An exact, unified distributional characterization of statistics used to test linear hypotheses in simple regression models

Parker, Thomas

University of Illinois at Urbana-Champaign

20 May 2010

An Exact, Unified Distributional Characterization of Statistics used to Test Linear Hypotheses in Simple Regression Models

Thomas Parker
Department of Economics, University of Illinois at Urbana-Champaign
Email: tmparker@illinois.edu

May 20, 2010

Abstract

The Wald, likelihood ratio and Lagrange multiplier test statistics are commonly used to test linear restrictions in regression models. It is shown that for testing these restrictions in the classical regression model, the exact densities of these test statistics are special cases of the generalized beta distribution introduced by McDonald (1984); McDonald and Xu (1995a). This unified derivation provides a method by which one can derive small sample critical values for each test. These results may be indicative of the behavior of such test statistics in more general settings, and are useful in visualizing how each statistic changes with different parameter values in the simple regression model. For example, the results suggest that Wald tests may severely underreject the null hypothesis when the sample size is small or a large number of restrictions are tested.

Keywords: Test of linear restrictions, Generalized beta distribution, Small-sample probability distribution, Regression model

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; i.e., hypotheses of the form

$$H_0 : R\beta = \delta$$

They are usually easily constructed; Engle (1984), for example, shows that these tests can be expressed as simple functions of regression residuals. Exact small sample distributions for these test statistics tend to be intractable, but given typical regularity conditions their distributions converge under the null hypothesis to a χ^2 variable that depends on the rank of the restrictions being tested. Because asymptotic χ^2 critical values are usually used for finite sample inference, the three statistics can lead to different decisions in hypothesis testing situations — see for example Berndt and Savin (1977) or Buse (1982).

In the classical regression setting, exact distributions for all three of these test statistics can be derived by ignoring the construction of the statistics themselves and focusing instead on the relationships between the distributions of the three statistics and the F distribution. McDonald and Xu (1995a) develop a parametric family of generalized beta distributions for use in econometric applications. The F distribution is a member of this family, and the distributions of all three test statistics can be expressed as special cases of the generalized beta family via their relation to the F distribution. The generalized beta family is closely related to the beta distribution, and quantiles of the beta distribution can easily be transformed to give corresponding quantiles of the distributions of all three test statistics, for any sample size, number of parameters, or rank of restrictions.

2 Definitions and relationships

Suppose we have the classical linear model $y = x'\beta + \varepsilon$ based a sample of n observations, $\beta \in \mathbb{R}^k$ and $\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$. Suppose furthermore that we would like to conduct a test of the linear restrictions

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

where the rank of R is r. Then 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:

$$W = (R\hat{\beta} - \delta)'\big(RI(\hat{\beta})^{-1}R'\big)^{-1}(R\hat{\beta} - \delta) \tag{2}$$

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

$$LM = d(\tilde{\beta})'I(\tilde{\beta})^{-1}d(\tilde{\beta}) \tag{4}$$

where *W, LR* and *LM* stand for the Wald, likelihood ratio and Lagrange multiplier test statistics respectively, and $\ell(\cdot), I(\cdot)$, and $d(\cdot)$ are the likelihood, information matrix and score evaluated at a certain parameter value. It can be shown that these three statistics converge (under the null hypothesis) to a χ^2_r random variable, where r is the rank of R, although they have different distributions in finite samples.

These three statistics are related to the F distribution in small samples. One way to illustrate the relationship between the F and Wald statistics is through a simple example in which the rank of R is 1. To test the hypothesis of a single linear restriction, for example a test of the location parameter $H_0 : \mu = \mu_0$ against a two sided alternative, under the null hypothesis (using the mle $\hat{\sigma}^2$), $\frac{\bar{x}-\mu_0}{\sqrt{\hat{\sigma}^2/(n-1)}}$ has a t_{n-1} distribution. As can be seen,

$$t^2 = (\bar{x} - \mu_0)'\big(\hat{\sigma}^2/(n-1)\big)^{-1}(\bar{x} - \mu_0) = \frac{n-1}{n}W \tag{5}$$

because in this case,

$$W = (\bar{x} - \mu_0)' \left(\hat{\sigma}^2/n\right)^{-1} (\bar{x} - \mu_0). \tag{6}$$

