

# The Exact Distribution of the White t-Ratio

Bruce E. Hansen\*

University of Wisconsin†

June, 2021

## Abstract

This paper presents new expressions for the exact finite sample distribution of the White (1980) heteroskedasticity-robust t-ratio under the assumption of normal heteroskedastic errors. The first expression shows that the distribution function equals the expectation of a nonlinear function of a weighted sum of chi-square random variables, with the weights an explicit function of the regressor matrix and error variances. The second expression shows that the distribution function equals a mixture of student t distribution functions. These are the first expressions for the exact distribution of the White t-ratio allowing for heteroskedastic error variances, other than expressions based on the numerical inversion of the characteristic function.

Our exact distribution function is inconvenient to evaluate in practice, so we recommend a simple approximation with excellent computational and approximation properties. The motivation is the first result described above that the distribution function is completely determined by a specific weighted sum of chi-squares. Using results from the recent literature on approximation of the distribution function of weighted sums of chi-squares, we obtain a practical approximation to the distribution function of the White t-ratio which is computationally fast in small to moderate samples and is exceedingly accurate.

---

\*Research support from the NSF and the Phipps Chair are gratefully acknowledged. My thanks to Grant Hillier for helpful discussions and for sharing preliminary drafts of his related research.

†Department of Economics, 1180 Observatory Drive, University of Wisconsin, Madison WI 53706.

# 1 Introduction

The most important discovery in the history of econometrics may be Gosset's derivation of the exact distribution of the studentized sample mean under normal sampling (Gosset, 1908). This result is elegant. It is also practical, as it readily extends to the homoskedastic regression model and provides a simple computational method to evaluate the sampling distribution. Gosset's result provides the foundation for most finite-sample inference in practical applied econometrics – the comparison of empirical t-ratios with the student t distribution function.

Unfortunately, the distributional result (that t-ratios are distributed student t) is incorrect when applied to modern “heteroskedasticity-robust” and “cluster-robust” t-ratios. This is because “robust” t-ratios do not have student t distributions in finite samples. Regardless, the student t distribution is still routinely applied to these t-ratios, based on the powerful heuristic of Gosset's result, conventional software implementations, and ingrained statistical thinking. The problem, fundamentally, is that there is no good alternative, as there is no known computable form of the exact distribution function. While it is possible to simulate the exact distribution and/or calculate it numerically by inverting the characteristic function, such techniques are not practical defaults due to computation cost. The lack of an explicit computable exact distribution has preclude routine use of exact distribution methods.

This paper, at least partially, solves this impasse. We present new expressions for the exact finite sample distribution of the White (1980) heteroskedasticity-robust t-ratio under the assumption of normal heteroskedastic errors. First, we show that the exact distribution equals the expectation of a nonlinear function of a specific weighted sum of independent chi-square random variables, where the weights are known computable functions of the regressor matrix and error variances. Second, we show that the exact distribution equals a weighted average of student t distribution functions, where the weights are explicit functions of the regressor matrix and error variances.

For practical application we recommend approximating the exact distribution function by a simpler version obtained by recent developments in the literature on approximating the distribution of weighted sums of chi-squares. Our approximations are near-exact, resulting in no loss of practical applicability, and are computationally fast, with p-value calculation taking less than 0.02 seconds on a personal computer for sample sizes under 500. For larger sample sizes, however, computation may require program-

ming improvements.

This exact distribution function just described depends on the unknown variances, so is infeasible, but replacement of the unknown variances by estimates provides a simple and reasonably accurate estimate of the distribution function.

The heteroskedasticity-robust covariance matrix estimator known as HC0 was introduced by Eicker (1963) and subsequently popularized in econometrics by White (1980). Alternative versions known as HC1, HC2, and HC3 were introduced by Hinkley (1977), Horn, Horn, and Duncan (1975), and MacKinnon and White (1985). All are programmed in packages such as Stata. The HC1 version is obtained by the popular “, r” covariance matrix option, and is thereby the de-facto default in contemporary applied econometrics.

The recognition that the finite-sample distribution of the White t-ratio can be severely distorted from the student t distribution has been long recognized. Early investigations include MacKinnon and White (1985), Chesher and Jewitt (1987), Chesher (1989), and Phillips (1993). Chesher and Austin (1991) proposed calculation of the exact distribution of the White t-ratio based on inversion of the characteristic function as in Imhof (1961); this proposal has been expanded recently by Chu, Lee, Ullah, and Xu (2021) to regression F statistics. While promising, the Imhof-based method has not gained traction in applied econometrics due to computation costs.

Our paper is closely related to the work of Bell and McCaffrey (2002), Imbens and Kolesár (2016), and Hillier and O’Brien (2019). Bell and McCaffrey (2002) and Imbens and Kolesár (2016) proposed approximating the distribution of the White t-ratio with a student t distribution whose scale and degree of freedom parameter are selected to match the first two moments of the distribution of the variance estimator under the assumption of homoskedastic normal errors. This “two-parameter” method works remarkably well in many applications, but has a few deficiencies. First, the approximation assumes homoskedastic errors, and is sensitive to its violation. Second, the approximation is imprecise in many cases because it is insufficiently flexible. Third, these authors did not explain the sense in which the two-parameter method is an approximation to the exact distribution. Hillier and O’Brien (2019) extend this approach<sup>1</sup> by suggesting a “four-parameter” approximation to the distribution of the variance estimator which is a two-component weighted sum of chi-squares. They clarify the role of the approximation, showing that the degree of approximation is controlled by the number of chi-

---

<sup>1</sup>Based in part on a preliminary draft of the present paper.

square components. As in Bell-McCaffrey and Imbens-Kolesár, however, they maintain the assumption of homoskedastic error variances.

Another closely related paper is Pötscher and Preinerstorfer (2021) who demonstrate that there exists a lower bound for the critical level of the White t test which uniformly controls its size across all possible error variance combinations. They show by example that this bound can be exceedingly high; their Table 1 indicates that the 5% bound for the HC1 t-ratio in one example is 1711, meaning that t-ratios would need to exceed 1711 in order to be deemed “significant”. Empirical implementation of such critical values does not appear practical.

We extend this literature by introducing “three-parameter” and “four-parameter” approximations to the *heteroskedastic* normal regression model. The extension to heteroskedastic variances is important, as the goal of heteroskedasticity-robust t-ratios is to allow for unknown heteroskedasticity. We demonstrate that our four-parameter approximation is near-perfect, even in highly leveraged regression designs. The three-parameter approach is nearly as accurate, and has the advantage that it is computationally robust to extremely leveraged designs. Consequently we recommend a hybrid implementation which makes use of both the three-parameter and four-parameter approximations.

The organization of the paper is as follows. Section 2 introduces the heteroskedastic normal regression model, the HC1 covariance matrix estimator, and the White t-ratio. Section 3 introduces the numerical examples which are used throughout the paper to illustrate the proposed methods. Simulation results show that the finite sample size of conventional inference methods for the White t-ratio – based on the student t, wild bootstrap, and Bell-McCaffrey distributions – can be highly distorted. Section 4 presents the main theoretical result of the paper – the exact finite sample distribution of the White t-ratio. Section 5 proposes approximations. Section 6 discusses computational implementation. Section 7 discusses feasible implementation using estimated error variances. Section 8 presents an extension of the methods to the HC0, HC2, and HC3 covariance matrix estimators. Appendix A presents the formula used to match the moments of the approximating models. Appendix B contains the mathematical proofs of the three theorems in the paper.

All numerical results presented in the paper were computed in Matlab. The code is posted on the author’s website. The paper contains a subset of the results focusing on the HC1 covariance matrix estimator. A Supplemental Appendix contains the full set

of numerical results, including the HC2 and HC3 covariance matrix estimators and an extensive set of figures.

## 2 Model

Take the heteroskedastic normal regression model

$$Y = X'\beta + e \quad (1)$$

$$e | X \sim N(0, \sigma^2(X)) \quad (2)$$

where  $X$  and  $\beta$  are  $k \times 1$ . Let  $(Y_i, X_i)$ ,  $i = 1, \dots, n$  be a random sample of observations from  $(Y, X)$ . Write the observations in standard matrix notation as  $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$ . Assume that

$$\mathbf{X}'\mathbf{X} > 0. \quad (3)$$

The standard estimator for  $\beta$  is least squares  $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y})$  with residuals  $\hat{e}_i = Y_i - X_i'\hat{\beta}$ . Covariance matrix estimation is most commonly performed with the HC1 estimator:

$$\hat{\mathbf{V}}_{\text{HC1}} = \left(\frac{n}{n-k}\right)(\mathbf{X}'\mathbf{X})^{-1} \left(\sum_{i=1}^n X_i X_i' \hat{e}_i^2\right) (\mathbf{X}'\mathbf{X})^{-1}.$$

HC1, first proposed by Hinkley (1977), is a degree-of-freedom-adjusted version of the HC0 covariance matrix estimator popularized in econometrics by White (1980). HC1 has become the de-facto default in applied econometrics as it is the method implemented by the “, r” covariance matrix option in Stata. We will focus on the HC1 estimator in this manuscript. All results extend to a variety of other estimators, including HC0, HC2 and HC3, as described in Section 8. The numerical analysis presented in the paper is for the HC1 estimator; analogous analysis for the HC2 and HC3 estimators are presented in the Supplemental Appendix.

