	0.09388	0.09286	0.09081	0.08893	0.08307	0.07564	0.06429		
$\alpha = 0.05$												
15	0.08495	0.08493	0.07634	0.06965	0.06365	0.16510	0.13217	0.09373	0.05730	0.05045		
30	0.05753	0.05753	0.05414	0.05135	0.04750	0.09580	0.08403	0.06722	0.04612	0.03694		
50	0.05181	0.05181	0.05015	0.04863	0.04562	0.07307	0.06671	0.05514	0.04236	0.03257		
100	0.04206	0.04206	0.04137	0.04078	0.03913	0.04502	0.04278	0.03728	0.03049	0.02424		
200	0.04237	0.04237	0.04208	0.04174	0.04114	0.03507	0.03388	0.03086	0.02759	0.02278		
$\alpha = 0.01$												
15	0.04003	0.03941	0.03388	0.03043	0.02638	0.06487	0.05027	0.03512	0.01971	0.01743		
30	0.01680	0.01680	0.01543	0.01417	0.01270	0.02667	0.02270	0.01740	0.01095	0.00877		
50	0.01280	0.01280	0.01222	0.01177	0.01075	0.01722	0.01539	0.01267	0.00949	0.00775		
100	0.00740	0.00740	0.00722	0.00699	0.00667	0.00766	0.00717	0.00607	0.00462	0.00353		
200	0.00700	0.00700	0.00686	0.00679	0.00655	0.00470	0.00442	0.00397	0.00332	0.00277		

Notes: See Table 1.

Table 5: Size of the  $F$  Statistic Using the Feasible Imhof Distribution with  $t(5)$  Distribution of Errors

Heteroskedastic $t(5)$ Error Terms ( $\sigma^2(x) = \exp(x_i^2)$ )												
Using Feasible Imhof Distribution						Using $F$ -table						
$n$	HC0	HC1	HC2	HC3	HC4	HC0	HC1	HC2	HC3	HC4		
$\alpha = 0.10$												
15	0.13923	0.13923	0.12863	0.12059	0.11487	0.26538	0.22373	0.16426	0.10631	0.09741		
30	0.11061	0.11061	0.10607	0.10276	0.09835	0.17723	0.15932	0.13044	0.09637	0.07869		
50	0.09452	0.09452	0.09210	0.09004	0.08667	0.13376	0.12412	0.10529	0.08200	0.06502		
100	0.09164	0.09164	0.09058	0.08972	0.08763	0.10280	0.09832	0.08759	0.07417	0.05862		
200	0.09644	0.09644	0.09605	0.09570	0.09461	0.09294	0.09077	0.08430	0.07649	0.06536		
$\alpha = 0.05$												
15	0.08355	0.08352	0.07307	0.06587	0.05950	0.17856	0.14588	0.09407	0.05443	0.04664		
30	0.05789	0.05789	0.05442	0.05173	0.04782	0.09719	0.08582	0.06845	0.04633	0.03683		
50	0.04873	0.04873	0.04698	0.04559	0.04263	0.06820	0.06266	0.05232	0.03975	0.03112		
100	0.04641	0.04641	0.04568	0.04495	0.04345	0.04928	0.04691	0.04112	0.03380	0.02567		
200	0.04291	0.04291	0.04244	0.04219	0.04136	0.03460	0.03339	0.03038	0.02700	0.02181		
$\alpha = 0.01$												
15	0.03466	0.03454	0.03050	0.02736	0.02434	0.05301	0.04063	0.02979	0.01637	0.01554		
30	0.01810	0.01810	0.01701	0.01580	0.01406	0.02905	0.02433	0.01882	0.01198	0.00971		
50	0.01155	0.01155	0.01088	0.01032	0.00936	0.01591	0.01420	0.01158	0.00824	0.00639		
100	0.00754	0.00754	0.00739	0.00719	0.00665	0.00795	0.00726	0.00590	0.00453	0.00351		
200	0.00693	0.00693	0.00686	0.00677	0.00657	0.00489	0.00465	0.00420	0.00362	0.00298		

Notes: See Table 1.