Because of the identity $t_{n-1}^2 = F_{1,n-1}$, we have the following equality:

$$W = \frac{n}{n-1} F_{1,n-1}. \tag{7}$$

This relationship, and the relationship between the F, likelihood ratio and Lagrange multiplier test statistics can be generalized to a rank k matrix of independent variables and rank $r \leq k$ null restrictions. The general F statistic with r and $n-k$ degrees of freedom has density function defined by

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

where B is the beta function. For general n, k and r, we have the following relationships (writing F, W, LR and LM as the random variables that take the value of the various test statistics.) The exact relation between the Wald and $F_{r,n-k}$ statistics is (Engle, 1984, p. 788 for example)

$$W = \frac{nr}{n-k} F_{r,n-k}. \tag{9}$$

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

$$LR = n \log\left(1 + \frac{W}{n}\right) = n \log\left(1 + \frac{r}{n-k} F_{r,n-k}\right). \tag{10}$$

Finally,

$$LM = \frac{W}{1 + \frac{W}{n}} = \frac{\frac{nr}{n-k} F_{r,n-k}}{1 + \frac{r}{n-k} F_{r,n-k}}. \tag{11}$$

The existence of the above relationships is not new, but these links allow one to express the distributions of all of the test statistics as special cases of one parent model, as will be shown below.

3 Test statistics and their densities

Using the relationships described in Section 2, the following basic formula from calculus will be used repeatedly. If random variable X has density $f_X(x)$ and $Y = G(X)$ (G a nonstochastic function), then the density of Y may be expressed as

$$f_Y(y) = f_X\left(G^{-1}(y)\right)\left|\frac{\mathrm{d}G^{-1}(y)}{\mathrm{d}y}\right|. \tag{12}$$

For example, calling $f_F(\cdot;r,n-k)$ the density function of the $F_{r,n-k}$ statistic and $f_W(\cdot;r,n-k)$ the density function of the Wald statistic (a convention that will be repeated in the sequel), we use the function $W = G(F)$ from equation (9), which is linear and therefore invertible:

$$f_W(x;r,n-k) = f_F\left(\frac{n-k}{nr}x\right)\frac{n-k}{nr} \tag{13}$$

where W is the value of the Wald statistic. A little algebra using (8) reveals that this can be compactly expressed as

$$f_W(x;r,n-k) = \frac{(\frac{x}{n})^{\frac{r}{2}-1}}{nB\left(\frac{r}{2},\frac{n-k}{2}\right)(1+\frac{x}{n})^{\frac{r+n-k}{2}}}, \qquad x \in [0,+\infty). \tag{14}$$

The same technique (that is, the use of (8), (10) and (12)) shows that the exact density of the likelihood ratio statistic is

$$f_{LR}(x;r,n-k) = \frac{e^{\frac{x}{n}}\left(e^{\frac{x}{n}}-1\right)^{\frac{r}{2}-1}}{nB\left(\frac{n-k}{2},\frac{r}{2}\right)\left(e^{\frac{x}{n}}\right)^{\frac{r+n-k}{2}}}, \qquad x \in [0,+\infty). \tag{15}$$

This transformation is possible because of the monotonicity of the relationship in equation (10).

Finally, the density of the Lagrange multiplier statistic can be derived similarly:

$$LM = \frac{\frac{nr}{n-k}F}{1+\frac{r}{n-k}F} \qquad \Longrightarrow \qquad F = \frac{LM}{\frac{nr}{n-k}-\frac{r}{n-k}LM} \tag{16}$$

implies that the Jacobian is

$$|J| = \frac{\frac{nr}{n-k}-\frac{r}{n-k}LM+\frac{r}{n-k}LM}{\left(\frac{nr}{n-k}-\frac{r}{n-k}LM\right)^2} = \frac{n}{\frac{r}{n-k}(n-LM)^2}. \tag{17}$$

This case is different than the previous two; it can be seen above that the domain of this relationship is