We are interested in a scalar parameter  $R'\beta$  for some non-zero  $k \times 1$  vector  $R$ . This includes individual coefficients and linear combinations. Its estimator is  $R'\hat{\beta}$ , variance estimator  $R'\hat{\mathbf{V}}R$ , and t-ratio

$$T = \frac{R'(\hat{\beta} - \beta)}{\sqrt{R'\hat{\mathbf{V}}_{\text{HC1}}R}}.$$

This is the most commonly reported t-ratio in applied econometric practice. We call  $T$

the “White t-ratio”, or the “HC1 t-ratio” when we want to be specific about the covariance matrix estimation method.

While the asymptotic distribution of  $T$  is standard normal under broad conditions, it is conventional to calculate confidence intervals and p-values using the student t distribution with  $n - k$  degrees of freedom. This approximation is heuristically motivated by the fact that the classical (non-homoskedastic) t-ratio has the  $t_{n-k}$  distribution. This heuristic motivation, however, is far from exact in finite samples.

### 3 Numerical Illustration

To illustrate the magnitude of the distortions we calculate the actual size of nominal 5% tests using student t critical values. The model is the simple normal regression

$$Y = \alpha + X\beta + e$$

$$e | X \sim N(0, \sigma^2(X))$$

with  $X \in \mathbb{R}$  fixed. We consider five regressor specifications. The first is the dummy variable specification  $X = \{2, 2, 2, 1, 1, \dots, 1\}$ . This is a treatment setting with 3 treated observations and the remaining observations untreated. This is the model examined, for example, in Angrist and Pischke (2009), Imbens and Kolesar (2016), and others. The other four models are based on probability distributions, spacing out the regressor values evenly on the quantiles<sup>2</sup> of a specified distribution. The four probability models are Pareto(2), Gamma(1/4, 1), logNormal(0, 1) and logNormal(0, 4). Each model generates outlier-like regressor patterns, and have similar distributions to many economic variables. These models are commonly used in econometric studies of the impact of high leverage on heteroskedastic covariance matrix estimation.

We consider two specifications for the error variance: (1) Homoskedastic:  $\sigma^2(X) = 1$ ; and (2) Heteroskedastic:  $\sigma^2(X) = 1 + X^2$ . We consider four sample sizes:  $n = \{30, 60, 120, 500\}$ .

We construct the t-ratio  $T$  for the coefficient  $\beta$  and calculate the actual size of a two-sided test based on the  $t_{n-2}$  critical value using 20,000 simulation replications. The results are reported in Table 1.

The results are striking, but will not be surprising to specialists. Under both homoskedasticity and heteroskedasticity, for all five regressor specifications, and regard-

---

<sup>2</sup>Specifically, on the quantiles  $j/(n+1)$  for  $j = 1, \dots, n$ .

Table 1: Actual size using 5%  $t_{n-k}$  critical values

	Homoskedastic Error				Heteroskedastic Error			
	$n = 30$	$n = 60$	$n = 120$	$n = 500$	$n = 30$	$n = 60$	$n = 120$	$n = 500$
Dummy	0.17	0.20	0.22	0.25	0.19	0.22	0.23	0.25
Pareto(2)	0.13	0.12	0.11	0.08	0.24	0.21	0.19	0.15
Gamma(1/4,1)	0.13	0.10	0.08	0.06	0.21	0.17	0.12	0.08
logNormal(0,1)	0.10	0.09	0.08	0.06	0.20	0.17	0.13	0.09
logNormal(0,4)	0.20	0.19	0.16	0.12	0.49	0.41	0.34	0.24

less of the sample size, the tests are over-sized. The rejection rates are particularly over-sized under heteroskedasticity, with the rejection rates reaching as high as 49%. The rejection rates improve as  $n$  increases for most specifications (though not the dummy variable specification) but remain unacceptably high even for  $n = 500$ . This table illustrates that the student t distribution is not a good approximation to the finite sample distribution of the White t-ratio  $T$ .

These distortions are well-known and alternative inference procedures have been proposed. The most popular is the bootstrap, for which there are several implementations. For the regression model a popular method is the wild bootstrap, which treats the regressors as fixed and resamples the errors from  $e_i^* = \hat{e}_i \xi_i$  where  $\xi_i \sim (0, 1)$  is an auxiliary random variable. Following Davidson and Flachaire (2008) we use the Rademacher<sup>3</sup> distribution. For each simulation replication we generate 1000 bootstrap samples to calculate the bootstrap p-value of the two-sided test based on the t-ratio  $T$ . The percentage of bootstrap p-values less than the nominal level of 0.05 is the simulation estimate of the actual size of the bootstrap test. The results are reported in Table 2. We see that the performance of the bootstrap is actually worse than the t-distribution in some designs. In general, the bootstrap tests are severely over-sized, with rejection rates as high as 35%.

An alternative proposal was made by Bell and McCaffrey (2002) and endorsed by Imbens and Kolesar (2016). They suggest basing inference on a scaled student t distribution  $at_\eta$ , where the scale parameter  $a$  and degree of freedom parameter  $\eta$  are selected based on a two-moment chi-square approximation to the variance estimator, the approximation calculated under the assumption of homoskedastic normality. We implemented their procedure and calculated by simulation the actual size of the test. The results are reported in Table 3.

---

<sup>3</sup> $\xi = \pm 1$  each with probability 0.5.

Table 2: Actual size using 5% wild bootstrap critical values

	Homoskedastic Error				Heteroskedastic Error			
	$n = 30$	$n = 60$	$n = 120$	$n = 500$	$n = 30$	$n = 60$	$n = 120$	$n = 500$
Dummy	0.10	0.10	0.11	0.10	0.10	0.10	0.10	0.19
Pareto(2)	0.15	0.12	0.10	0.08	0.19	0.14	0.11	0.07
Gamma(1/4,1)	0.14	0.10	0.07	0.05	0.17	0.11	0.07	0.05
logNormal(0,1)	0.11	0.09	0.07	0.05	0.14	0.09	0.07	0.05
logNormal(0,4)	0.24	0.21	0.17	0.11	0.35	0.25	0.19	0.11

When the errors are homoskedastic the Bell-McCaffrey rejection rates are uniformly below the nominal level of 5%. Instead of over-rejection there is under-rejection, with rejection rates as low as 2%. In the heteroskedastic design, however, the Bell-McCaffrey rejection rates are over-sized, with rejection rates as high as 18%.

From this simple investigation we can see that the finite sample distribution of the White t-ratio is not well approximated by the student t distribution, the wild bootstrap distribution, nor the Bell-McCaffrey distribution.

Table 3: Actual size using 5% Bell-McCaffrey critical values

	Homoskedastic Error				Heteroskedastic Error			
	$n = 30$	$n = 60$	$n = 120$	$n = 500$	$n = 30$	$n = 60$	$n = 120$	$n = 500$
Dummy	0.03	0.03	0.04	0.05	0.04	0.05	0.05	0.05
Pareto(2)	0.03	0.04	0.04	0.04	0.10	0.10	0.09	0.09
Gamma(1/4,1)	0.03	0.04	0.04	0.05	0.09	0.08	0.07	0.06
logNormal(0,1)	0.04	0.04	0.04	0.05	0.11	0.09	0.08	0.07
logNormal(0,4)	0.02	0.03	0.03	0.04	0.18	0.15	0.13	0.11

One limitation of Tables 1-3 is that their focus on 5% size leaves open the possibility that the distributional approximations may differ at other quantiles. To investigate this we display in Figure 1 plots of the exact CDF of the t-ratio  $T$ , calculated from one million simulation draws, along with the student t and Bell-McCaffrey distributions<sup>4</sup>, for the logNormal(0,4) regressor design with  $n = 30$  observations. Panel (a) is for case of homoskedastic errors and panel (b) for the case of heteroskedastic errors. Panel (a) is displayed for  $x \in [0, 4]$  and panel (b) for  $x \in [0, 8]$ . We can see that in these plots, the distributions are strictly ranked for all values of  $x$ , so that the sign of the distortions from Tables 1-3 are invariant to the nominal significance level. In these plots we can see

<sup>4</sup>We do not display the bootstrap distribution as it is random and depends on the realized values of  $Y$ .



that the student t distribution lies above the exact distribution and thus produces over-sized tests, while the Bell-McCaffrey distribution lies under the exact distribution in the homoskedastic case, and the reverse in the heteroskedastic case. The Bell-McCaffrey p-value therefore produces under-sized and over-sized tests in the homoskedastic and heteroskedastic designs, respectively. We constructed similar plots for the other sample sizes and designs, and are displayed in the Supplemental Appendix. The qualitative nature of Figure 1 generally holds for all cases, with the exception that with the HC3 covariance matrix estimator the exact and Bell-McCaffrey distribution functions cross in some cases.

