Testing Special Cases of the GB2 Distribution

William E. Griffiths
University of Melbourne
R. Carter Hill
Louisiana State University

Abstract

We examine the power of hypothesis tests leading to four widely used special cases of the four-parameter generalized beta distribution of the second kind (GB2). These are the Singh-Maddala, Dagum, beta-2 and Fisk distributions. For the Singh-Maddala, Dagum, and Fisk distributions, the power of Wald tests is nonmonotonic when the true parameter values are greater than those in the null hypothesis. As the difference between the hypothesized values of the parameters and their actual values increases, the power increases and then declines. Wald tests on log-transformations of the parameters are more powerful than those on the original parameters; their power functions are also nonmonotonic but the decline in power begins at higher values of the parameters. The likelihood ratio and Lagrange multiplier tests have power functions that increase monotonically for true parameter values beyond the hypothesized values, and are more powerful than their Wald counterparts, but only marginally so relative to the Wald test on the log-transformed parameters. For parameter values less than the hypothesized values there are no large differences in the power of all four tests. Tests for the remaining distribution, the beta-2 distribution, have power functions that are well behaved and exhibit only small differences. We provide an extensive online supplement containing several theoretical results as well as further insights into the power behavior of the tests.

KEY WORDS: Wald test; Likelihood ratio test; nonmonotonic power; log-transformed parameters.

Corresponding author:
William Griffiths
Department of Economics
University of Melbourne
wegrif@unimelb.edu.au

Testing Special Cases of the GB2 Distribution

William E. Griffiths
University of Melbourne
R. Carter Hill
Louisiana State University

1. Introduction

We consider the generalized beta distribution of the second kind (GB2) for a random variable $x > 0$ with probability density function (pdf)

$$f(x) = \frac{ax^{ap-1}}{b^{ap}B(p,q)\left[1+\left(x/b\right)^{a}\right]^{p+q}} \tag{1}$$

where $\boldsymbol{\theta} = \left(a\ b\ p\ q\right)'$ are positive parameters, and $B(p,q)$ is the beta function. We are particularly interested in the properties of test statistics designed to test the special cases $p=1$ (a Singh-Maddala distribution), $q=1$ (a Dagum distribution), $a=1$ (a beta-2 or B2 distribution), and $p=q=1$ (a Fisk distribution). The standard test statistics used for this purpose are the Wald, Lagrange multiplier (LM) and likelihood ratio (LR) tests. Although test statistics for these tests all have the same limiting (large sample) distribution when the null hypothesis is true, their distributions under the alternative hypothesis do not coincide, except in special and artificial scenarios when only local alternatives are considered. Thus, the performance of the three tests, evaluated in terms of their power, can vary considerably. We carried out a thorough investigation of the power of the Wald, LM and LR tests for testing the null hypotheses (*i*) $H_0: p=1$, (*ii*) $H_0: q=1$, (*iii*) $H_0: a=1$ and (*iv*) $H_0: p=q=1$. Because it turned out that the power properties of the LM and LR statistics were similar, and the LM test statistic is not automatically computed by standard statistical software, in the paper we focus on the LR and Wald tests leaving discussion of the LM test to supplementary material. Results for testing $H_0: a=1$ are also left for the supplementary material because, unlike the results for the other hypotheses, in this instance all three tests had similar and conventional power properties. Focusing on the LR and Wald tests for the Singh-Maddala, Dagum and Fisk distributions, our study reveals several interesting features. Five general conclusions are:

(1) The LR test is superior to the Wald test.

(2) The powers of Wald tests for log-transformations of the parameters p and q, that is, $H_0: \ln(p) = 0$, $H_0: \ln(q) = 0$, and $H_0: \ln(p) = \ln(q) = 0$, are substantially different from the powers of the Wald tests for testing that the original parameters are equal to one. For parameter values less than one, the Wald tests on the original parameters have slightly greater power; for parameter values greater than one, Wald tests on the log-transformations have greater power.

(3) For parameter values greater than one, the power functions for Wald tests for p and q do not increase monotonically. For $H_0: p = 1$ and $H_0: q = 1$, they increase up to a point, and then decline. For $H_0: \ln(p) = 0$ and $H_0: \ln(q) = 0$ the downturn occurs at a larger parameter value and is less pronounced.

(4) The power function of the Wald test for $H_0: p = q = 1$ is particularly poorly behaved.

(5) The power functions from assuming asymptotic local alternatives are poor approximations of actual power when testing $H_0: p = 1$ and $H_0: q = 1$.

Computationally, using a log-transformation of the parameters has some significant numerical advantages and is a common way of ensuring nonnegative parameter estimates.

In Section 2 we describe the background to the GB2 distribution and other studies where nonmonotonic power was investigated. We develop notation in Section 3, and we describe some general results. We consider the tests and test results for each of the Singh-Maddala, Dagum, and Fisk cases in Sections 4, 5, and 6, respectively. Section 7 contains concluding remarks.

2. Background

The GB2 distribution and its special cases have attracted a great deal of attention in the literature and have been widely used for modelling income distributions. The Singh-Maddala, Dagum, beta-2 and Fisk distributions were proposed in Singh and Maddala (1976), Dagum (1977), Vartia and Vartia (1980) and Fisk (1961), respectively.[1] The popularity of the GB2 distribution can be attributed to a series of

[1] We consider the "Dagum 1" distribution. Other variants proposed by Dagum are discussed in Kleiber (2008).

papers by McDonald and his co-authors. McDonald (1984) shows how several well-known distributions, namely, beta of the first and second kinds, Singh-Maddala, lognormal, gamma, generalized gamma, Fisk and exponential, can be viewed as special cases of the four-parameter distributions GB1 and GB2.[2] A five-parameter distribution that nests both the GB1 and GB2 and over 30 distributions is introduced in McDonald and Xu (1995). The Dagum distribution, overlooked in the 1984 paper, is recognized as one of the special cases, but by its other name, a type 3 Burr distribution. Properties of the GB2 relevant for measuring inequality are considered in Butler and McDonald (1989), McDonald and Ransom (2008), McDonald et al. (2011), Jenkins (2009) and Sarabia, et al. (2017). A survey of GB2 estimation methodology and applications, and details of how GB2 parameters can be used to estimate inequality, poverty and pro-poor growth appears in Chotikapanich et al. (2018). Kleiber and Kotz (2003) provide a detailed description of the GB2 and its special cases.

