Finite Sample Distributions of the Wald, Likelihood Ratio and Lagrange Multiplier Test Statistics in the Classical Linear Model

Thomas Parker

University of Waterloo

August 26, 2015

Abstract

In this note it is shown that the finite sample distributions of the Wald, likelihood ratio and Lagrange multiplier statistics in the classical linear regression model are members of the generalized beta model introduced by McDonald and Xu (1995a). This is useful for examining properties of these test statistics. For example, this characterization makes it easy to find distribution-, quantile- and density functions for each test statistic, makes it clear why Wald tests may overreject the null hypothesis using asymptotic critical values and formalizes the fact that the Lagrange multiplier statistic follows a distribution with bounded support.

Keywords: Linear regression model, Generalized beta distribution, Finite sample distribution

1 Introduction

The Wald, likelihood ratio and Lagrange multiplier (or Rao score) tests are quite well known methods of testing linear hypotheses in regression analysis. Engle (1984) shows that in linear regressions, all three test statistics can be expressed as simple functions of regression residuals. Exact finite-sample distributions for these test statistics are generally intractable, but typically their distributions converge under the null hypothesis to a χ^2 variable. Because asymptotic χ^2 critical values are used for finite-sample inference, the three statistics can lead to different decisions in hypothesis tests — see for example Berndt and Savin (1977) or Buse (1982).

McDonald (1984) introduced the *generalized beta distribution* as a parametric model for use in investigating income distributions. This very flexible parametric family was followed by further generalizations in McDonald and Xu (1995a), with applications to the distribution of income and financial return distributions. More recently, Mauler and McDonald (2015) revisited the modeling of option price distributions using flexible parametric distributions and found that the generalized beta offers superior predictive performance over several of its proposed competitors, many of which are its own special or limiting cases. Researchers modeling financial returns continue to find novel applications of this distribution, for example its use in exponential general autoregressive conditional heteroskedastic (EGARCH) models, in which the exponential generalized beta (a variant that will be discussed below) is attractive due to its robustness to outlying observations, as shown in Caivano and Harvey (2014).

In this note, we derive exact finite-sample distributions for all three test statistics when they are based on residuals from a classical linear regression model, and it is furthermore shown that they are all members of the generalized beta model. The standard beta distribution is also a member of this family, and it is shown that quantiles and cumulative probabilities of the standard beta distribution can easily be transformed to give corresponding quantities for all three test statistics. These results may not see much use in direct application — these test statistics are most appropriate for use in nonlinear models, at which point the distributions derived here would no longer be exact — but the addition of a few new members to the family of generalized beta distributions may be of independent interest, and these results may be indicative of the behavior of the classical test statistics in more complex models.

2 Definitions and functional relationships

Suppose the model is $y = X^\top \beta + \varepsilon$ based a sample of n iid observations with $\beta \in \mathbb{R}^k$ and $\varepsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_n)$ is independent of X. We label this collection of assumptions the classical linear model for brevity's sake. Consider a test of the linear hypothesis

$$H_0 : R\beta = \delta, \tag{1}$$

where the rank of R is r. Letting $\tilde{\beta}$ and $\hat{\beta}$ be maximum likelihood estimates of β respectively with and without the restrictions implied by the null hypothesis, define the following test statistics:

$$W = n(R\hat{\beta} - \delta)^\top \left(RI^{-1}(\hat{\beta})R^\top\right)^{-1} (R\hat{\beta} - \delta) \tag{2}$$

$$LR = 2\big(\ell(\hat{\beta}) - \ell(\tilde{\beta})\big) \tag{3}$$

$$LM = \frac{1}{n} s(\tilde{\beta})^\top I^{-1}(\tilde{\beta}) s(\tilde{\beta}), \tag{4}$$

respectively the Wald, likelihood ratio and Lagrange multiplier test statistics. In the above definitions the loglikelihood function associated with the model is $\ell(\cdot) = \sum_i \log f(y_i, \cdot|X_i)$, the information matrix is $I(\cdot) = -\frac{1}{n}\mathrm{E}\left[\nabla^2_{\beta\beta}\ell\right]$, and $s(\cdot) = \sum_i \nabla_\beta \ell(y_i, \cdot|X_i)$ is the average of individual score contributions. It is a standard exercise to show that these three statistics converge asymptotically to a χ^2_r random variable under the null hypothesis, although they have different distributions in finite samples.