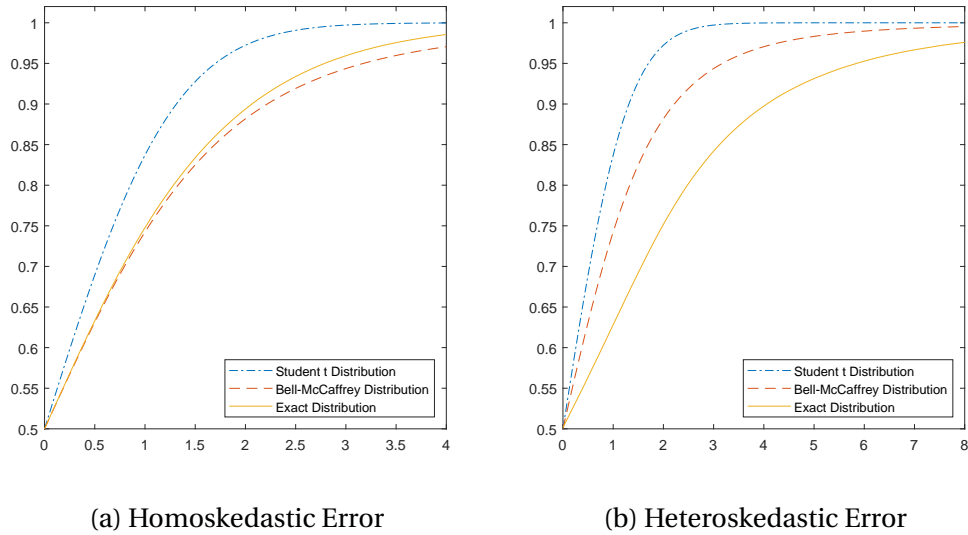


Figure 1: Student t, Bell-McCaffrey, and Exact Distribution of t-ratio.  $\log\text{Normal}(0,4)$  Regressor Design,  $n = 30$

## 4 Exact Distribution

In this section we derive the exact finite sample distribution of the White t-ratio  $T$ , conditional on  $\mathbf{X}$ :

$$G(x | \mathbf{X}) = \mathbb{P}[T \leq x | \mathbf{X}].$$

Our analysis is for a fixed value of  $x$ .

Define the annihilator matrix  $\mathbf{M} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ , the scalar transformed regres-

for  $Z_i = R'(\mathbf{X}'\mathbf{X})^{-1}X_i$ , its stacked vector  $\mathbf{Z}$ , the matrix

$$\mathbf{B}_1 = \left(\frac{n}{n-k}\right) \text{diag}(Z_1^2, \dots, Z_n^2), \quad (4)$$

the covariance matrix

$$\Sigma = \text{diag}(\sigma^2(X_1), \dots, \sigma^2(X_n)),$$

and the matrix

$$\mathbf{C} = \Sigma^{1/2} \mathbf{Z} \mathbf{Z}' \Sigma^{1/2} - x^2 \Sigma^{1/2} \mathbf{M} \mathbf{B}_1 \mathbf{M} \Sigma^{1/2}. \quad (5)$$

Set  $N = n - k$ . Let  $\lambda_0 \geq \lambda_1 \geq \dots \geq \lambda_N$  be the non-zero eigenvalues of  $\mathbf{C}$ . These satisfy  $\lambda_0 > 0$  and  $\lambda_j < 0$  for  $j = 1, \dots, N$ . Set  $w_j = -\lambda_j / \lambda_0$  for  $j = 1, \dots, N$ . They satisfy  $0 < w_1 \leq \dots \leq w_N$ .

**Theorem 1** *In model (1)-(3) the conditional distribution of  $T$  given  $\mathbf{X}$  equals*

$$G(x | \mathbf{X}) = \mathbb{E} \left[ \Phi \left( \sqrt{Q(w)} \right) \mid \mathbf{X} \right] \quad (6)$$

where

$$Q(w) = \sum_{j=1}^N w_j Q_j, \quad (7)$$

and  $Q_j \sim \chi_1^2$  are mutually independent. The cumulative distribution function (6) equals

$$G(x | \mathbf{X}) = \sum_{m=0}^{\infty} c_m F_{N+2m} \left( \sqrt{w_1 (N+2m)} \right) \quad (8)$$

where the coefficients  $c_m$  equal

$$c_0 = \prod_{j=1}^N \left( \frac{w_1}{w_j} \right)^{1/2} \quad (9)$$

$$c_m = \frac{1}{m} \sum_{j=0}^{m-1} c_j d_{m-j} \quad (10)$$

$$d_m = \frac{1}{2} \sum_{j=1}^N \left( 1 - \frac{w_1}{w_j} \right)^m, \quad (11)$$

and  $F_r(x)$  is the student  $t$  cumulative distribution function with  $r$  degrees of freedom.

Theorem 1 provides two new and useful expressions. Expression (6) shows that the

distribution function  $G(x)$  of the White t-ratio  $T$  equals the expectation of the nonlinear function  $\Phi(\sqrt{x})$  of a weighted sum of chi-square random variables with weights  $w_j$ . The distribution  $G(x)$  simplifies to the conventional student t distribution when the weights are all equal:  $w_j = 1/N$ . Otherwise, when the weights  $w_j$  are heterogeneous the distribution deviates from the student t distribution.

The second expression is (8). It shows that the distribution function  $G(x)$  of the White t-ratio  $T$  equals a weighted average of student t distribution functions. The series (8) is convergent, and is thus a valid representation for  $G(x)$ .

Expression (8) is based on an representation for the distribution of a weighted sum of chi-square variables due to Ruben (1962) and Farebrother (1984). Other representations exist, including that of Davies (1980). However, the Farebrother algorithm is viewed as computationally more efficient. Furthermore, it is unclear how the Davies algorithm, for example, could be used in conjunction with the method of Theorem 1 to obtain a useful expression for the distribution function.

## 5 Approximate Distribution

While (8) provides an explicit formula for computation of the distribution  $G(x)$ , it is not recommended for empirical practice. The formula is sensitive to the magnitude of the weight ratio  $w_1/w_K$  which can be arbitrarily small in some empirical contexts. Instead, we recommend an approximation obtained by replacing the weighted sum of chi-squares  $Q(\lambda)$  in (7) with a smaller number of components. There is a substantial statistical literature developing approximations to the distribution of a weighted sum of chi-squares; for a recent review and evaluation see Bodenham and Adams (2015). We focus on approximations to  $Q(\lambda)$  which are another weighted sum of chi-squares with a smaller number of components. These turn out to provide excellent approximations with greatly improved computational properties.

We consider three approximations. The simplest is the one-component model due to Welch (1938) and Satterthwaite (1946):

$$Q(w) \simeq Q_2 = a\chi_\eta^2$$

where we use the symbol “ $\simeq$ ” to denote “approximately distributed as”. This use of  $Q_2$  to approximate  $Q(w)$  is the basis of the Bell-McCaffrey distribution recommended by

Imbens and Kolesar (2016).

Our second approximation is the shifted one-component model of Hall (1983) and Buckley and Eagleson (1988):

$$Q(w) \simeq Q_3 = a + b\chi_\eta^2.$$

Bodenham and Adams (2015) provide substantial numerical evidence that  $Q_3$  is a superior approximation to  $Q(w)$  than  $Q_2$ .

Our third approximation is the two-component model of Hillier and O'Brien (2019):

$$Q(w) \simeq Q_4 = a_1\chi_{\eta_1}^2 + a_2\chi_{\eta_2}^2.$$

Hillier and O'Brien provide detailed numerical evidence that  $Q_4$  is a highly accurate approximation to  $Q(w)$  in a wide range of designs.

Let  $Q_r$  denote  $Q_2$ ,  $Q_3$ , or  $Q_4$ . Replacing  $Q(w)$  in (6) with  $Q_r$  we obtain the distributional approximation

$$G(x | \mathbf{X}) \simeq G_r(x | \mathbf{X}) = \mathbb{E} \left[ \Phi \left( \sqrt{Q_r} \right) \middle| \mathbf{X} \right].$$

We call  $G_r$  the “ $r$ -parameter approximation” to  $G$ . As  $r$  increases the approximation accuracy increases, with the approximation error eliminated for  $r = N$ .

The  $G_2$  distribution in the homoskedastic case equals the Bell-McCaffrey distribution, which we observed in Table 3 is undersized. Seeking an improved approximation we do not pursue  $G_2$  further.

The three-parameter approximation  $G_3$  can be written as the integral

$$G_3(x | \mathbf{X}) = \int_0^\infty \Phi \left( \sqrt{a + bt} \right) f_\eta(t) dt \tag{12}$$

where  $f_\eta(t)$  is the  $\chi_\eta^2$  density function. An algebraic solution to this integral is not available, but numerical evaluation is straightforward and computationally fast.

The distribution  $G_3$  is nested within  $G_4$ . To see this, take the limit of  $Q_4$  as  $\eta_1 \rightarrow \infty$  with  $a_1\eta_1 \rightarrow a$ . We find  $Q_4 = a_1\chi_{\eta_1}^2 + a_2\chi_{\eta_2}^2 \xrightarrow{p} a + a_2\chi_{\eta_2}^2 = Q_3$ . Thus  $G_3$  is a limiting case of  $G_4$ .