$[0,n)$ and not $[0,\infty)$. Performing the same process as above for the other two statistics, we have

$$f_{LM}(x;r,n-k) = \frac{\left(\frac{r}{n-k}\right)^{\frac{r}{2}} \left(\frac{x}{\frac{nr}{n-k}-\frac{r}{n-k}x}\right)^{\frac{r}{2}-1}}{B\left(\frac{r}{2},\frac{n-k}{2}\right)\left(1+\frac{r}{n-k}\frac{x}{\frac{nr}{n-k}-\frac{r}{n-k}x}\right)^{\frac{r+n-k}{2}}} \cdot \frac{n}{\frac{r}{n-k}(n-x)^2} \tag{18}$$

$$= \frac{\left(\frac{x}{n-x}\right)^{\frac{r}{2}-1}}{B\left(\frac{r}{2},\frac{n-k}{2}\right)\left(\frac{n}{n-x}\right)^{\frac{r+n-k}{2}}} \cdot \frac{n}{(n-x)^2} \tag{19}$$

and some rearrangement implies the exact density function for the Lagrange multiplier statistic is

$$f_{LM}(x;r,n-k) = \frac{\frac{x}{n}^{\frac{r}{2}-1}}{nB\left(\frac{r}{2},\frac{n-k}{2}\right)(1-\frac{x}{n})^{\frac{2-n+k}{2}}}, \qquad x \in [0,n). \tag{20}$$

4 The generalized beta distribution

The densities derived in the previous Section are all closely related. This section identifies those distributions as special cases of a generalized beta distribution, introduced to the econometrics literature by McDonald (1984) and generalized by McDonald and Xu (1995a,b).

4.1 The generalized beta distribution

McDonald (1984) defines two distributions for applications to the analysis of national income distributions. This work is generalized in McDonald and Xu (1995a) and McDonald and Xu (1995b) to a single distribution that nests both cases from McDonald (1984); this single distribution will be referred to as the generalized beta distribution. As will be seen below, all three test statistics may be expressed as special cases of this distribution. The "usual" beta distribution, for reference, is the density

$$f_B(x;p,q) = \frac{x^{p-1}(1-x)^{q-1}}{B(p,q)}, \quad 0 < x < 1 \tag{21}$$

for $p,q \in [0,\infty)$. The generalized beta distribution is similar, but defined with five parameters. Its density is defined by

$$f_{GB}(x;a,b,c,p,q) = \frac{|a|x^{ap-1}(1-(1-c)(x/b)^a)^{q-1}}{b^{ap}B(p,q)(1+c(x/b)^a)^{p+q}}, \quad 0 < x^a < \frac{b^a}{(1-c)} \tag{22}$$

for $a \in \mathbb{R},\ b,p,q \in [0,\infty)$ and $c \in [0,1]$. This model is extremely flexible and nests many well-known models such as the gamma, F, t, χ^2, Pareto, Weibull, exponential and normal distributions. The usual beta distribution satisfies $f_B(x;p,q) = f_{GB}(x;a=1,b=1,c=0,p,q)$. Below it will be shown that the distributions of the three test statistics described above are special cases of this model. It is unsurprising that the distributions of the test statistics are special cases of this model; they are distributions that are closely related to the F distribution and that converge to the χ^2 distribution, two other special cases of the model.

4.2 Wald and Lagrange multiplier statistics

It can be seen that by setting $a = 1, b = n, c = 1, p = \frac{r}{2}$ and $q = \frac{n-k}{2}$ in expression (22), one obtains

$$f_{GB}\left(x; a = 1, b = n, c = 1, p = \frac{r}{2}, q = \frac{n-k}{2}\right) = \frac{x^{\frac{r}{2}-1}}{n^{\frac{r}{2}} B(\frac{r}{2}, \frac{n-k}{2})(1 + \frac{x}{n})^{\frac{r+n-k}{2}}} \tag{23}$$

which is the density of the Wald statistic given in equation (14).

Similarly, the density of the Lagrange multiplier statistic can be expressed in terms of the generalized beta distribution. By letting $a = 1, b = n, c = 0, p = \frac{r}{2}$, and $q = \frac{n-k}{2}$, we obtain

$$f_{GB}\left(x; a = 1, b = n, c = 0, p = \frac{r}{2}, q = \frac{n-k}{2}\right) = \frac{\left(\frac{x}{n}\right)^{\frac{r}{2}-1}\left(1 - \frac{x}{n}\right)^{\frac{n-k}{2}-1}}{nB(\frac{r}{2}, \frac{n-k}{2})} \tag{24}$$

which is identical to equation (20). Here it becomes evident that the only difference between the Wald and Lagrange multiplier statistics, in terms of the parameters of the generalized beta model, is the value of c.