In the classical linear model, nonstochastic relationships between the four statistics are known (Engle, 1984, p. 788), and these can be used to derive finite-sample distributions under the null hypothesis for the three test statistics. Let F, W, LR and LM denote the realized value of the respective test statistics. In linear regression models the Wald and $F_{r,n-k}$ statistics are related by the following equation:

$$W = \frac{nr}{n-k} F. \tag{5}$$

The relationship between the likelihood ratio statistic and the $F_{r,n-k}$ statistic is

$$LR = n \log\left(1 + \frac{W}{n}\right) = n \log\left(1 + \frac{r}{n-k} F\right). \tag{6}$$

Finally,

$$LM = \frac{W}{1 + \frac{W}{n}} = \frac{\frac{nr}{n-k} F}{1 + \frac{r}{n-k} F}. \tag{7}$$

These relationships could be used to find quantiles for each distribution, since all are monotone functions of the F statistic. However, they also provide the means to show that the distributions of all of these statistics are members of one parent model, as will be shown below.

3 Exact finite-sample densities

Theorem 1 below exploits the nonstochastic relationships outlined above and the density of the $F_{r,n-k}$ distribution. For reference, the F distribution with r and $n-k$ degrees of freedom has the density function

$$f_F(x;r,n-k) = \frac{\left(\frac{r}{n-k}\right)^{r/2} x^{r/2-1}}{B\left(\frac{r}{2},\frac{n-k}{2}\right)\left(1+\frac{r}{n-k}x\right)^{(r+n-k)/2}}, \qquad x \in [0,\infty) \tag{8}$$

where B is the beta function.

Theorem 1. *Let f_W, f_{LR} and f_{LM} respectively denote the density functions of the Wald statistic (2), the likelihood ratio statistic (3) and the Lagrange multiplier statistic (4) in the classical linear model. Then*

$$f_W(x;r,n-k) = \frac{\left(\frac{x}{n}\right)^{r/2-1}}{nB\left(\frac{r}{2},\frac{n-k}{2}\right)\left(1+\frac{x}{n}\right)^{(r+n-k)/2}}, \qquad x \in [0,\infty), \tag{9}$$

$$f_{LR}(x;r,n-k) = \frac{\left(e^{-x/n}\right)^{(n-k)/2}\left(1-e^{-x/n}\right)^{r/2-1}}{nB\left(\frac{n-k}{2},\frac{r}{2}\right)}, \qquad x \in [0,\infty) \tag{10}$$

and

$$f_{LM}(x;r,n-k) = \frac{\left(\frac{x}{n}\right)^{r/2-1}\left(1-\frac{x}{n}\right)^{(n-k)/2-1}}{nB\left(\frac{r}{2},\frac{n-k}{2}\right)}, \qquad x \in [0,n). \tag{11}$$

Interestingly, it can be seen in equation (11) that the distribution of the Lagrange multiplier statistic has bounded support. This fact is reflected in the great many Lagrange multiplier statistics in regression models that may be expressed as nR^2 from an artificial regression (see, for example, Davidson and MacKinnon (1990)). Those tests imply the finite support of the distribution in finite samples, which can be observed directly in theory here. In fact, if X is a Lagrange multiplier statistic, then X/n follows a standard beta distribution with $r/2$ and $(n-k)/2$ degrees of freedom, which can be seen in (11), a fact noted by Fisher and McAleer (1984).