The parameters of  $G_3$  can be selected by matching the first three moments of  $Q_3$  to those of  $Q(w)$ , as recommended by Hall (1983) and Buckley and Eagleson (1988). Simi-

Table 4: Actual size using 5% critical values from infeasible three-parameter and four-parameter distributions

	$n = 30$	$n = 60$	$n = 120$	$n = 250$	$n = 500$
	Three-Parameter Distribution, Homoskedastic Case				
Dummy	0.05	0.05	0.05	0.05	0.05
Pareto(2)	0.05	0.05	0.05	0.05	0.05
Gamma(1/4,1)	0.05	0.05	0.05	0.05	0.05
logNormal(0,1)	0.05	0.05	0.05	0.05	0.05
logNormal(0,4)	0.05	0.06	0.05	0.05	0.05
	Four-Parameter Distribution, Homoskedastic Case				
Dummy	0.05	0.05	0.05	0.05	0.05
Pareto(2)	0.05	0.05	0.05	0.05	0.05
Gamma(1/4,1)	0.05	0.05	0.05	0.05	0.05
logNormal(0,1)	0.05	0.05	0.05	0.05	0.05
logNormal(0,4)	0.05	0.05	0.05	0.05	0.05
	Three-Parameter Distribution, Heteroskedastic Case				
Dummy	0.05	0.05	0.05	0.05	0.05
Pareto(2)	0.06	0.06	0.06	0.06	0.06
Gamma(1/4,1)	0.06	0.06	0.06	0.05	0.05
logNormal(0,1)	0.06	0.06	0.06	0.05	0.06
logNormal(0,4)	0.07	0.07	0.07	0.06	0.06
	Four-Parameter Distribution, Heteroskedastic Case				
Dummy	0.05	0.05	0.05	0.05	0.05
Pareto(2)	0.05	0.05	0.05	0.05	0.05
Gamma(1/4,1)	0.05	0.05	0.05	0.05	0.05
logNormal(0,1)	0.05	0.05	0.05	0.05	0.05
logNormal(0,4)	0.05	0.04	0.05	0.04	0.05

larly, the parameters of  $G_4$  can be selected by matching the first four moments of  $Q_4$  to those of  $Q(w)$ , as recommended by Hillier and O'Brien (2019). Details are provided in Appendix A.

The four-parameter approximation  $G_4$  has an explicit solution which we now present.

**Theorem 2** *The exact four-moment cumulative distribution function is*

$$G_4(x | \mathbf{X}) = \sum_{m=0}^{\infty} b_m F_{\eta_1 + \eta_2 + 2m} \left( \sqrt{a_1 (\eta_1 + \eta_2 + 2m)} \right) \quad (13)$$

where the coefficients  $b_m$  equal

$$b_m = \left(\frac{a_1}{a_2}\right)^{\eta_2/2} \left(1 - \frac{a_1}{a_2}\right)^m \frac{\Gamma\left(m + \frac{\eta_2}{2}\right)}{\Gamma\left(\frac{\eta_2}{2}\right) m!} \quad (14)$$

and  $F_r(x)$  is the student  $t$  cumulative distribution function with  $r$  degrees of freedom.

In contrast to the recursion (10) for  $c_m$ , (14) provides an explicit formula for the coefficients  $b_m$ . Consequently, they are computationally quick to calculate and the distribution function (13) is numerically fast to evaluate in most cases. Computational issues are discussed in the next section.

It may be useful to observe that the coefficients  $b_m$  equal the probability density function of the Negative Binomial distribution with parameters  $\eta_2/2$  and  $a_1/a_2$ . One implication is  $\sum_{m=0}^{\infty} b_m = 1$ . Another is that (13) is convergent, and thus a valid representation for  $G_4(x)$ .

We investigate by simulation the accuracy of the approximations  $G_3$  and  $G_4$  to the exact distribution. We used the same design as in Tables 1-3 with the addition of  $n = 250$ . The results are reported in Table 4. For the homoskedastic designs the three and four-parameter approximations  $G_3$  and  $G_4$  are excellent, with size equal to 5% for  $G_3$  for all but one case, and for  $G_4$  for every case. For heteroskedastic designs the performance is more varied. We can see that the three-parameter distribution  $G_3$  has 5%-6% size for all designs except logNormal(0,4), where size equals 7% for  $n \leq 120$ . The four-parameter distribution  $G_4$  has excellent performance, with 5% size for all but two cases and no over-rejections.

To illustrate the approximations see Figure 2. Here we plot the exact distribution of the  $t$ -ratio, along with the approximations  $G_3$  and  $G_4$ , for the logNormal(0,4) regressor design with  $n = 120$ . Panel (a) is the homoskedastic case and panel (b) the heteroskedastic case. We can see in both panels that the three-parameter distribution  $G_3$  lies strictly above (but is close to) the exact distribution, and the four-parameter distribution  $G_4$  lies very close to the exact distribution. This means that  $p$ -values based on  $G_3$  slightly over-reject, and those based on  $G_4$  are essentially correct. Plots for the other regressor designs are qualitatively similar, though in most cases with considerably reduced differences between the distribution functions.

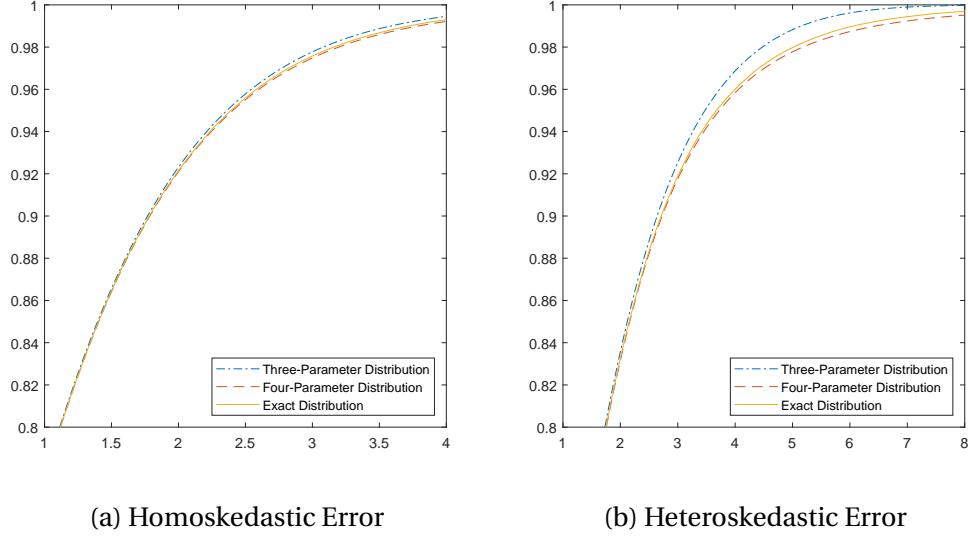


Figure 2: Three-Parameter and Four-Parameter Distributions. logNormal(0,4) Regressor Design,  $n = 120$

## 6 Computation

In practice, the four-moment distribution  $G_4(x)$  defined in (13) is evaluated with a finite number of terms

$$G_4^M(x | \mathbf{X}) = \sum_{m=0}^M b_m F_{\eta_1 + \eta_2 + 2m} \left( \sqrt{a_1 (\eta_1 + \eta_2 + 2m)} \right). \quad (15)$$

The difference is the truncation error  $T_M = |G_4^M(x | \mathbf{X}) - G_4(x | \mathbf{X})|$ . By selecting the number of series terms  $M$  sufficiently large this truncation error can be made arbitrarily small.

**Theorem 3** For any  $\tau > 0$  the truncation error satisfies  $T_m \leq \tau$  if

$$M \geq \frac{1}{2c} \left( q_{\eta_2} \left( 1 - \left( \left( \frac{1}{\delta} - 1 \right) c \right)^{\eta_2/2} \frac{\tau}{1 - \delta} \right) - c\eta_2 \right) \quad (16)$$

where  $\delta = a_1 / a_2$ ,  $c = \log(1/(1 - \delta))$ , and  $q_{\eta}(p)$  is the quantile function of the  $\chi_{\eta}^2$  distribution.

The right-side of (16) is a function of  $a_1$ ,  $a_2$ ,  $\eta_2$ , and  $\tau$ . Thus in any application,  $M$  can be selected so that the truncation error is smaller than a specified tolerance  $\tau$ . In

our numerical work we set  $\tau = 0.0001$ .