An undesirable property that we uncover is the nonmonotonic power of the Wald tests for three of the special cases. The source of the nonmonotonicity is like that described by Nelson and Savin (1990) for a simple exponential regression model. As the deviation of the parameter value from its hypothesized value increases, the variance of the parameter estimator also increases such that the denominator in the Wald statistic can increase at a greater rate than its numerator. Other circumstances where nonmonotonic power has been encountered include logit, probit, and Tobit models (Nelson and Savin 1988, Savin and Würtz 1999), testing for structural change in time series (Vogelsang 1997, 1999, Crainiceanu and Vogelsang 2007) and multiple testing (Cao et al. 2013).

Two other GB2 special cases where testing is likely to be of interest, but is not straightforward, are for the generalized gamma distribution and the lognormal distribution. In these cases, the violation of a boundary condition means the likelihood ratio test statistic does not have a chi-square distribution under the null hypothesis. See McDonald and Xu (1992) for a discussion and a Monte Carlo experiment.

[2] The GB1 (generalized beta of the first kind) distribution is an alternative to the GB2 with a finite support.

3. Preliminaries

We assume we have an independent random sample $\mathbf{x} = (x_1 \dots x_N)'$ of N observations taken from the GB2 distribution or one of its special cases. Using the GB2 density $f(x_i)$ in (1), the log-likelihood for a single observation x_i is $\ln L_i = \ln f(x_i)$ and the log-likelihood for all observations is $\ln L = \sum_{i=1}^{N} \ln L_i$. Let the maximum likelihood estimator (MLE) of the parameters $\boldsymbol{\theta} = (a\ b\ p\ q)'$ be $\hat{\boldsymbol{\theta}} = (\hat{a}\ \hat{b}\ \hat{p}\ \hat{q})'$. The MLE is consistent and asymptotically normally distributed under some regularity conditions[3]. We denote the asymptotic distribution as $\hat{\boldsymbol{\theta}} \overset{a}{\sim} N\left[\boldsymbol{\theta}, \operatorname{var}(\hat{\boldsymbol{\theta}})\right]$, where $\operatorname{var}(\hat{\boldsymbol{\theta}}) = \mathfrak{I}^{-1}/N$ and $\mathfrak{I}$ is the information matrix for a single observation.[4] We refer to the covariance matrix of this distribution, $\operatorname{var}(\hat{\boldsymbol{\theta}})$, as the asymptotic variance of $\hat{\boldsymbol{\theta}}$. We use the estimator based on the matrix of second derivatives of the log-likelihood function, the Hessian[5]. That is

$$\widehat{\operatorname{var}}(\hat{\boldsymbol{\theta}}) = -\left(\frac{1}{N}\sum_{i=1}^{N} \partial^2 \ln L_i / \partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'\right)^{-1}\Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \tag{2}$$

3.1 Wald Tests

For testing J linear hypotheses $H_0 : \mathbf{R}\boldsymbol{\theta} = \mathbf{r}$, the Wald test statistic is

$$W = \left(\mathbf{R}\hat{\boldsymbol{\theta}} - \mathbf{r}\right)' \left(\mathbf{R}\widehat{\operatorname{var}}(\hat{\boldsymbol{\theta}})\mathbf{R}'\right)^{-1} \left(\mathbf{R}\hat{\boldsymbol{\theta}} - \mathbf{r}\right) \tag{3}$$

This statistic has a limiting $\chi^2_{(J)}$-distribution when H_0 is true.

Maximum likelihood estimates are invariant under one-to-one transformations of the parameters, but Wald tests are not. One transformation that turns out to be convenient both computationally and because of the power properties of the resulting Wald tests is a log-transformation. That is, in addition to Wald tests for $\boldsymbol{\theta}$, we also consider Wald tests for $\boldsymbol{\phi} = (\ln(a)\ \ln(b)\ \ln(p)\ \ln(q))'$. The hypotheses of

---

[3] Greene (2018, 542)

[4] We give the details of the information matrix and the derivatives of the log-likelihood in the supplement.

[5] See Greene (2018, 549-550) for a discussion of the common estimators of the MLE covariance matrix. Which one that we use matters little in our examples.

interest are $H_0:\ln(p)=0$, $H_0:\ln(q)=0$, and $H_0:\ln(p)=\ln(q)=0$ [6]. Using the delta method, the Wald statistic in (3) is

$$W=\left(\mathbf{R}\hat{\boldsymbol{\phi}}\right)'\left(\mathbf{R}\hat{\mathbf{D}}\widehat{\operatorname{var}}\left(\hat{\boldsymbol{\theta}}\right)\hat{\mathbf{D}}\mathbf{R}'\right)^{-1}\mathbf{R}\hat{\boldsymbol{\phi}} \tag{4}$$

where $\hat{\mathbf{D}}=\operatorname{diag}\left(\hat{a}^{-1}\ \hat{b}^{-1}\ \hat{p}^{-1}\ \hat{q}^{-1}\right)$.