4 Characterization as generalized beta distributions

The above exact densities are all straightforward consequences of the relationships that the test statistics have with the standard F statistic. However, these densities can be expressed as members of the generalized beta distribution. The distribution referred to here as the generalized beta distribution was introduced to the econometrics literature by McDonald and Xu (1995a,b), extending work started in McDonald (1984), intended for applications to the analysis of income distributions.

This distribution is defined using five parameters. Its density function is

$$f_{GB}(x;a,b,c,p,q) = \frac{|a| x^{ap-1}\left(1-(1-c)\left(\frac{x}{b}\right)^a\right)^{q-1}}{b^{ap}B(p,q)\left(1+c\left(\frac{x}{b}\right)^a\right)^{p+q}}, \quad 0 < x^a < \frac{b^a}{(1-c)} \tag{12}$$

for $a \in \mathbb{R}$, $b, p, q \in [0, \infty)$ and $c \in [0,1]$. The model is very flexible and nests many well-known models such as the F, χ^2, gamma, t, Pareto, Weibull, exponential and normal distributions. The standard beta distribution satisfies $f_B(x;p,q) = f_{GB}(x;a=1,b=1,c=0,p,q)$.

McDonald and Xu (1995a) also define a variant of the generalized beta distribution. If Y has a generalized beta distribution and $X = \ln(Y)$, then the distribution of X is called the exponential generalized beta distribution, and it can be defined via the generalized beta distribution:

$$f_{EGB}(x;\delta,\sigma,c,p,q) = f_{GB}(e^x; a = 1/\sigma, b = e^\delta, c, p, q) \cdot e^x \tag{13}$$

or more explicitly

$$f_{EGB}(x;\delta,\sigma,c,p,q) = \frac{e^{p(x-\delta)/\sigma}\left(1-(1-c)e^{(x-\delta)/\sigma}\right)^{q-1}}{|\sigma| B(p,q)\left(1+ce^{(x-\delta)/\sigma}\right)^{p+q}}, \quad -\infty < \frac{x-\delta}{\sigma} < \ln\left(\frac{1}{1-c}\right). \tag{14}$$

For example, the lognormal distribution is an exponential generalized beta distribution.

Theorem 2 shows the characterization of the Wald and Lagrange multiplier distributions as generalized beta distributions, and the likelihood ratio distribution as an exponential generalized beta distribution.

Theorem 2. *Using the same assumptions and notation as in Theorem 1,*

$$f_W(x;r,n-k) = f_{GB}\left(x; a=1, b=n, c=1, p=\frac{r}{2}, q=\frac{n-k}{2}\right), \qquad x \in [0,\infty), \tag{15}$$

$$f_{LM}(x;r,n-k) = f_{GB}\left(x; a=1, b=n, c=0, p=\frac{r}{2}, q=\frac{n-k}{2}\right), \qquad x \in [0,n) \tag{16}$$

and

$$f_{LR}(x;r,n-k) = f_{EGB}\left(-x;\delta=0,\sigma=n,c=0,p=\frac{n-k}{2},q=\frac{r}{2}\right) \tag{17}$$

$$= e^{-x}\cdot f_{GB}\left(e^{-x};a=\frac{1}{n},b=1,c=0,p=\frac{n-k}{2},q=\frac{r}{2}\right), \qquad x\in[0,\infty). \tag{18}$$

The difference between the Wald and Lagrange multiplier statistics is captured by a single parameter of the generalized beta model. Also, for the likelihood ratio statistic $p=(n-k)/2$ and $q=r/2$ — for the other two distributions these parameters are switched.

5 Quantiles and cumulative probabilities via the beta distribution

The cumulative distribution function and quantile functions of generalized beta models can be found easily from the standard beta model, which implies this should be possible for the statistics described above. McDonald and Xu (1995a) note that if $x \sim GB(a,b,c,p,q)$ and the transformation y of x is defined by

$$y = \frac{\left(\frac{x}{b}\right)^a}{1+c\left(\frac{x}{b}\right)^a} \tag{19}$$

then $y \sim \mathrm{Beta}(p,q)$. This fact is exploited in Theorem 3.