Table 6: Empirical Application: Rejection Probabilities with Different HC Estimators

	HC0	HC1	HC2	HC3	HC4
$H_0 : \beta_1 = 0$					
<i>F</i> -Statistic	45.14748	42.06925	40.18185	35.11130	28.61864
Imhof <i>p</i> -value	0.00000	0.00000	0.00000	0.00000	0.00000
<i>F</i> -table <i>p</i> -value	0.00000	0.00000	0.00000	0.00000	0.00000
Wild bootstrap <i>p</i> -value (Mammen)	0.00020	0.00020	0.00020	0.00040	0.00100
Wild bootstrap <i>p</i> -value (Rademacher)	0.00000	0.00000	0.00000	0.00000	0.00000
$H_0 : \beta_2 = 0$					
<i>F</i> -Statistic	1.50145	1.39908	1.26527	1.05235	0.79717
Imhof <i>p</i> -value	0.31279	0.31279	0.32210	0.32686	0.31671
<i>F</i> -table <i>p</i> -value	0.22396	0.24030	0.26394	0.30798	0.37455
Wild bootstrap <i>p</i> -value (Mammen)	0.34153	0.34153	0.33653	0.33623	0.32983
Wild bootstrap <i>p</i> -value (Rademacher)	0.35874	0.35874	0.35264	0.34823	0.34133
$H_0 : \beta_3 = 0$					
<i>F</i> -Statistic	0.07409	0.06904	0.05382	0.03754	0.01686
Imhof <i>p</i> -value	0.82260	0.82260	0.83461	0.84445	0.86335
<i>F</i> -table <i>p</i> -value	0.78615	0.79340	0.81713	0.84684	0.89700
Wild bootstrap <i>p</i> -value (Mammen)	0.81308	0.81308	0.80928	0.80498	0.79718
Wild bootstrap <i>p</i> -value (Rademacher)	0.81768	0.81768	0.80868	0.80008	0.78628
$H_0 : \beta_4 = 0$					
<i>F</i> -Statistic	0.49756	0.46364	0.44222	0.38857	0.33664
Imhof <i>p</i> -value	0.52759	0.52759	0.53208	0.53444	0.51954
<i>F</i> -table <i>p</i> -value	0.48257	0.49785	0.50792	0.53478	0.56336
Wild bootstrap <i>p</i> -value (Mammen)	0.53135	0.53135	0.52975	0.52855	0.52545
Wild bootstrap <i>p</i> -value (Rademacher)	0.51805	0.51805	0.51535	0.51265	0.50585
$H_0 : \beta_5 = 0$					
<i>F</i> -Statistic	1.50869	1.40582	1.40069	1.29511	1.29898
Imhof <i>p</i> -value	0.25141	0.25141	0.25179	0.25059	0.23238
<i>F</i> -table <i>p</i> -value	0.22285	0.23918	0.24003	0.25842	0.25772
Wild bootstrap <i>p</i> -value (Mammen)	0.25493	0.25493	0.25543	0.25463	0.25453
Wild bootstrap <i>p</i> -value (Rademacher)	0.25733	0.25733	0.25733	0.25663	0.25683
$H_0 : \beta_2 + \beta_5 = 0$					
<i>F</i> -Statistic	3.35403	3.12534	3.04230	2.73146	2.44761
Imhof <i>p</i> -value	0.08505	0.08505	0.08654	0.08587	0.07170
<i>F</i> -table <i>p</i> -value	0.07067	0.08080	0.08487	0.10221	0.12156
Wild bootstrap <i>p</i> -value (Mammen)	0.08621	0.08621	0.08201	0.07991	0.07921
Wild bootstrap <i>p</i> -value (Rademacher)	0.08921	0.08921	0.08711	0.08621	0.08561

Notes: Reported are the *F* statistics testing for each of the six different single restrictions. Four different *p*-values of each *F* statistic are computed from using the Imhof distribution, the *F*-table, and two different ways of the wild bootstrap methods.