3.2 Likelihood Ratio Test

The likelihood ratio test statistic, based on the difference between the restricted and unrestricted maximized log-likelihood functions, is

$$\mathrm{LR}=2\left[\sum_{i=1}^{N}\ln L_i\bigg|_{\hat{\boldsymbol{\theta}}}-\sum_{i=1}^{N}\ln L_i\bigg|_{\tilde{\boldsymbol{\theta}}}\right] \tag{5}$$

The estimator $\tilde{\boldsymbol{\theta}}$ is the MLE obtained assuming the null hypothesis is true. Like the Wald statistic, the LR statistic has a limiting $\chi^2_{(J)}$ distribution when the null hypothesis is true. This test is invariant with respect to the log-transformation. For many models the likelihood ratio test is the best alternative and the same will be true in the GB2 model, as we will show.[7]

3.3 Test Power Considerations

To evaluate test power, the probability of rejecting a false hypothesis, we need the distributions of the test statistics under the alternative hypothesis $H_1:\mathbf{R}\boldsymbol{\theta}\neq\mathbf{r}$. Although the test statistics have the same limiting distribution under H_0, they do not have the same limiting distribution under H_1, except for the artificial special case of local alternatives, where H_1 is written $H_1:\mathbf{R}\boldsymbol{\theta}-\mathbf{r}=\boldsymbol{\delta}/\sqrt{N}$ [8]. In this case the limiting distribution of the test statistics is a noncentral chi-square distribution with noncentrality parameter $\lambda=N\left(\mathbf{R}\boldsymbol{\theta}-\mathbf{r}\right)'\left(\mathbf{R}\mathfrak{I}^{-1}\mathbf{R}'\right)^{-1}\left(\mathbf{R}\boldsymbol{\theta}-\mathbf{r}\right)$. Because the specification of local alternatives leads to

---

[6] Jenkins (2014) developed a Stata (www.stata.com) module for maximizing the reparameterized log-likelihood function. The Stata module is GB2LFIT. The reparameterized log-likelihood function is in the supplement to this paper. Another Stata module has been developed by McDonald and Triplett (2024). It can be used to estimate the GB2 distribution under a variety of scenarios including censored, interval and grouped data and regression models.

[7] The other common test when using maximum likelihood estimation is the Lagrange multiplier (LM) test, also known as the Score test. In practice this test is not likely to be used in the GB2 context and its performance is like that of the likelihood ratio test. The paper supplement includes a full discussion of the LM test.

[8] See, for example, Ruud (2000, p.404) or Bickel and Doksum (2015, p.400).

tractable limiting distributions from which power can be readily calculated, many authors have used these power calculations to approximate the power of the test statistics in finite samples.[9] However, in practice, alternative hypotheses cannot be written as $H_1 : \mathbf{R}\boldsymbol{\theta} - \mathbf{r} = \boldsymbol{\delta}/\sqrt{N}$. The discrepancy between $\mathbf{R}\boldsymbol{\theta}$ and $\mathbf{r}$ does not decrease with an increasing sample size and the test statistics do not have the same limiting distribution under the more general alternative hypotheses, $H_1 : \mathbf{R}\boldsymbol{\theta} \neq \mathbf{r}$. Furthermore, as noted by Nelson and Savin (1988, 1990), a local power approximation can be poor in finite samples, particularly in circumstances where the noncentrality parameter is nonmonotonic, a situation we encounter here.

We plot histograms of Monte Carlo generated test statistic values under both the null and alternative hypotheses in the supplementary material (Section S3). For values computed when the null hypotheses are true the histograms agree well with central chi-square distributions. However, for values computed under the alternative, we find (1) the noncentral chi-square distribution implied by an asymptotic local power assumption is a poor approximation of the distribution of the test statistic values, and (2) the test statistics have quite different distributions under the alternative, even for very large sample sizes.

Having different distributions of the test statistics under H_1, as well as an interest in their relative finite-sample power, means Monte Carlo experiments are useful for assessing their power. We will consider the results of these experiments in the next sections. We also compute asymptotic local power, which we refer to as "hypothetical power."

4. Testing for a Singh-Maddala distribution ($p = 1$)

The Wald test statistic for testing $H_0 : p = 1$, obtained by setting $\mathbf{R} = \left(0\,0\,1\,0\right)$ and $\mathbf{r} = (1)$ in equation (3), is given by

$$W_{SM} = \frac{\left(\hat{p} - 1\right)^2}{\widehat{\operatorname{var}}\left(\hat{p}\right)} \tag{6}$$

[9] See McManus (1991) for a discussion of the origins of local power analysis, and references to examples.

The Wald statistic for testing $H_0: \ln(p) = 0$ uses the asymptotic variance of $\hat{\gamma} = \ln(\hat{p})$, which is $\text{var}(\hat{\gamma}) = \text{var}(\hat{p})/p^2$, and the statistic in equation (6) becomes

$$W_{SM}^{\log} = \frac{[\hat{p}\ln(\hat{p})]^2}{\widehat{\text{var}}(\hat{p})} \tag{7}$$

The LR test statistic for testing $H_0: p = 1$ use the restricted estimator $\tilde{\boldsymbol{\theta}}' = (\tilde{a}\;\tilde{b}\;1\;\tilde{q})$.

4.1 Monte Carlo experiments for testing *p*

To estimate the finite sample power when testing $p = 1$, and $\ln(p) = 0$, we carried out a Monte Carlo experiment[10] with 10,000 samples, a sample size of $N = 4000$, the settings $a = 5$, $b = 1$, $q = 1.75$, and values of *p* from 0.5 to 4.0 in increments of 0.1.[11] A sample size of 4000 may seem large for investigating finite sample properties, but it would not be considered a large sample for estimating income distributions and it is small enough to illustrate the properties of the tests. The settings for *a* and *b* have little bearing on the power of the tests because $\text{var}(\hat{p})$ does not depend on the settings for *a* and *b*.[12] The value $q = 1.75$ is a reasonable representation of values that occur empirically.[13] Experimentation suggests larger values of *q* increase power by a small amount when $p > 1$ and have a negligible effect for $p < 1$. With 10,000 samples, the maximum standard error of an estimate of the power (calculated using properties of the binomial distribution) is $\sqrt{0.5^2/10000} = 0.005$. A 5% significance level was used and hence power was estimated as the proportion of test values that exceeded $\chi^2_{(0.95,1)} \cong 3.841$.

[10] We used Gauss Version 24, Aptech Systems, Inc. (2023) www.aptech.com, and the application OPTMUM to conduct the simulations. We created the graphs using Stata 18.0 (StataCorp LLC).

[11] We illustrate the GB2 density functions for these settings in Section S5 of the supplementary material.

[12] See Theorem 1 in Section S4 of the supplementary material.