Theorem 3. *Let $B(\tau)$ denote the τ^{th} quantile of the* $\mathrm{Beta}(r/2,(n-k)/2)$ *distribution. Then the τ^{th} quantile of the distributions of the Wald, likelihood ratio and Lagrange multiplier statistics may be written as*

$$W(\tau) = \frac{nB(\tau)}{1-B(\tau)} \tag{20}$$

$$LR(\tau) = -n\log(1-B(\tau)) \tag{21}$$

$$LM(\tau) = nB(\tau). \tag{22}$$

Let F_B be the cumulative distribution function of the $\mathrm{Beta}(r/2,(n-k)/2)$ *distribution. Then*

$$F_W(x;r,n-k) = F_B\left(x/(x+n)\right) \tag{23}$$

$$F_{LR}(x;r,n-k) = 1-F_B\left(e^{-x/n}\right) \tag{24}$$

$$F_W(x;r,n-k) = F_B\left(x/n\right). \tag{25}$$

Appealing to the basic inequality $x \leq -\log(1-x) \leq \frac{x}{1-x}, 0 < x < 1$ (see for example Davidson and MacKinnon (1993, p. 456-457)), it is immediate that $LM \leq LR \leq W$, which is noted, for example, by Engle (1984, p. 792). However, this exact characterization makes it easy to examine the magnitude the effect over the whole distribution, for any configuration of parameters.

6 The effect of parameters

The effect of changes in the values of parameters n and r can be seen in Figures 1 and 2 below (we use density functions but quantile functions could also be enlightening). A decrease in $n-k$ or an increase in r tends to result in a greater divergence between the appropriate critical values for each test from the asymptotic χ^2 value. Since n and k usually enter the density formulas as a difference, an increase in k usually has a very similar effect to a decrease in n (n is introduced into the formulas in isolation via the transformations (5)-(7)). Because of this, only changes in n and r are depicted in Figures 1 and 2. The effect of a change in r is somewhat more pronounced than effects from changes in the other two parameters, heuristically because the value of r is often small.

The densities of the three finite-sample densities are compared to the asymptotic distribution in Figure 1 for two different values of n, illustrating the effect of sample size on the appropriate critical values for tests. The left panel shows that when n is small, the Wald statistic is likely to be larger than its asymptotic counterpart: in the left panel of Figure 1, the 95$^{\text{th}}$ percentile of the asymptotic χ^2_3 distribution is the 78$^{\text{th}}$ percentile of the finite-sample distribution.

Figure 2 shows the effect that different values of r has on the shape of the distributions. When r is relatively large compared to k, the upper quantile of the likelihood ratio statistic can be less than the χ^2 approximation. This is illustrated in the bottom two panels of Figure 2.

The quantile inequality between the test statistics is plain to see in the figures, represented by the relative position of the theoretical critical values on each graph. Generally it seems that for small samples, the size of the Lagrange multiplier test does not suffer very much from its χ^2 approximation, but the other two tests do, in particular the Wald test. The asymptotic approximation for the Wald test is particularly poor in situations where $n-k$ is small — see, for example, the left panel of Figure 1 or the bottom row of panels of Figure 2. In cases where $n-k$ is extremely small or r is large, the use of an exact critical value would improve the theoretical size of Wald and likelihood ratio tests (very much so