4.3 Likelihood ratio

McDonald and Xu (1995a) also define a variant of the generalized beta distribution given in equation (22) via a simple transformation. If Y is distributed with a generalized beta distribution and $X = \ln(Y)$, then the distribution of X is called the exponential generalized beta distribution, and it is defined via the generalized beta:

$$f_{EGB}(x; \delta, \sigma, c, p, q) = f_{GB}(e^x; a = 1/\sigma, b = e^\delta, c, p, q) \cdot e^x \tag{25}$$

or more explicitly

$$f_{EGB}(x; \delta, \sigma, c, p, q) = \frac{e^{p(x-\delta)/\sigma}(1 - (1-c)e^{(x-\delta)/\sigma})^{q-1}}{|\sigma| B(p, q)(1 + ce^{(x-\delta)/\sigma})^{p+q}}, \quad -\infty < \frac{x-\delta}{\sigma} < \ln\left(\frac{1}{1-c}\right). \tag{26}$$

The likelihood ratio statistic follows an exponential generalized beta distribution. First, restrict the parameters of the generalized beta distribution as follows: let $\sigma = n,\ \delta = 0,\ c = 0,\ p = \frac{n-k}{2}$ and $q = \frac{r}{2}$. Then x should be replaced by $-x$ to obtain a statistic with the desired domain. We then have the following:

$$f_{EGB}\left(-x;\delta=0,\sigma=n,c=0,p=\frac{n-k}{2},q=\frac{r}{2}\right)=\frac{\left(e^{\frac{-x}{n}}\right)^{\frac{n-k}{2}}\left(1-e^{\frac{-x}{n}}\right)^{\frac{r}{2}-1}}{nB\left(\frac{n-k}{2},\frac{r}{2}\right)}. \tag{27}$$

Multiply top and bottom by $\left(e^{x/n}\right)^{\frac{r}{2}-1}$:

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

and it can be seen that this is equal to equation (15).

5 Quantiles via the beta distribution

To summarize the previous Section, the densities of the $F_{r,n-k}$, Wald, likelihood ratio, Rao score densities can be expressed as cases of the generalized beta density given n, r and k:

$$\begin{aligned}
f_F(x;r,n-k) &= f_{GB}\left(x;a=1,b=\frac{n-k}{r},c=1,p=\frac{r}{2},q=\frac{n-k}{2}\right)\\
f_W(x;r,n-k) &= f_{GB}\left(x;a=1,b=n,c=1,p=\frac{r}{2},q=\frac{n-k}{2}\right)\\
f_{LR}(x;r,n-k) &= f_{GB}\left(e^{-x};a=\frac{1}{n},b=1,c=0,p=\frac{n-k}{2},q=\frac{r}{2}\right)\cdot e^{-x}\\
f_{LM}(x;r,n-k) &= f_{GB}\left(x;a=1,b=n,c=0,p=\frac{r}{2},q=\frac{n-k}{2}\right).
\end{aligned}$$

Quite conveniently, as generalized beta distributions, their cumulative distributions, and therefore critical values used for testing, can easily be found. McDonald and Xu (1995a) note that if $x \sim GB(a,b,c,p,q)$, then the distribution of the transformation of x defined by

$$y=(x/b)^a/(1+c(x/b)^a) \tag{29}$$

follows the usual beta distribution[1]. This makes the calculation of exact quantiles very easy.

The easiest quantiles to obtain are the Lagrange multiplier quantiles because $c=0$ in its formulation as a generalized beta variable; letting $B_{1-\alpha}$ be the $(1-\alpha)$th quantile of the usual $B(\frac{r}{2},\frac{n-k}{2})$ distribution, the transform $B_{1-\alpha}\mapsto nB_{1-\alpha}$ is extremely simple — this statistic is simply a rescaled version of the usual beta distribution.

Quantiles for the Wald statistic can be determined using the relationship $y=(x/n)/(1+(x/n))=x/(n+x)$, which is distributed as a beta distribution. That is, if $B_{1-\alpha}$ is the $(1-\alpha)$th quantile of the

[1] This is provable using formulas (21), (22), (12) and lots of algebra.

$B(\frac{r}{2},\frac{n-k}{2})$ distribution, the same quantile of the Wald statistic is found through the transform $B_{1-\alpha} \mapsto nB_{1-\alpha}/(1-B_{1-\alpha})$.