Table 5: Computation Cost

	$n = 30$	$n = 60$	$n = 120$	$n = 250$	$n = 500$	$n = 1000$	$n = 2000$
Three-Parameter Distribution, Computation Time (Seconds)							
Dummy	0.001	0.002	0.002	0.002	0.012	0.067	0.352
Pareto(2)	0.001	0.001	0.003	0.003	0.011	0.061	0.356
Gamma(1/4,1)	0.001	0.001	0.002	0.003	0.011	0.060	0.357
logNormal(0,1)	0.001	0.001	0.003	0.004	0.010	0.063	0.362
logNormal(0,4)	0.001	0.003	0.002	0.003	0.012	0.061	0.365
Four-Parameter Distribution, Computation Time (Seconds)							
Dummy	0.001	0.003	0.004	0.010	0.056	0.267	1.067
Pareto(2)	0.001	0.002	0.002	0.003	0.013	0.044	0.249
Gamma(1/4,1)	0.001	0.001	0.001	0.002	0.010	0.045	0.247
logNormal(0,1)	0.001	0.002	0.002	0.002	0.010	0.050	0.246
logNormal(0,4)	0.001	0.002	0.001	0.003	0.012	0.046	0.244
Four-Parameter Distribution, Number of series terms ( $M$ )							
Dummy	740	3319	14,003	62,429	252,776	1,017,415	4,076,592
Pareto(2)	65	65	64	66	71	77	85
Gamma(1/4,1)	63	57	57	61	69	81	98
logNormal(0,1)	55	61	65	71	79	90	104
logNormal(0,4)	90	69	58	55	56	59	63

When  $\eta_1 + \eta_2 + 2m$  is large, the student t distribution function in (15) is close to the standard normal, and this substitution can be made with little added error. As the standard normal distribution is considerably faster to evaluate this can greatly reduce computation time in contexts with large  $M$ . A numerical investigation shows that the student t and normal distributions satisfy the inequality  $\sup_x |F_r(x) - \Phi(x)| \leq \tau$  if  $r \geq 0.158/\tau$ . Thus for a given error tolerance  $\tau$  the normal distribution can be substituted into (15) for  $\eta_1 + \eta_2 + 2m \geq 0.158/\tau$ . In our numerical work we set  $\tau = 0.0001$ .

Our formula for computation of any of the approximate distributions  $G_2$ ,  $G_3$ , or  $G_4$  require the weights  $w_1, \dots, w_K$ , which are based on the eigenvalues of the matrix  $\mathbf{C}$  in (5). When  $n$  is small this is a reasonable calculation but as  $n$  increases this is computationally costly. It is not strictly necessary, however, to actually compute all the eigenvalues. As shown in Appendix A, the coefficients of the approximations are functions only of the averages  $\mu_r = \sum_{j=1}^K w_j^r$ , which can be calculated more efficiently as follows. First, calculate the largest eigenpair  $(\lambda_0, h_0)$  of the matrix  $\mathbf{C}$ . In Matlab, for example, this can be accomplished using the `eigs` command with the `largestreal` option. Set



$\mathbf{C}_1 = (h_0 h_0' \lambda_0 - \mathbf{C}) / \lambda_0$ . This is identical to  $\mathbf{C}$ , but stripped of its largest eigenvalue  $\lambda_0$ , normalized by  $\lambda_0$ , and multiplied by  $-1$ . The non-zero eigenvalues of  $\mathbf{C}_1$  equal the weights  $(w_1, \dots, w_K)$ . The averages  $\mu_r$  are then found as  $\mu_r = \text{tr}(\mathbf{C}_1^r)$ . These matrix operations are computationally more efficient than eigenvalue calculation, especially when the sample size is large.

To illustrate computation cost, in Table 5 we display the computation time to calculate the distribution functions  $G_3(x | \mathbf{X})$  and  $G_4(x | \mathbf{X})$  at  $x = 3$  for the five regressor designs under homoskedasticity and a range of values of  $n$ . Computation time is displayed in seconds, and was calculated using the `timeit` function in Matlab. Computation was done in Matlab R2020b on a personal computer with an i7-4790 3.60GHz processor running Windows 10 Pro. Examine the first panel of Table 5, which displays the computation time for the  $G_3$  distribution. We can see that computation time is roughly invariant across the regressor designs, but is strongly increasing in  $n$  for  $n > 250$ . Computation time is about 0.002 seconds for  $n = 100$ , increases to 0.01 seconds for  $n = 500$ , and to 0.06 seconds for  $n = 1000$ . This computation cost is mostly due to the matrix manipulations needed to calculate the parameters of the approximation; very little is due to the numerical integral (12).

Now examine the second panel of Table 5, which displays the computation time for the  $G_4$  distribution. If we ignore the first row (the Dummy design), we see that the computation cost is essentially identical to the  $G_3$  distribution. This is because most of the computation time is due to the same matrix manipulations. However, the computation time required to compute  $G_4$  for the Dummy design is considerably different from the others. The computation time is small for small  $n$ , but greatly increases as  $n$  increases. For large  $n$  the computation time becomes unreasonable for conventional application.

The reason why the Dummy design has a high computation time for  $G_4$  can be seen from the third panel of Table 5, which displays the number of series terms  $M$  needed to obtain a small approximation error (we used the tolerance  $\tau = 0.0001$ ). We can see that for all designs excepting the Dummy design, only a smaller number (less than 100) of series terms are required, but for the Dummy design the number of series terms is large and increasing in  $n$ . The reason is because in the approximating model  $Q_4 = a_1 \chi_{\eta_1}^2 + a_2 \chi_{\eta_2}^2$  the ratio  $a_1 / a_2$  is extremely small, leading to slow convergence of (13).

This evidence suggests the practical rule: Use the  $G_4$  approximation if the number of required series terms  $M$  from (16) is not too large; otherwise use the  $G_3$  approximation. From Table 5 we see that computation time is less than 0.02 seconds if

$M \leq 100,000$ , which we propose as a practical maximum. Implementation of this rule (use  $G_4$  if  $M \leq 100,000$ , and otherwise use  $G_3$ ) implies in our examples that  $G_4$  will be used for all designs except for the Dummy design with  $n > 250$  under homoskedasticity and  $n > 120$  under heteroskedasticity (Table 5 in the heteroskedastic case is presented in the Supplemental Appendix). From Table 4 we can see that this leads to nominal 5% rejection rates essentially equal to 5% for all cases.

Overall, we deduce that the combined  $G_3/G_4$  implementation is near-exact, with minimal computation time (less than 0.01 second) for sample sizes up to 500. While computation time is higher for larger  $n$ , more efficient programming may be able to reduce these costs.

## 7 Estimation of Approximate Distribution

The  $G_3$  and  $G_4$  distributions depend on the unknown error variances  $\sigma_i^2 = \sigma^2(X_i)$  and are thus infeasible. One feasible implementation replaces the unknown variances by estimates. In principle any nonparametric estimator  $\hat{\sigma}_i^2 = \hat{\sigma}^2(X_i)$  can be used. For illustration we use the following estimator.

Take the leave-one-out prediction errors  $\tilde{e}_i = Y_i - X_i' \tilde{\beta}_{-i}$  and their squares  $\tilde{e}_i^2$ . Take sample mean estimator  $\tilde{\sigma}^2 = n^{-1} \sum_{i=1}^n \tilde{e}_i^2$  and the regression prediction estimator  $\tilde{\sigma}_i^2 = \tilde{\gamma}_0 + \tilde{\gamma}_1 X_i^2$  where  $(\tilde{\gamma}_0, \tilde{\gamma}_1)$  are obtained by least-squares of  $\tilde{e}_i^2$  on  $X_i^2$ . Take the weighted average  $\bar{\sigma}_i^2 = \phi_1 \tilde{\sigma}^2 + \phi_2 \tilde{\sigma}_i^2$  for non-negative weights whose sum is bounded by one. Pick the weights to minimize the Mallows criterion of Hansen (2007). The resulting estimator is a shrinkage-type averaging estimator which shrinks the regression estimator towards the sample mean and towards zero. Finally, trim the estimator using the rule  $\hat{\sigma}_i^2 = \max[\bar{\sigma}_i^2, \tilde{\sigma}^2/100]$ , to bound the estimated variances away from zero.

We investigate the accuracy of this feasible distribution rule by simulation using the same designs as in the previous sections. The actual size of nominal 5% tests are calculated using the recommended  $G_3/G_4$  implementation (as described in the previous section) and the Mallows variance estimator as described in the previous paragraph. This is a fully data-dependent feasible implementation. The results are reported in Table 6. We can see that under homoskedastic errors the size of the White t-ratio  $T$  is near-exact, equalling 5% for most cases. Under heteroskedastic errors the test is over-sized for the smallest samples but has excellent size for  $n \geq 120$ .