[13] We consulted the UQICDV3.0 data base (https://uqicd.economics.uq.edu.au/ (Rao et al. 2022)) to see what parameter values are common empirically. They report GB2 distribution estimates for 148 countries in each of 50 years. Within the context explained later in this section, the median estimate for *q* in the UQICD database is 1.52 and the mean estimate is 2.31.

4.2 Power results of Wald tests for *p*

We first consider the power of the Wald tests, $H_0: p = 1$ and $H_0: \ln(p) = 0$, and then in Section 4.3 compare their power with that of the LR test. Wald tests are routinely produced by statistical software, so it is useful to consider their power in some depth. The hypothesis $H_0: p = 1$ is relevant if software maximizes the log-likelihood with respect to the original parameters; the hypothesis $H_0: \ln(p) = 0,$ which is $H_0: \gamma = 0$ in terms of the log-transformed parameters, is relevant if the log-likelihood is maximized with respect to the log-transformed parameters α, β, γ and δ, as in Jenkins (2014). Figure 1 displays the estimated power for both hypotheses along with the hypothetical power predicted by the corresponding asymptotic local approximations. In the following three subsections we consider three properties of the Wald test that are revealed in Figure 1, the relative power of the two tests, the nonmonotonic nature of the Wald test, and the accuracy of the asymptotic local approximation, as well as some computational issues and an illustration of the effect of changing the sample size.

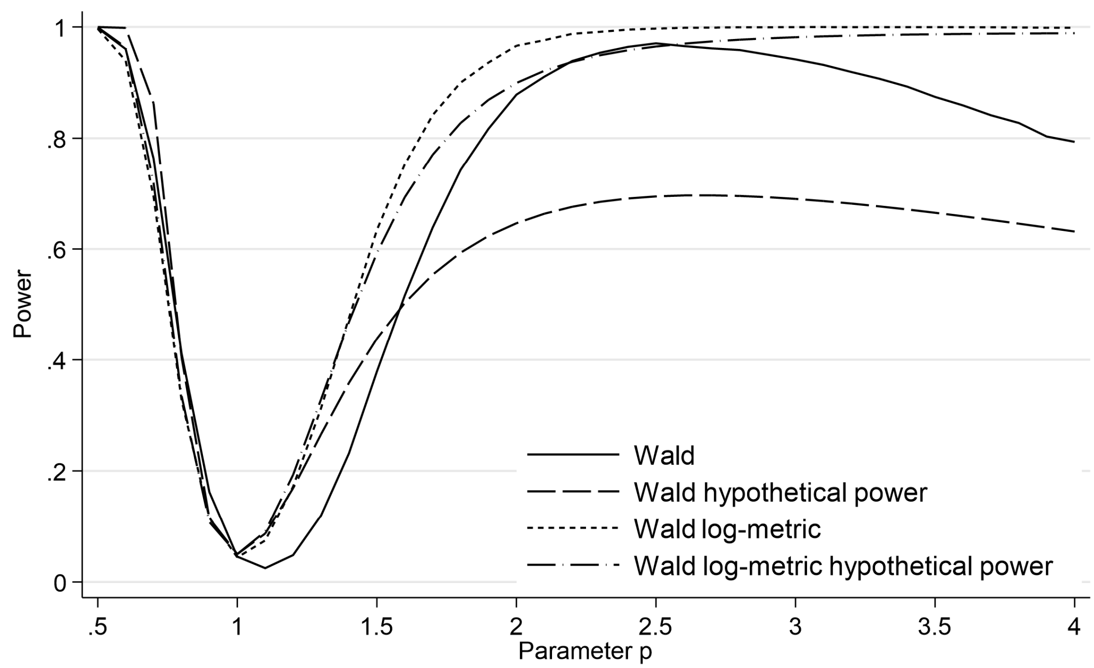


Figure 1 A comparison of the powers of the Wald tests. These are the estimated power functions for the tests $H_0: p = 1$ and $H_0: \ln(p) = 0$. We use a Monte Carlo experiment with 10,000 samples, and N = 4000, a = 5, b = 1 and q = 1.75 to estimate the finite sample powers.

Relative power of the two Wald tests

Testing $H_0 : \ln(p) = 0$ is more powerful than testing $H_0 : p = 1$ when $p > 1$, and less powerful when $p < 1$, but the power difference is very slight in the latter case. This property can be explained by considering the relative magnitudes of the numerators of the two test statistics $W_{SM} = (\hat{p} - 1)^2 \big/ \widehat{\mathrm{var}}(\hat{p})$ and $W_{SM}^{\log} = [\hat{p}\ln(\hat{p})]^2 \big/ \widehat{\mathrm{var}}(\hat{p})$. When $\hat{p} > 1$, $(\hat{p} - 1)^2 > [\hat{p}\ln(\hat{p})]^2$ and $W_{SM}^{\log} > W_{SM}$, implying $W_{SM}^{\log}$ will have greater power. When $\hat{p} < 1$, $(\hat{p} - 1)^2 < [\hat{p}\ln(\hat{p})]^2$ and $W_{SM}^{\log} < W_{SM}$, implying W_{SM} will have greater power.

Lack of power monotonicity

The power of the Wald test for $H_0 : p = 1$ does not increase monotonically for $p > 1$. Also, although it is not evident from Figure 1, it turns out the Wald test for $H_0 : \ln(p) = 0$ has the same property, but the power declines very slowly after its maximum is reached, and the maximum power is at a larger value for *p* than the maximum reached when testing $H_0 : p = 1$. Consider the two Wald statistics, but with the estimates replaced by their true parameter values. That is, $\lambda_{SM} = (p - 1)^2 \big/ \mathrm{var}(\hat{p})$ and $\lambda_{SM}^{\log} = [p\ln(p)]^2 \big/ \mathrm{var}(\hat{p})$. In the supplement we show that as *p* increases the denominator $\mathrm{var}(\hat{p})$ increases at an increasing rate. However, the numerator increases linearly; $d\left[(p-1)^2\right] \big/ dp = 2(p-1)$. The value λ_{SM} increases at first while the rate of change of the numerator is greater than the rate of change of the denominator, then it reaches a maximum when the rates of change are equal, and then it declines when the rate of change in $\mathrm{var}(\hat{p})$ exceeds that of $(p-1)^2$.[14] For the settings in Figure 1, the maximum power is reached at $p = 2.663$ and that is reflected in the power behavior illustrated in the figure. A similar phenomenon is observed for $\lambda_{SM}^{\log} = [p\ln(p)]^2 \big/ \mathrm{var}(\hat{p})$, but in this case the numerator and denominator rates of change are equal at the larger value of $p = 4.356$.