The likelihood ratio is only slightly more complex — as McDonald and Xu (1995a, Appendix A.2.2) point out, if $X \sim F_{EGB}$ and $X = \ln(Y)$, then $F_{EGB}(x) = \mathrm{P}(X \leq x) = \mathrm{P}(e^Y \leq e^y) = F_{GB}(e^y)$. However, this is complicated by the domain issue noted above. The transformation $y = (e^x)^{\frac{1}{n}}$ implies that the $(1-\alpha)$th quantile of the likelihood ratio statistic is equal to $-n\ln\left(B_\alpha\right)$, where B_α is the αth quantile of the $B(\frac{n-k}{2},\frac{r}{2})$ distribution. Using the identity $B_\alpha(p,q) = 1 - B_{1-\alpha}(q,p)$, we arrive at the mapping $B_{1-\alpha} \mapsto -n\log(1-B_{1-\alpha})$.

With an appeal to a basic inequality for exponents or logarithms,[2] it is immediately clear that quantiles of the three distributions satisfy the relationship $LM \leq LR \leq W$, which is noted, for example, in Engle (1984, p. 792).

In summary, the quantiles of these distributions can be expressed through mappings from quantiles of the beta$(r/2,(n-k)/2)$ distribution: letting $B \equiv B_{1-\alpha}(\frac{r}{2},\frac{n-k}{2})$, the $(1-\alpha)^{\text{th}}$ quantile of the distribution of each test statistic is given by

Wald:	$\frac{nB}{1-B}$
likelihood ratio:	$-n\log(1-B)$
Lagrange multiplier:	nB

6 The effect of parameter values visually

The effect of the parameters n and k can be seen in Figures 1, 2 and 3. A decrease in n or an increase in k or 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 often enter the density formulas as a difference, a decrease in n is effectively the same as an increase in k in most cases. However, the effect of a change in r is slightly different and somewhat more pronounced than effects from changes in the other two parameters, effectively because the value of r is often small.

The exact quantiles for each of the distributions are, by and large, greater than the asymptotic value. However, n and k do not always enter the parameterizations together. This means that changes in n and k are not entirely analogous to one another, even if quantiles can be derived from a beta distribution that only depends on r and $n-k$. In particular, as k becomes small — equal to or just greater than r — the upper quantile of the likelihood ratio statistic can be less than the χ^2 approximation. This is illustrated in the top two panels of Figure 2. Despite these minor complications, the quantile inequality between the test statistics is plain to see in the figures, represented by the relative placement of the theoretical critical values on each graph.

In particular, it can be seen that the likelihood ratio statistic does not suffer very greatly from an asymptotic χ^2 approximation, but the other two test statistics, particularly the Wald, do. In fact,

[2]That is, since

$$nB \leq -n\log(1-B) \leq \frac{nB}{1-B}, \qquad 0 < B < 1$$

see, for example Davidson and MacKinnon (1993, p. 456-457).

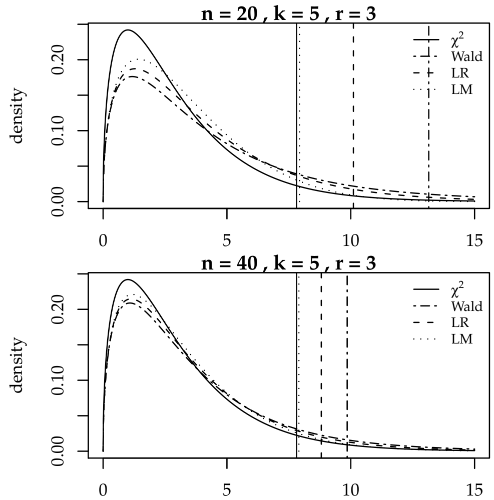

Figure 1: The effect of n on densities and critical values: the 95th percentile of the likelihood ratio statistic is uniformly closer to the asymptotic χ^2 value than the other two statistics. However, this is for a moderately large value of k, and this can change for smaller values of k. The Wald statistic will be rejected more often than it should be in small samples; more so than the other two statistics.