The performance of the feasible distribution for the smallest samples in Table 6 is

Table 6: Actual size using HC1 covariance matrix and 5% critical values from feasible  $G_3/G_4$  distribution

	$n = 30$	$n = 60$	$n = 120$	$n = 250$	$n = 500$
	Homoskedastic Error				
Dummy	0.05	0.05	0.05	0.05	0.05
Pareto(2)	0.05	0.05	0.05	0.05	0.05
Gamma(1/4,1)	0.05	0.05	0.05	0.05	0.05
logNormal(0,1)	0.05	0.05	0.05	0.05	0.05
logNormal(0,4)	0.05	0.05	0.05	0.05	0.05
	Heteroskedastic Error				
Dummy	0.07	0.06	0.06	0.05	0.05
Pareto(2)	0.10	0.08	0.06	0.05	0.05
Gamma(1/4,1)	0.09	0.07	0.06	0.05	0.05
logNormal(0,1)	0.09	0.06	0.05	0.05	0.05
logNormal(0,4)	0.11	0.07	0.05	0.05	0.05

disappointing, as nominal 5% rejection rates are as high as 11%. To improve these rejection rates we recommend using the HC3 covariance matrix along with our proposed feasible finite sample distribution. The HC3 estimator and modification to the distributional calculations are described in the following section. The actual size of nominal 5% tests calculated using the HC3 covariance matrix estimator and our recommended feasible  $G_3/G_4$  implementation are reported in table 7. The performance of the test is greatly improved relative to Table 6 and other feasible methods. Under homoskedastic errors the size is exactly 5% for all cases. For heteroskedastic errors the tests have size 5% for  $n \geq 250$  but are over-sized for  $n \leq 120$ . For  $n = 120$  the actual size is 5%-6%. For  $n = 60$  the size is 6%-7%. For  $n = 30$  the size is 6%-8%. Hence for small samples (and extremely leveraged regressor designs) the test remains over-sized, but not dramatically so, and greatly reduced relative to the excessive size distortions of the existing feasible methods (Tables 1-3).

Overall, Tables 6-7 show that feasible p-values can be implemented with generally excellent accuracy except in the most extreme cases, especially if the HC3 covariance matrix estimator is used. Accuracy is better than other existing methods, such as the student t distribution, the wild bootstrap, and the Bell-McCaffrey distribution.

Table 7: Actual size using HC3 covariance matrix and 5% critical values from feasible  $G_3/G_4$  distribution

	$n = 30$	$n = 60$	$n = 120$	$n = 250$	$n = 500$
	Homoskedastic Error				
Dummy	0.05	0.05	0.05	0.05	0.05
Pareto(2)	0.05	0.05	0.05	0.05	0.05
Gamma(1/4,1)	0.05	0.05	0.05	0.05	0.05
logNormal(0,1)	0.05	0.05	0.05	0.05	0.05
logNormal(0,4)	0.05	0.05	0.05	0.05	0.05
	Heteroskedastic Error				
Dummy	0.06	0.06	0.05	0.05	0.05
Pareto(2)	0.08	0.07	0.06	0.05	0.05
Gamma(1/4,1)	0.08	0.07	0.06	0.05	0.05
logNormal(0,1)	0.07	0.06	0.05	0.05	0.05
logNormal(0,4)	0.08	0.06	0.05	0.05	0.05

## 8 Alternative Covariance Matrix Estimators

Theorems 1 and 2 are written for the White t-ratio  $T$  computed with the HC1 covariance matrix estimator. The result immediately generalizes to several other covariance matrix estimators. For the HC0, HC2, and HC3 estimators, the only difference is that the matrix  $\mathbf{B}_1$  in (4) has an alternative scaling, as we now describe.

**HC0.** This is the estimator proposed by White (1980).

$$\hat{\mathbf{V}}_{\text{HC0}} = (\mathbf{X}'\mathbf{X})^{-1} \left( \sum_{i=1}^n X_i X_i' \hat{e}_i^2 \right) (\mathbf{X}'\mathbf{X})^{-1}.$$

For the HC0 estimator, replace  $\mathbf{B}_1$  in (4) with

$$\mathbf{B}_0 = \text{diag}(Z_1^2, \dots, Z_n^2).$$

**HC2.** This is the estimator proposed by Horn, Horn and Duncan (1975).

$$\hat{\mathbf{V}}_{\text{HC2}} = (\mathbf{X}'\mathbf{X})^{-1} \left( \sum_{i=1}^n X_i X_i' \frac{\hat{e}_i^2}{(1 - h_{ii})} \right) (\mathbf{X}'\mathbf{X})^{-1}$$

where  $h_{ii} = X_i' (\mathbf{X}'\mathbf{X})^{-1} X_i$  are the leverage values. For the HC2 estimator, replace  $\mathbf{B}_1$  in

(4) with

$$\mathbf{B}_2 = \text{diag}\left(\frac{Z_1^2}{(1-h_{11})}, \dots, \frac{Z_n^2}{(1-h_{nn})}\right).$$

**HC3.** This is the estimator derived by MacKinnon and White (1985) from the jack-knife principle and by Andrews (1991) based on the principle of leave-one-out cross-validation.

$$\hat{\mathbf{V}}_{\text{HC3}} = (\mathbf{X}'\mathbf{X})^{-1} \left( \sum_{i=1}^n X_i X_i' \frac{\hat{e}_i^2}{(1-h_{ii})^2} \right) (\mathbf{X}'\mathbf{X})^{-1}$$

For the HC3 estimator, replace  $\mathbf{B}_1$  in (4) with

$$\mathbf{B}_3 = \text{diag}\left(\frac{Z_1^2}{(1-h_{11})^2}, \dots, \frac{Z_n^2}{(1-h_{nn})^2}\right).$$

## 9 Appendix A: Moment Matching

The parameters of the three-parameter and four-parameter distributions are obtained by selecting the parameters to match the moments of  $Q(w)$ , which are determined by the scaled cumulants  $\mu_r = \sum_{j=1}^K w_j^r$ . The solution are as follows.

**Three-Parameter Distribution.** (Hall (1983) and Buckley and Eagleson (1988))

$$Q_3 = a\chi_\eta^2 + b.$$

Set  $a = \mu_3/\mu_2$ ,  $b = \mu_1 - \mu_2^2/\mu_3$ , and  $\eta = \mu_2^3/\mu_3^2$ .

**Four-Parameter Distribution.** (Hillier and O'Brien (2019)).

$$Q_4 = a_1\chi_{\eta_1}^2 + a_2\chi_{\eta_2}^2.$$

The parameters are normalized as  $a_1 \leq a_2$ . Set

$$\begin{aligned} a_2 &= \frac{2}{\rho - \sqrt{\rho^2 - 4\psi}} \\ a_1 &= \frac{\mu_3 - \mu_2 a_2}{\mu_2 - \mu_1 a_2} \\ \eta_1 &= \frac{\mu_1 a_2 - \mu_2}{a_1(a_2 - a_1)} \\ \eta_2 &= \frac{\mu_1 - \eta_1 a_1}{a_2} \end{aligned}$$

where

$$\rho = \frac{\mu_1 \mu_4 - \mu_2 \mu_3}{\mu_2 \mu_4 - \mu_3^2}, \quad \psi = \frac{\mu_1 \mu_3 - \mu_2^2}{\mu_2 \mu_4 - \mu_3^2}.$$

## 10 Appendix B: Proofs

**Proof of Theorem 1:** All probability calculations are conditional on the regressor matrix  $\mathbf{X}$ . To simplify the notation we will not make this conditioning explicit.

The t-ratio equals

$$\begin{aligned} T &= \frac{R'(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{e})}{\sqrt{\left(\frac{n}{n-k}\right) R'(\mathbf{X}'\mathbf{X})^{-1} \left(\sum_{i=1}^n X_i X_i' \hat{\varepsilon}_i^2\right) (\mathbf{X}'\mathbf{X})^{-1} R}} \\ &= \frac{\mathbf{Z}'\mathbf{e}}{\sqrt{\left(\frac{n}{n-k}\right) \left(\sum_{i=1}^n Z_i^2 \hat{\varepsilon}_i^2\right)}} \\ &= \frac{\mathbf{Z}'\mathbf{e}}{\sqrt{\hat{\boldsymbol{\varepsilon}}' \mathbf{B}_1 \hat{\boldsymbol{\varepsilon}}}} \\ &= \frac{\mathbf{Z}'\boldsymbol{\Sigma}^{1/2}\boldsymbol{\varepsilon}}{\sqrt{\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{1/2} \mathbf{M} \mathbf{B}_1 \mathbf{M} \boldsymbol{\Sigma}^{1/2} \boldsymbol{\varepsilon}}} \end{aligned}$$

where  $\boldsymbol{\varepsilon} = \boldsymbol{\Sigma}^{-1/2} \mathbf{e} \sim \mathbf{N}(0, \mathbf{I}_n)$ . This expression is an odd function of  $\boldsymbol{\varepsilon}$ , so is symmetrically distributed about zero. Thus for  $x > 0$ ,

$$\mathbb{P}[T \leq x] = \frac{1}{2} (\mathbb{P}[|T| \leq x] + 1).$$

We calculate that for  $x > 0$ ,

$$\begin{aligned}
\mathbb{P}[|T| \leq x] &= \mathbb{P}[T^2 \leq x^2] \\
&= \mathbb{P}\left[\frac{(\mathbf{Z}'\Sigma^{1/2}\boldsymbol{\varepsilon})^2}{\boldsymbol{\varepsilon}'\Sigma^{1/2}\mathbf{M}\mathbf{B}_1\mathbf{M}\Sigma^{1/2}\boldsymbol{\varepsilon}} \leq x^2\right] \\
&= \mathbb{P}[\boldsymbol{\varepsilon}'\Sigma^{1/2}\mathbf{Z}\mathbf{Z}'\Sigma^{1/2}\boldsymbol{\varepsilon} \leq x^2\boldsymbol{\varepsilon}'\Sigma^{1/2}\mathbf{M}\mathbf{B}_1\mathbf{M}\Sigma^{1/2}\boldsymbol{\varepsilon}] \\
&= \mathbb{P}[\boldsymbol{\varepsilon}'\mathbf{C}\boldsymbol{\varepsilon} \leq 0]
\end{aligned}$$