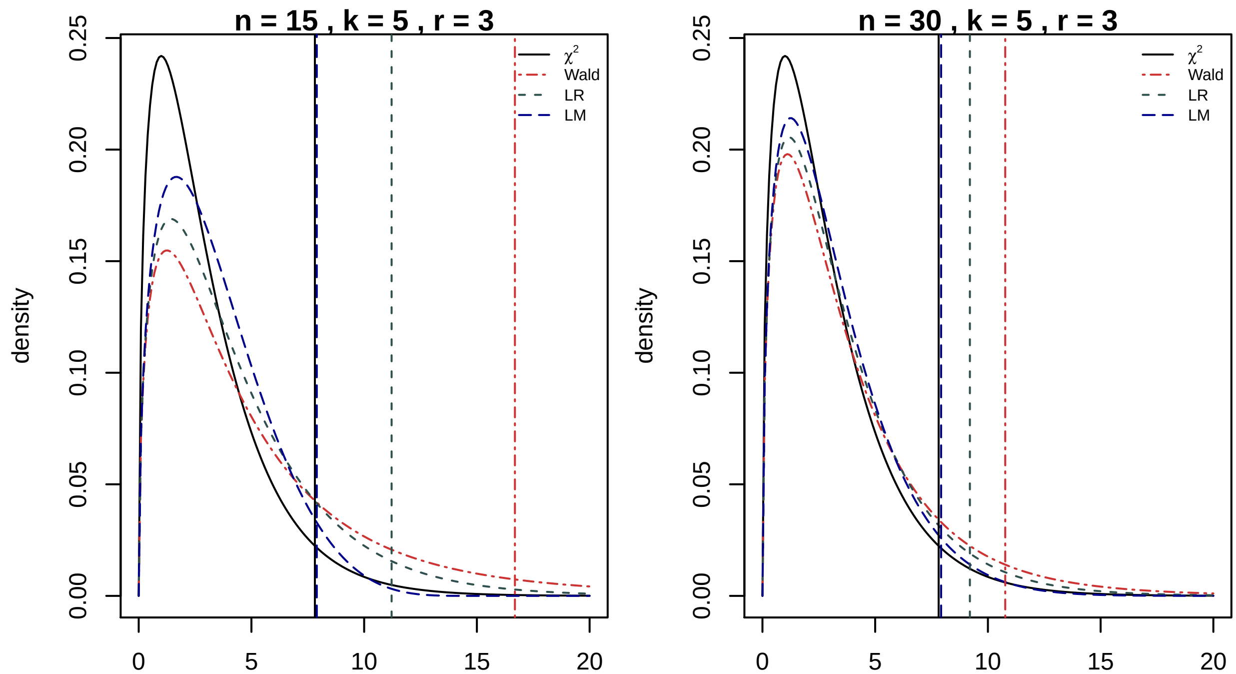

Figure 1: The effect of changes in n on densities and critical values: the 95th percentile of the likelihood ratio statistic is much closer to the asymptotic χ^2 value than the other two statistics, and is extremely close to the asymptotic value for these relatively extreme parameter configurations. In small samples, hypotheses based on tests using the Wald statistic would be rejected much more often than in tests using the other two statistics.

for the Wald test).

7 Conclusion

The generalized beta distribution provides a unified characterization of the distributions of the standard test statistics in the classical linear model. This characterization formalizes some well-known features of these test statistics and provides simple rules for finding values of their distribution-, quantile- and density functions.

Appendix: proofs

Proof of Theorem 1. The densities of all three test statistics follow from the nonstochastic relationships they share with the $F_{n-k,r}$ density, and the result that for $Y = G(X)$, $f_Y(y) = f_X(G^{-1}(y))\left|\frac{\mathrm{d}}{\mathrm{d}y}G^{-1}(y)\right|$. The Wald and likelihood ratio statistics are straightforward applications of this technique. We include