[14] Further details are given in the supplement (Section S6).

Support for using log-transformation

Both its power and numerical advantages suggest estimating the log-transformed parameters is preferable to estimating the original parameters. There is a potential loss in power from using the log transformation when $p<1$ but it is small relative to the potential gain when $p>1$. To determine if values of p greater than one are more prevalent than values less than one in applied studies, we extracted estimates of p from the UQICDV3.0 database (Rao et al., 2023). In all country-year combinations on the database, there were 3,776 unique estimates of the GB2 parameters (a,p,q).[15] Discarding values of p that were greater than 15 left a total of 2,996 estimates for p.[16] They are displayed in the histogram in Figure 4; 71% of them are greater than 1 and 29% are less than 1. The mean is 2.48 and the median is 1.62.

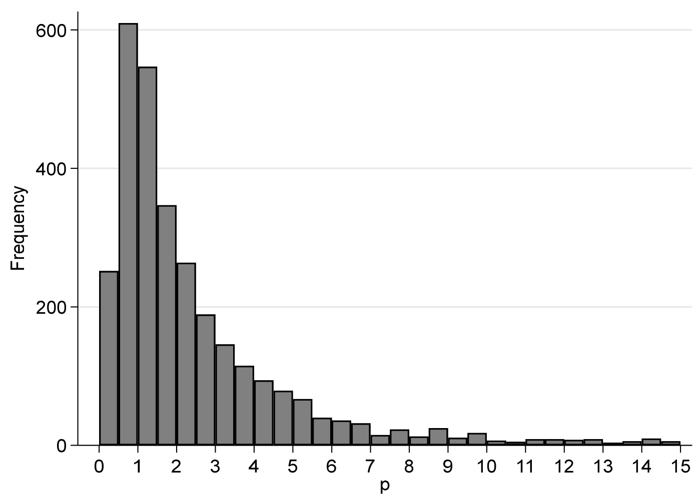


Figure 2. Histogram of subset of unique estimates of *p* from the UQICD database. The subset is those values less than 15.

Accuracy of the asymptotic local approximation

The remaining result from Figure 1 that we consider is the discrepancy between the Monte-Carlo estimated power of the Wald tests and that from the noncentral chi-square distribution which holds under the asymptotic local alternatives assumption. For $p>1$, the finite sample power of the Wald tests

[15] We refer to "unique" estimates because, for a given country, the same income shares, but different mean incomes, are often used to estimate the parameters for different years. When the income shares do not change, estimates for a, p and q do not change. Only the estimate for b changes.

[16] The GB2 distribution approaches an inverted generalized gamma distribution as $p \to \infty$ (Chotikapanich and Griffiths, 2023).

is worse than that predicted by their hypothetical power functions when p is close to 1 and better for larger values of p. Also, the differences are much less pronounced when testing $H_0 : \ln(p) = 0$ than they are when testing $H_0 : p = 1$. For $p < 1$, the Monte-Carlo estimated powers and the hypothetical power functions almost coincide.

To explain these differences, we compare the finite sample distributions of the test statistics with the relevant noncentral chi-square distributions. In Figure 3 we plot histograms of the finite sample distributions estimated from the Monte-Carlo generated test statistics alongside the noncentral chi-square distributions. Two settings of p were used: $p = 1.3$ where the hypothetical power for testing $H_0 : p = 1$ is greater than the finite sample power, and $p = 2.5$ where hypothetical power is less than the finite sample power. We plot comparisons for these two settings in Figures 3(a) and 3(b), respectively. In Figure 3(a) there are very few histogram values beyond 6. On the other hand, the corresponding noncentral chi-square distribution has a long tail to the right, extending well beyond 6 and leading to a higher proportion of values greater than the critical value of 3.841. The opposite phenomenon occurs in Figure 3(b) where $p = 2.5$. In this case, very few histogram values are less than two and the vast majority are greater than the critical value of 3.841. However, the noncentral chi-square distribution includes a substantial tail to the left, leading to a relatively large proportion of values that fail to reject the null hypothesis.

When we make the same comparisons for testing $H_0 : \ln(p) = 0$, we find, from Figure 3(c), that the noncentral chi-square distribution is a good approximation to the histogram when $p = 1.3$. This outcome is consistent with what we observe in Figure 1 where the two power functions are almost identical for $1 \le p \le 1.5$. In Figure 3(d) where $p = 2.5$, the noncentral chi-square distribution is no longer a good approximation to the histogram. The proportion of test values less than 3.841 is less than that predicted by the noncentral chi-square distribution and hence the finite sample power is greater than the hypothetical power. The discrepancy is not as great as that which occurs when testing $H_0 : p = 1$, however, a result that is confirmed by comparing Figures 3(b) and 3(d).