where  $\mathbf{C}$  is defined in (5). By the spectral decomposition,  $\mathbf{C} = \mathbf{H}'\Lambda\mathbf{H}$ , where  $\mathbf{H}'\mathbf{H} = \mathbf{I}_n$  and  $\Lambda = \text{diag}(\lambda_0, \lambda_1, \dots, \lambda_{n-1})$  are the eigenvalues of  $\mathbf{C}$ . As described in the text,  $\lambda_0 > 0$ ,  $\lambda_j < 0$  for  $j = 1, \dots, N$ , and the remainder are zero. Set  $\zeta = \mathbf{H}\boldsymbol{\varepsilon} \sim \mathbf{N}(0, \mathbf{I}_n)$  and partition  $\zeta$  in conformity with  $\Lambda$ . Then

$$\boldsymbol{\varepsilon}'\mathbf{C}\boldsymbol{\varepsilon} = \zeta'\Lambda\zeta = \lambda_0\zeta_0^2 + \sum_{j=1}^N \lambda_j\zeta_j^2 = \lambda_0\zeta_0^2 + \sum_{j=1}^N \lambda_j Q_j$$

where  $Q_j = \zeta_j^2 \sim \chi_1^2$  are mutually independent and independent of  $\zeta_0 \sim \mathbf{N}(0, 1)$ . Hence

$$\begin{aligned}
\mathbb{P}[|T| \leq x] &= \mathbb{P}[\boldsymbol{\varepsilon}'\mathbf{C}\boldsymbol{\varepsilon} \leq 0] \\
&= \mathbb{P}\left[\lambda_0\zeta_0^2 \leq -\sum_{j=1}^N \lambda_j Q_j\right] \\
&= \mathbb{P}\left[\zeta_0^2 \leq \sum_{j=1}^N w_j Q_j\right] \\
&= \mathbb{P}\left[|\zeta_0| \leq \sqrt{Q(w)}\right] \\
&= 2\mathbb{P}\left[\zeta_0 \leq \sqrt{Q(w)}\right] - 1.
\end{aligned}$$

The final equality uses the fact that the distribution is symmetric about zero.

Using this expression, the law of iterated expectations, and the definition  $\Phi(t) = \mathbb{P}[\zeta_0 \leq t]$ , we deduce that

$$\begin{aligned}
\mathbb{P}[T \leq x] &= \mathbb{P}\left[\zeta_0 \leq \sqrt{Q(w)}\right] \\
&= \mathbb{E}\left(\mathbb{P}\left[\zeta_0 \leq \sqrt{Q(w)} \mid Q(w)\right]\right) \\
&= \mathbb{E}\left[\Phi\left(\sqrt{Q(w)}\right)\right]. \tag{17}
\end{aligned}$$

This is (6).

As derived by Ruben (1962) and Farebrother (1984), the density function of  $Q = Q(w)$  when  $Q_j^2 \sim \chi_{\eta_j}^2$  for general degrees of freedom  $\eta_j$  is

$$g_Q(t) = \sum_{m=0}^{\infty} \frac{c_m}{w_1} g_{\eta+2m}(t/w_1) \quad (18)$$

where  $g_r(t)$  is the chi-square density function with  $r$  degrees of freedom,  $\eta = \prod_{j=1}^N \eta_j$ , and the coefficients  $c_m$  are defined as

$$\begin{aligned} c_0 &= \prod_{j=1}^N \left( \frac{w_1}{w_j} \right)^{\eta_j/2} \\ c_m &= \frac{1}{m} \sum_{j=0}^{m-1} c_j d_{m-j} \\ d_m &= \sum_{j=1}^N \frac{\eta_j}{2} \left( 1 - \frac{w_1}{w_j} \right)^m, \end{aligned}$$

This specializes to  $\eta = N$  and (9)-(11) when  $\eta_j = 1$  as for  $Q(w)$ .

Using the law of iterated expectations and the definition  $\Phi(t) = \mathbb{P}[\zeta \leq t]$ , (17) equals

$$\mathbb{P}[\zeta \leq \sqrt{Q}] = \mathbb{E} \left[ \mathbb{P}[\zeta \leq \sqrt{Q} \mid Q] \right] = \mathbb{E} \left[ \Phi(\sqrt{Q}) \right].$$

Writing (17) using the density (18), integrating term-by-term, and making the change of variables  $s = t/w_1$ , we find

$$\begin{aligned} \mathbb{E} \left[ \Phi(\sqrt{Q}) \right] &= \int_0^{\infty} \Phi(\sqrt{t}) \sum_{m=0}^{\infty} \frac{c_m}{w_1} g_{N+2m}(t/w_1) dt \\ &= \sum_{m=0}^{\infty} c_m \int_0^{\infty} \Phi(\sqrt{s}\sqrt{w_1}) g_{N+2m}(s) ds \\ &= \sum_{m=0}^{\infty} c_m F_{N+2m} \left( \sqrt{w_1(N+2m)} \right) \end{aligned} \quad (19)$$

which is the stated result. The final equality is

$$\int_0^{\infty} \Phi(\sqrt{sa}) g_r(s) ds = F_r(\sqrt{ra}) \quad (20)$$

which we now establish. Let  $\phi(t)$  and  $f_r(t)$  denote the standard normal and the student



t density functions. Since  $\Phi(s) = \int_{-\infty}^s \phi(t) dt$ ,

$$\begin{aligned}
\int_0^\infty \Phi(\sqrt{s}a) g_r(s) ds &= \int_0^\infty \left( \int_{-\infty}^s \sqrt{s} \phi(\sqrt{st}) dt \right) g_r(s) ds \\
&= \int_{-\infty}^s \left( \int_0^\infty \frac{\exp(-st^2/2)}{\sqrt{2\pi}} \sqrt{s} \frac{s^{r/2-1} \exp(-s/2)}{2^{r/2} \Gamma(\frac{r}{2})} ds \right) dt \\
&= \int_{-\infty}^s \left( \int_0^\infty \frac{s^{(r+1)/2-1} \exp(-s(1+t^2)/2)}{\sqrt{\pi} 2^{(r+1)/2} \Gamma(\frac{r}{2})} ds \right) dt \\
&= \int_{-\infty}^s \frac{\Gamma(\frac{r+1}{2})}{\sqrt{\pi} \Gamma(\frac{r}{2}) (1+t^2)^{(r+1)/2}} dt \\
&= \int_{-\infty}^s \sqrt{r} f_r(\sqrt{r}t) dt \\
&= F_r(\sqrt{r}a)
\end{aligned}$$

as claimed. ■

**Proof of Theorem 2:** The random variable  $Q_4 = a_1 \chi_{\eta_1}^2 + a_2 \chi_{\eta_2}^2$  is a special case of that studied by Ruben (1962) and Farebrother (1984), which has density (18). Thus by the argument of the proof of Theorem 1, the distribution  $G_4$  is equal to that as specified, with the coefficients  $b_m$  equal to

$$\begin{aligned}
b_0 &= \left( \frac{a_1}{a_2} \right)^{\eta_2/2} \\
b_m &= \frac{1}{m} \sum_{j=0}^{m-1} b_j \frac{\eta_2}{2} \left( 1 - \frac{a_1}{a_2} \right)^{m-j}.
\end{aligned}$$

The derivation is completed by showing that  $b_m$  satisfies (14). The proof is by induction. Notice that  $b_0$  satisfies (14) for  $m = 0$ . Assume that  $b_m$  satisfies (14) for  $j < m$ . Then

$$\begin{aligned}
b_m &= \frac{1}{m} \sum_{j=0}^{m-1} b_j \frac{\eta_2}{2} \left( 1 - \frac{a_1}{a_2} \right)^{m-j} \\
&= \frac{1}{m} \sum_{j=0}^{m-1} \frac{\left( \frac{a_1}{a_2} \right)^{\eta_2/2} \left( 1 - \frac{a_1}{a_2} \right)^j \Gamma(j + \frac{\eta_2}{2})}{\Gamma(\frac{\eta_2}{2}) j!} \frac{\eta_2}{2} \left( 1 - \frac{a_1}{a_2} \right)^{m-j} \\
&= \frac{\left( \frac{a_1}{a_2} \right)^{\eta_2/2} \left( 1 - \frac{a_1}{a_2} \right)^m}{\Gamma(\frac{\eta_2}{2})} \frac{1}{m} \sum_{j=0}^{m-1} \frac{\Gamma(j + \frac{\eta_2}{2})}{j!} \frac{\eta_2}{2}.
\end{aligned}$$