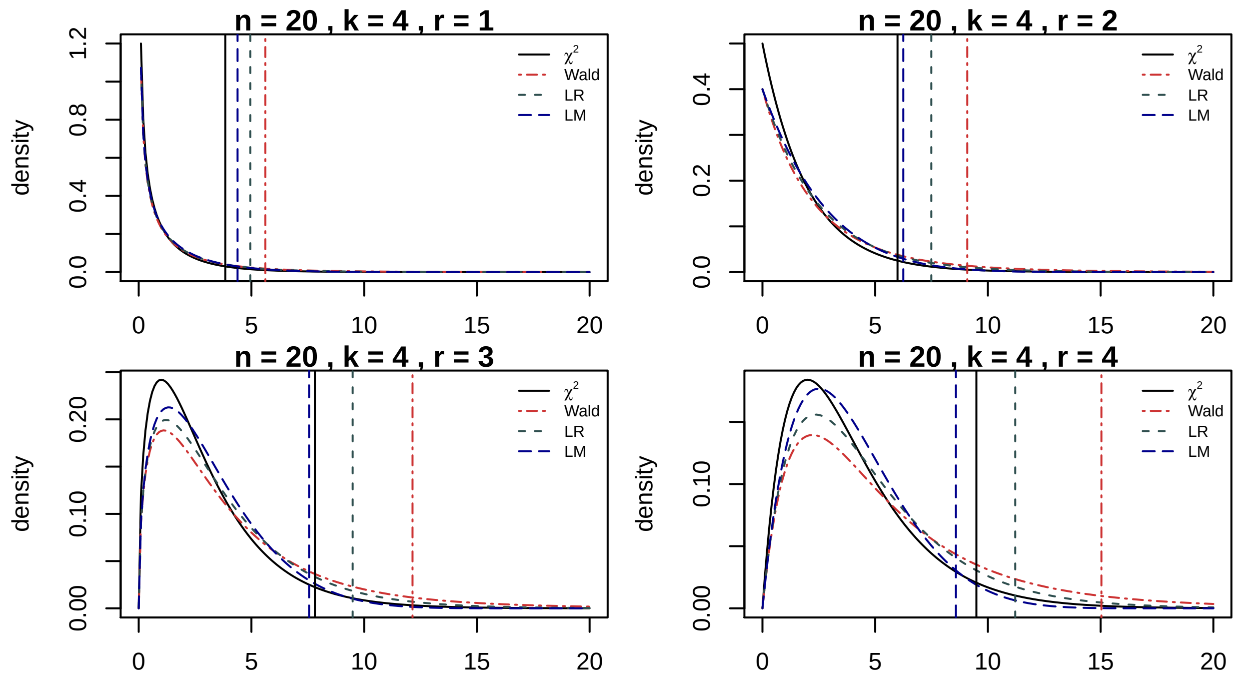

Figure 2: An increase in the number of restrictions in a test can have a pronounced effect. This is in part because r is much smaller than $n-k$ in most settings so a unit change in r when its value is small induces a drastic change in the shape of the distributions. Once again, the Wald test statistic is most affected by changes in r, is the case with respect to changes in n.

detail for the score statistic, which involves a bit more manipulation than the other two. Inverting equation (7) and taking a derivative results in

$$|J| = \frac{\frac{nr}{n-k} - \frac{r}{n-k}LM + \frac{r}{n-k}LM}{\left(\frac{nr}{n-k} - \frac{r}{n-k}LM\right)^2} = \frac{n}{\frac{r}{n-k}\left(n-LM\right)^2}. \tag{26}$$

Then the density of the statistic is

$$f_{LM}(x;r,n-k) = \frac{\left(\frac{r}{n-k}\right)^{r/2}\left(\frac{x}{\frac{nr}{n-k}-\frac{r}{n-k}x}\right)^{r/2-1}}{B\left(\frac{r}{2},\frac{n-k}{2}\right)\left(1+\frac{r}{n-k}\frac{x}{\frac{nr}{n-k}-\frac{r}{n-k}x}\right)^{(r+n-k)/2}} \cdot \frac{n}{\frac{r}{n-k}(n-x)^2} \tag{27}$$

$$= \frac{\left(\frac{x}{n-x}\right)^{r/2-1}}{B\left(\frac{r}{2},\frac{n-k}{2}\right)\left(\frac{n}{n-x}\right)^{(r+n-k)/2}} \cdot \frac{1}{n\left(\frac{n-x}{n}\right)^2} \tag{28}$$

$$= \frac{\left(\frac{x}{n-x}\right)^{r/2-1}}{nB\left(\frac{r}{2},\frac{n-k}{2}\right)\left(\frac{n}{n-x}\right)^{(1-r/2)+(1-(n-k)/2)}} \tag{29}$$

$$= \frac{\left(\frac{x}{n}\right)^{r/2-1}\left(1-\frac{x}{n}\right)^{(n-k)/2-1}}{nB\left(\frac{r}{2},\frac{n-k}{2}\right)} \tag{30}$$

which is (11). ■

Proof of Theorem 2. Verify by substituting parameters into (12) or (14) and checking against the densities given in Theorem 1. One must only take care when defining the domain of the LR statistic. ■