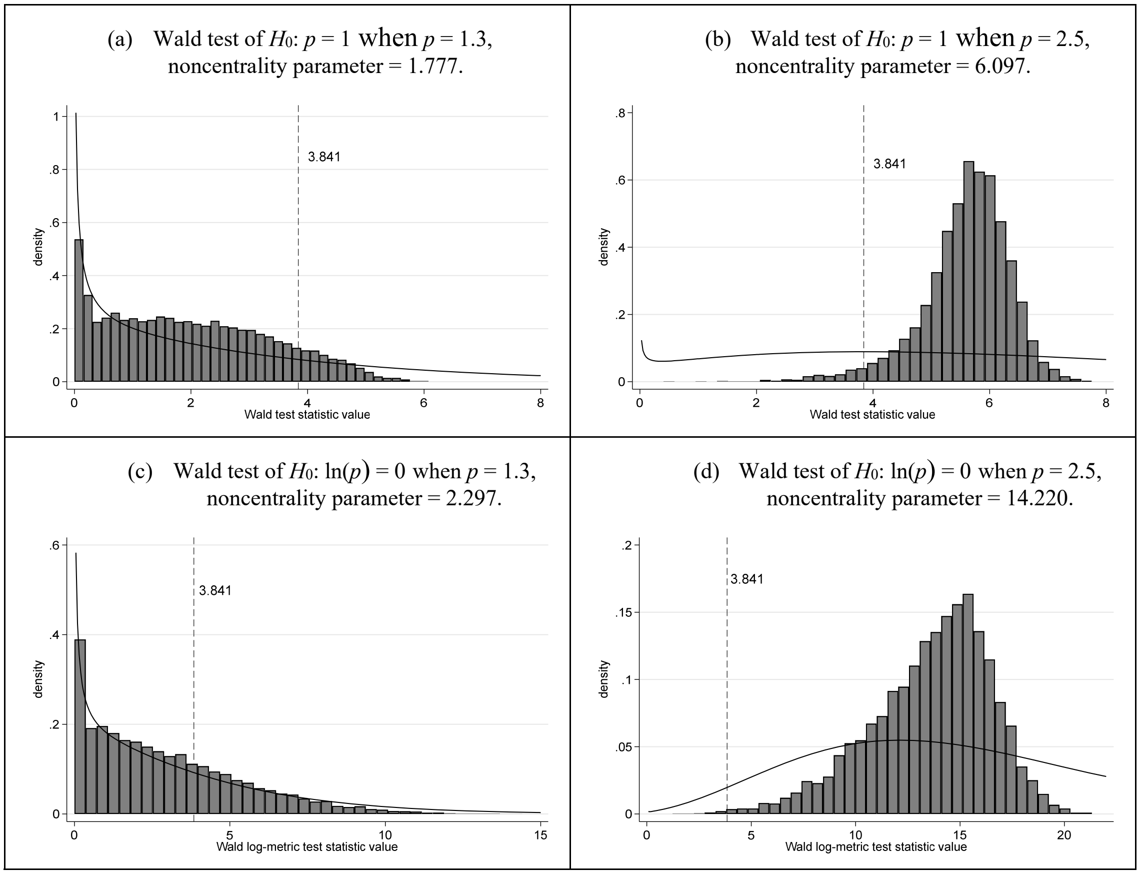
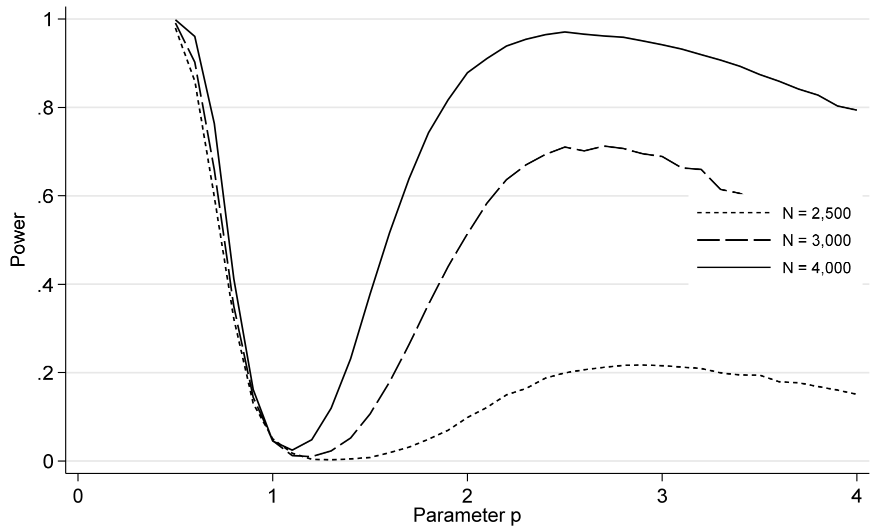


Figure 3 Comparing Monte Carlo histograms to hypothetical distributions. Comparisons of histograms of Wald test statistic values for testing $H_0: p = 1$ and $H_0: \ln(p) = 0$ with noncentral chi-square distributions given by asymptotic local approximations. We used 10,000 samples of size 4000 with settings a = 5, b = 1 and q = 1.75 to generate the test statistic values. Two settings for p were considered, $p = 1.3$ and $p = 2.5$.

Changing sample size

We finish this section on characteristics of the Wald test with illustrations of the effects of changing sample size. In Figures 4(a) and 4(b), we plot the estimated power functions for $H_0: p = 1$ and $H_0: \ln(p) = 0$, respectively, for sample sizes 2500, 3000 and 4000. It is evident from Figure 4(a) that, for $p > 1$, the power for testing $H_0: p = 1$ deteriorates rapidly when the sample size falls, but it changes very little for $p < 1$. Figure 4(b) shows that the reduction in power is much less dramatic when testing $H_0: \ln(p) = 0$.