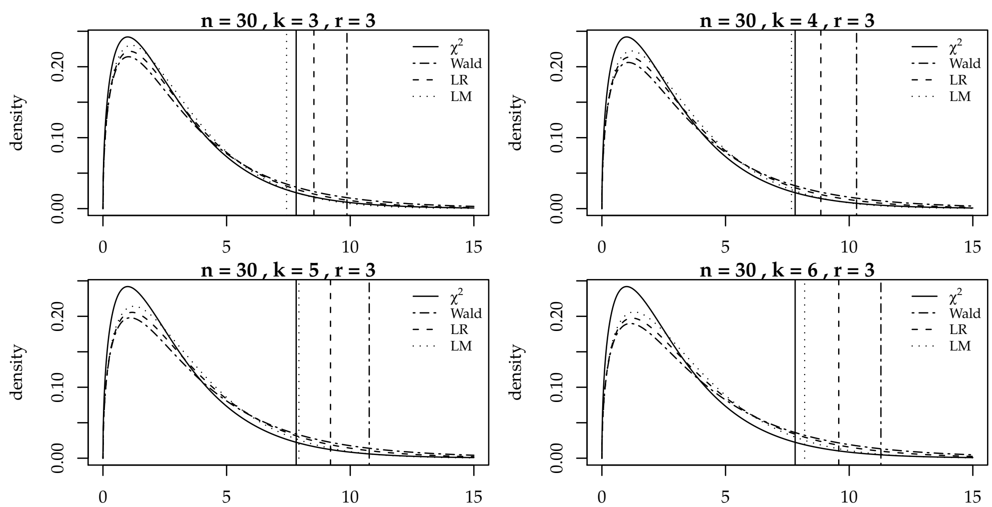

Figure 2: The effect of k on densities and critical values shows roughly the same pattern as that in Figure 1, because n and k often enter as a difference in the parameterization of the exact distributions. The apparent changes across panels are smaller than in Figure 1 because k is typically much smaller than n, and the parameter values have been chosen to reflect this. Very small values of k relative to r can lead to likelihood ratio distributions that have smaller upper quantiles than the asymptotic approximation, which is not usually the case for any other arrangements of parameter values.

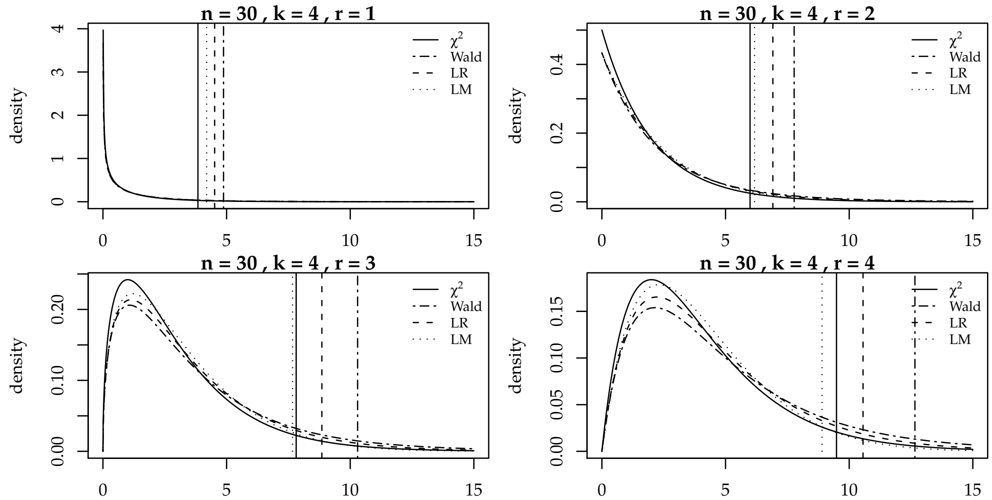

Figure 3: An increase in the number of restrictions in a test, r, has a more pronounced effect than from a decrease in n or an increase in k. This is in part because r is much smaller than $n-k$ in most settings, and in part because a change in r when its value is small induces such a drastic change in the shape of the distributions.

Figure 3 reveals that the difference between the asymptotic and the true likelihood ratio quantiles can be inversely proportional to the number of restrictions, unlike the Wald and Lagrange multiplier tests, which behave more as is intuitively expected. Figures 1 and 2 also reveal that the asymptotic approximation is particularly harmful to small the Wald statistic in situations where $n-k$ is small — see, for example, the top left panel of Figure 1 or the bottom right panel of Figures 2 and 3. 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 tests in particular.

7 Conclusion

The relationship $LR \leq LR \leq W$ is well known, but the generalized beta distribution provides a unified distributional characterization of these test statistics in the simple linear regression setting. For cases in which the asymptotic χ^2 approximation is a poor one, for example when the sample size is small or the number of restrictions is large, this distribution also provides more appropriate critical values for tests.

References

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

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

R. Davidson and J. 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.

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

J.B. McDonald and Y.J. Xu. A generalization of the beta distribution with applications. *Journal of Econometrics*, 69:133–152, 1995a.

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