Proof of Theorem 3. The inverse transformation of (19) is Quantiles for the distributions of the Wald and Lagrange multiplier statistics can be determined by substituting the parameter values given in Theorem 2 into the inverse of equation (19), which is $x = b\left(\frac{y}{1+cy}\right)^{1/a}$. This results in (20) and (22).

The likelihood ratio is an exponential generalized beta variable. McDonald and Xu (1995a, Appendix A.2.2) note that if $X \sim EGB$ and $X = \log Y$, then

$$F_{EGB}(x) = \mathrm{P}\{X \leq x\} = \mathrm{P}\left\{e^Y \leq e^y\right\} = F_{GB}(e^y).$$

Combining the transformations (19) and $x = e^{-z}$, it follows that if Z has an exponential generalized beta distribution, then the transformed variable $y = \frac{\left(\frac{e^{-z}}{b}\right)^a}{1+c\left(\frac{e^{-z}}{b}\right)^a}$ has a standard beta distribution. However, because the distribution lies in $\mathbb{R}_+$, the τ^{th} quantile of z is mapped to the $(1-\tau)^{\text{th}}$ quantile of y. Substituting the required values from the third part of Theorem 2 ($a = \frac{1}{n}$, $b = 1$ and $c = 0$) implies $LR(\tau) = -n \log B'(1-\tau)$, where $B'(1-\tau)$ is the $(1-\tau)^{\text{th}}$ quantile of the Beta$\left(\frac{n-k}{2}, \frac{r}{2}\right)$ distribution. Using the identity $B'(1-\tau) = 1 - B(\tau)$ one obtains (21). To find cumulative distribution functions for each test statistic, use the transformation (19), with one extra transformation needed for the likelihood ratio statistic, in the appropriate beta distribution. ■

Acknowledgements

The author would like to thank Anil Bera for helpful discussions.

References

E. Berndt and N.E. Savin. Conflict among criteria for testing hypotheses in the multivariate linear regression model. *Econometrica*, 45(5):1263–1277, 1977.

A. Buse. The likelihood ratio, Wald, and Lagrange multiplier tests: An expository note. *The American Statistician*, 36(3):153–157, 1982.

M. Caivano and A. Harvey. Time series models with an EGB2 conditional distribution. *Journal of Time Series Analysis*, 35:558–571, 2014.

R. Davidson and J.G. MacKinnon. Specification tests based on artificial regressions. *The Journal of the American Statistical Association*, 85(409):220–227, 1990.

R. Davidson and J.G. MacKinnon. *Estimation and Inference in Econometrics*. Oxford, 1993.

R.F. Engle. Wald, likelihood ratio, and Lagrange multiplier tests in econometrics. In Z. Griliches and M.D. Intriligator, editors, *Handbook of Econometrics*, volume II, pages 775–826. North-Holland, 1984.

G. Fisher and M. McAleer. The geometry of specification error. *Australian Journal of Statistics*, 26(3): 310–322, 1984.

D.J. Mauler and J.B. McDonald. Option pricing and distribution characteristics. *Computational Economics*, 45:579–595, 2015.

J.B. McDonald. Some generalized functions for the size distribution of income. *Econometrica*, 52(3): 647–664, 1984.

J.B. McDonald and Y.J. Xu. A generalization of the beta distribution with applications. *Journal of Econometrics*, 66:133–152, 1995a.

J.B. McDonald and Y.J. Xu. Errata — a generalization of the beta distribution with applications. *Journal of Econometrics*, 69:427–428, 1995b.