(a) Power functions for testing H_0: $p = 1$



(b) Power functions for testing H_0: $\ln(p) = 0$

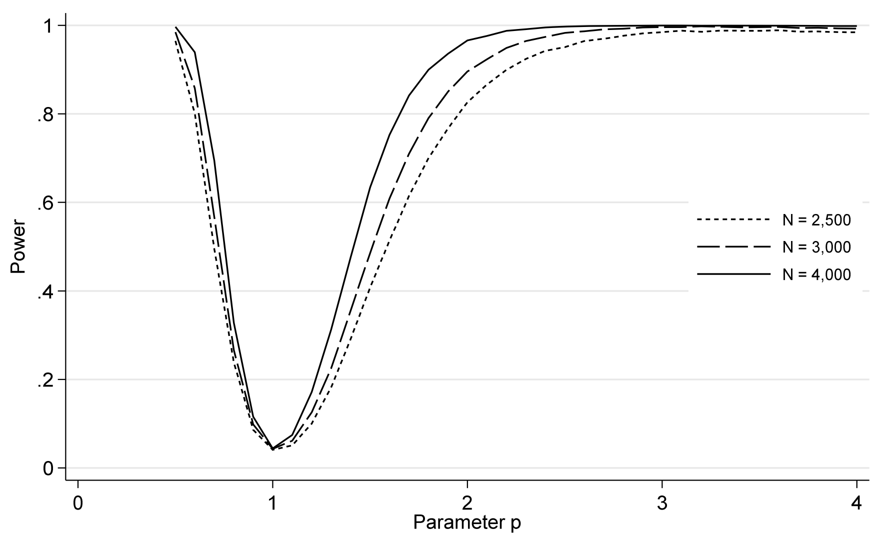


Figure 4 Monte-Carlo estimated power functions for different sample sizes. Monte-Carlo estimated power functions for Wald tests of $H_0 : p = 1$ and $H_0 : \ln(p) = 0$ for sample sizes $N = 2500$, $N = 3000$ and $N = 4000$. We used 10,000 Monte Carlo samples with settings $a = 5, b = 1$ and $q = 1.75$ to obtain the estimates.

4.3 Power comparison for the Wald and LR tests

In this subsection we compare the Monte Carlo estimated power of the LR and Wald tests for $H_0 : p = 1$ and the Wald test for $H_0 : \ln(p) = 0$. We display the power functions for these tests in Figure 5. There is little difference in the power of all tests for values $p < 1$. For $p > 1$ the powers of the LR test and the Wald test for $H_0 : \ln(p) = 0$ are virtually indistinguishable. The power of the Wald test for

$H_0 : p = 1$ is less than that of the other three tests and, as mentioned earlier, suffers from a declining power for larger values of p. The power of the LR test does increase monotonically; that for the Wald test for $H_0 : \ln(p) = 0$ does eventually decline but at much larger values of p than that from testing $H_0 : p = 1$. Reductions in power for the LR test from decreasing the sample size to 3000 and 2500 were like those in Figure 4(b) for the Wald test of $H_0 : \ln(p) = 0$.

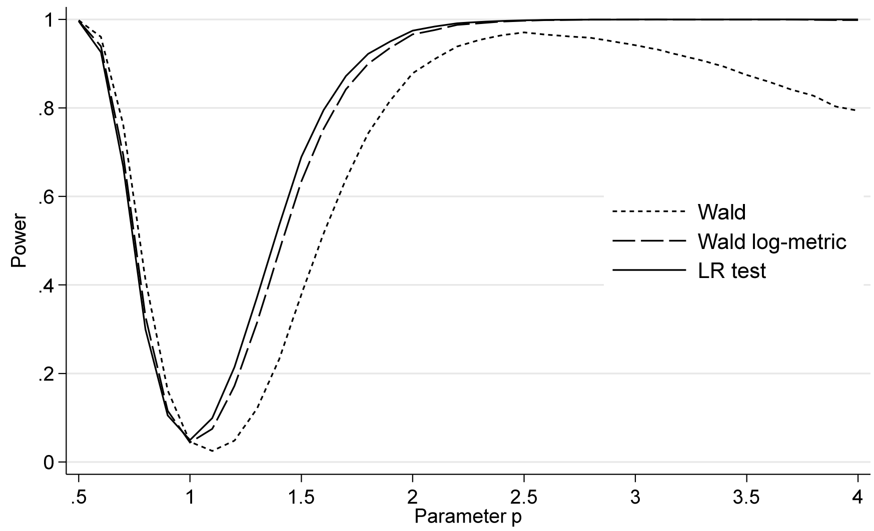


Figure 5 Power comparisons. Monte Carlo estimated power for the LR and Wald tests for $H_0 : p = 1$ and $H_0 : \ln(p) = 0$ with $N = 4000, b = 1, q = 1.75$, $a = 5$ and using 10,000 samples.

4.4 Concluding remark

Our results suggest the standard Wald test with null hypothesis $H_0 : p = 1$ should be avoided when testing for the Singh-Maddala distribution. The other tests, LR and Wald for $H_0 : \ln(p) = 0$, are reasonable alternatives. We prefer the LR test. It does not suffer from the nonmonotonic power that occurs with the Wald test. The disadvantage of the LR test is that we must estimate both restricted and unrestricted models.

5. Testing for a Dagum distribution $(q = 1)$

The Wald test statistic for testing $H_0 : q = 1$, obtained by setting $\mathbf{R} = (0\,0\,0\,1)$ and $\mathbf{r} = (1)$ in equation (3) is given by $W_D = (\hat{q} - 1)^2 \big/ \widehat{\text{var}}(\hat{q})$. For the alternative Wald statistic for testing

$H_0: \ln(q) = 0$, the asymptotic variance of $\hat{\delta} = \ln(\hat{q})$ is $\operatorname{var}(\hat{\delta}) = \operatorname{var}(\hat{q})/q^2$, and the statistic in equation (4) becomes

$$W_D^{\log} = \frac{\hat{\delta}^2}{\widehat{\operatorname{var}}(\hat{q})/\hat{q}^2} = \frac{[\hat{q}\ln(\hat{q})]^2}{\widehat{\operatorname{var}}(\hat{q})} \tag{8}$$

The LR test statistic for testing $H_0: q = 1$ uses the restricted estimator $\tilde{\boldsymbol{\theta}}' = (\tilde{a}\ \tilde{b}\ \tilde{p}\ 1)$.

5.1 Power results

A Monte Carlo experiment with the same settings as those used to test for a Singh-Maddala distribution were used except this time we set $p = 1.75$ and q was varied from 0.5 to 4 in increments of 0.1. Examples of the density functions implied by these settings are illustrated in Section S5 of the supplementary material for selected values of q. It turns out that the Wald tests for testing $q = 1$ (or $\ln(q) = 0$) have the same properties as those for testing for the Singh-Maddala distribution. This result comes from the following result which we prove in the supplementary material (corollary to Theorem 2 in Section S4).