The proof is completed by showing that

$$\frac{1}{m} \sum_{j=0}^{m-1} \frac{\Gamma(j + \frac{\eta_2}{2})}{j!} \frac{\eta_2}{2} = \frac{\Gamma(m + \frac{\eta_2}{2})}{m!}. \quad (21)$$

Take the binomial identity

$$\binom{a}{n} = \sum_{j=0}^n \binom{j + a - n - 1}{j}$$

(see, e.g., equation 26.3.8 of Olver, Lozier, Boisvert, and Clark (2010)). Set  $n = m - 1$  and  $a = \eta_2/2 - m + 1$ . This implies

$$\sum_{j=0}^{m-1} \binom{j + \frac{\eta_2}{2} - 1}{j} = \binom{\frac{\eta_2}{2} + m - 1}{m - 1}$$

and thus

$$\sum_{j=0}^{m-1} \frac{\Gamma(j + \frac{\eta_2}{2})}{\Gamma(\frac{\eta_2}{2}) j!} = \frac{\Gamma(m + \frac{\eta_2}{2})}{\Gamma(\frac{\eta_2}{2} + 1)(m - 1)!}$$

which is (21) if we multiply both sides by  $\Gamma(\frac{\eta_2}{2} + 1)/m$ . This completes the proof.  $\blacksquare$

**Proof of Theorem 3:** Set  $\delta = a_1/a_2$ . Expression (14) and the inequality  $\Gamma(m + r)/m! \leq (m + r - 1)^{r-1}$  imply

$$b_m = \frac{\delta^{\eta_2/2} (1 - \delta)^m \Gamma(m + \frac{\eta_2}{2})}{\Gamma(\frac{\eta_2}{2}) m!} \leq \frac{\delta^{\eta_2/2} (1 - \delta)^m}{\Gamma(\frac{\eta_2}{2})} \left(m + \frac{\eta_2}{2} - 1\right)^{\eta_2/2 - 1}. \quad (22)$$

Using the triangle inequality, the fact  $|F_\eta(x)| \leq 1$ , expression (22), the change-of-index  $s = m + \eta_2/2 - 1$ , the equality  $(1 - \delta)^s = \exp(-cs)$ , and the change-of-variables

$t = 2cs$ , we find that

$$\begin{aligned}
T_M &= \left| \sum_{m=M+1}^{\infty} b_m F_{\eta_1+\eta_2} \left( \sqrt{a_1(\eta_1+\eta_2+2m)} \right) \right| \\
&\leq \sum_{m=M+1}^{\infty} \frac{\delta^{\eta_2/2} (1-\delta)^m}{\Gamma\left(\frac{\eta_2}{2}\right)} \left(m + \frac{\eta_2}{2} - 1\right)^{\eta_2/2-1} \\
&= \sum_{s=M+\eta_2/2}^{\infty} \frac{\delta^{\eta_2/2} (1-\delta)^{s+1-\eta_2/2}}{\Gamma\left(\frac{\eta_2}{2}\right)} s^{\eta_2/2-1} \\
&\leq \frac{\delta^{\eta_2/2}}{(1-\delta)^{\eta_2/2-1}} \int_{M+\eta_2/2}^{\infty} \frac{\exp(-cs) s^{\eta_2/2-1}}{\Gamma\left(\frac{\eta_2}{2}\right)} ds \\
&= \frac{\delta^{\eta_2/2}}{(1-\delta)^{\eta_2/2-1} c^{\eta_2/2}} \int_{2cM+c\eta_2}^{\infty} \frac{\exp(-t/2) t^{\eta_2/2-1}}{2^{\eta_2/2} \Gamma\left(\frac{\eta_2}{2}\right)} dt \\
&= \frac{\delta^{\eta_2/2}}{(1-\delta)^{\eta_2/2-1} c^{\eta_2/2}} (1 - H_{\eta_2}(2cM + c\eta_2)) \\
&\leq \tau
\end{aligned}$$

where  $H_{\eta}(x)$  is the  $\chi_{\eta}^2$  distribution function, and the final inequality holds because condition (16) implies

$$H_{\eta_2}(2cM + c\eta_2) \geq 1 - \frac{c^{\eta_2/2} (1-\delta)^{\eta_2/2-1}}{\delta^{\eta_2/2}} \tau.$$

■

## References

- [1] Andrews, Donald W. K. (1991): “Asymptotic normality of series estimators for non-parametric and semiparametric regression models,” *Econometrica*, 59, 307-345.
- [2] Angrist, Joshua D. and Jörn-Steffen Pischke (2009): *Mostly Harmless Econometrics: An Empiricists Companion*, Princeton University Press.
- [3] Bell, Robert M., and Daniel F. McCaffrey (2002): “Bias reduction in standard errors for linear regression with multi-stage samples,” *Survey Methodology*, 28, 169-181.
- [4] Bodenham, Dean A. and Niall M. Adams (2015): “A comparison of efficient approximations for a weighted sum of chi-squared random variables,” *Statistics and Computing*, 26, 917-928.
- [5] Buckley, M. J., and G. K. Eagleson (1988): “An approximation to the distribution of quadratic forms in normal random variables,” *Australian Journal of Statistics*, 30, 150-159.
- [6] Chesher, Andrew D. (1989): “Hájek inequalities, measures of leverage, and the size of heteroskedasticity robust Wald tests,” *Econometrica*, 57, 971-977.
- [7] Chesher, Andrew D. and Gerard Austin (1991): “The finite-sample distributions of heteroskedasticity robust Wald statistics,” *Journal of Econometrics*, 47, 153-173.
- [8] Chesher, Andrew D. and Ian D. Jewitt (1987): “The bias of the heteroskedasticity consistent covariance matrix estimator,” *Econometrica*, 55, 1217-1272.
- [9] Chu, Jianghao, Tae-Hwy Lee, Aman Ullah, and Haifeng Xu (2021): “Exact distribution of the F-statistic under heteroskedasticity of unknown form for improved inference,” *Journal of Statistical Computation and Simulation*.
- [10] Davidson, Russell and Emmanuel Flachaire (2008): “The wild bootstrap, tamed at last,” *Journal of Econometrics*, 146, 162-169.
- [11] Davies, Robert B. (1980): “Algorithm AS 155: The distribution of a linear combination of  $\chi^2$  random variables,” *Journal of the Royal Statistical Society, Series C*, 29, 323-333.

- [12] Eicker, Friedhelm (1963): "Asymptotic normality and consistency of the least squares estimators for families of linear regressions," *Annals of Mathematical Statistics*, 34, 447-456.
- [13] Farebrother, R. W. (1984): "Algorithm AS 204: The distribution of a positive linear combination of  $\chi^2$  random variables," *Journal of the Royal Statistical Society, Series C*, 33, 332-339.
- [14] Gosset, William S. (a.k.a. "Student") (1908): "The probable error of a mean," *Biometrika*, 6, 1-25.
- [15] Hall, Peter (1983): "Chi squared approximations to the distribution of a sum of independent random variables," *Annals of Probability*, 11, 1028-1036.
- [16] Hansen, Bruce E. (2007): "Least squares model averaging," *Econometrica*, 75, 1175-1189.
- [17] Hillier, Grant and Raymond O'Brien (2019): "A simple four-moment approximation to the distribution of a positive definite quadratic form, with applications to testing," working paper.
- [18] Hinkley, David V. (1977): "Jackknifing in unbalanced situations," *Technometrics*, 19, 285-292.
- [19] Horn, S. D., R. A. Horn, and D. B. Duncan. (1975): "Estimating heteroscedastic variances in linear model," *Journal of the American Statistical Association*, 70, 380-385.
- [20] Imbens, Guido W. and Michal Kolesár (2016): "Robust standard errors in small samples: Some practical advice," *Review of Economics and Statistics*, 98, 701-712.
- [21] Imhof, J. P. (1961): "Computing the distribution of quadratic forms in normal variables," *Biometrika*, 48, 419-426.
- [22] MacKinnon, James G. and Halbert White (1985): "Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties," *Journal of Econometrics*, 29, 305-325.
- [23] Olver, Frank W. J., Daniel W. Lozier, Ronald F. Boisvert, and Charles W. Clark (2010): *NIST Handbook of Mathematical Functions*, Cambridge University Press.

- [24] Phillips, Peter C. B. (1993): "Operational algebra and regression t-tests," in Peter C. B. Phillips (editor), *Models, Methods and Applications of Econometrics: Essays in Honor of A. R. Bergstrom*, Blackwell.
- [25] Pötscher, Benedikt M. and David Preinerstorfer (2021): "Valid heteroskedasticity robust testing," working paper.
- [26] Ruben, Harold (1962): "Probability content of regions under spherical normal distributions, IV: The distribution of homogeneous and non-homogeneous quadratic functions of normal variables," *Annals of Mathematical Statistics*, 33, 542-570.
- [27] Satterthwaite, F. E. (1946): "An approximate distribution of estimates of variance components," *Biometrics Bulletin*, 2, 110-114.
- [28] Welch, B. L. (1938): "The significance of the difference between two means when the population variances are unequal," *Biometrika*, 29, 350-362.
- [29] White, Halbert (1980): "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity," *Econometrica*, 48, 817-838.