> The asymptotic variances for the maximum likelihood estimators for p and q are symmetric in the sense that
>
> $$\operatorname{var}\left(\sqrt{N}\hat{p} \mid p = c_1, q = c_2\right) = \operatorname{var}\left(\sqrt{N}\hat{q} \mid p = c_2, q = c_1\right)$$
>
> It follows that the properties of Wald tests for p when $p = c_1$ and $q = c_2$ will be the same as those for tests for q when $p = c_2$ and $q = c_1$.

This symmetric relationship does not necessarily carry over for the LR test. However, Monte Carlo and graphical evidence[17] suggests that the power behavior of the LR test for $q = 1$ is like that for testing $p = 1$.

[17] Provided in Section S7 of the supplementary material. The graph for testing $q = 1$ (Figure S20) is virtually identical to that in Figure 5 for testing $p = 1$.

6. Testing for a Fisk distribution ($q=1$ and $p=1$)

Testing for a Fisk distribution involves testing the joint null hypothesis $H_0: p=q=1$ or, in the case of the Wald test for the log-transformed parameters, $H_0: \ln(p)=\ln(q)=0$. The Wald statistics for the two cases are given by (3) and (4), with $\mathbf{r}=(1 1)'$ and partitioned $\mathbf{R}$ into two 2×2 submatrices, $\mathbf{R}=\begin{pmatrix}\mathbf{0} & \mathbf{I}\end{pmatrix}$. [18]

6.1 Power results for testing for a Fisk distribution ($q=1$ and $p=1$)

We examine power of tests for $H_0: p=q=1$ under three different scenarios:

(i) $p=q$ with p and q varying from 0.5 to 4 in increments of 0.1.

(ii) p varies from 0.5 to 4 in increments of 0.1 with q set at its hypothesized value, $q=1$.

(iii) p varies from 0.5 to 4 in increments of 0.1 with q set at $q=1.75$.

The symmetry between p and q implied by Theorem 2 in Section S4 of the supplement means that similar results would be obtained if the roles of p and q were reversed. The other settings were $a=5$, $b=1$ and 10,000 samples of size 4000 were used to estimate power.

Results for $0.5\le p=q\le 4.0$

In Figure 6, we plot the estimated power functions for the case where both p and q are changing such that $p=q$. The most dramatic outcome is the extremely poor performance of the Wald test. For values of p and q greater than the hypothesized value of one, its power first declines to values less than the 0.05 significance level, then it increases, but not reaching 0.05 until $p=q=1.6$, and never exceeding 0.2 before it starts to decline again. All the tests perform well when $p=q<1$. The Wald test for $H_0: \ln(p)=\ln(q)=0$ has a slight advantage over the LR test when $p=q>1$, and the Wald test on the original parameters is best for $p=q<1$, but when the complete range of (p,q) values is considered, the LR test is a better choice.

[18] Detailed expressions for these Wald tests are in the supplement (Section S9.1).

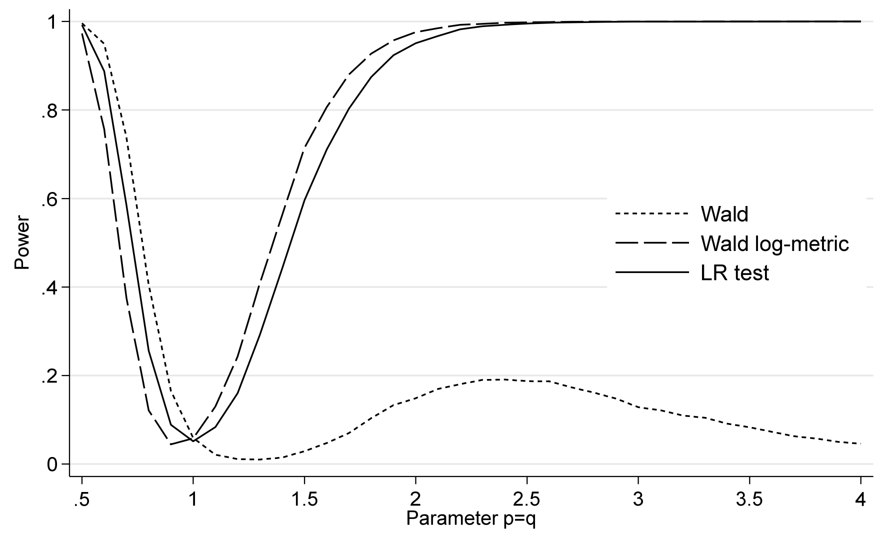


Figure 6 Power comparisons for the Wald and LR test. The hypotheses are $H_0 : p = q = 1$ and $H_0 : \ln(p) = \ln(q) = 0$ for the log-metric Wald test. Settings are $0.5 \le p = q \le 4.0$, $a = 5$, $b = 1$, $N = 4000$ and 10,000 samples.

Results for $0.5 \le p \le 4.0$ *with* $q = 1$

Under the label "conditional on $q = 1$", we plot the powers of the LR test and the Wald test for $H_0 : p = q = 1$ in Figure 7. We have omitted the power curve for the log-metric Wald test because it is almost identical to that for the LR test. Also, we have included two additional results. One is the power of the LR test reproduced from Figure 6 where the null hypothesis is $H_0 : p = q = 1$ and *p* and *q* range over the interval such that $p = q$; *q* is not restricted. Comparing this curve with those conditional on $q = 1$ shows how conditioning on $q = 1$, which implies the joint hypothesis $H_0 : p = q = 1$ contains a redundant component, increases the power of the joint test. The second curve that we added is the power curve for the LR test for the single hypothesis $H_0 : p = 1$, computed assuming $q = 1$. The power from this test is less than that of the LR test for $H_0 : p = q = 1$ conditional on $q = 1$, illustrating a point related to that made by Griffiths and Hill (2022), that adding a redundant, superfluous parameter restriction can sometimes increase the power of an *F*-test. It is also true here for the chi-square test.

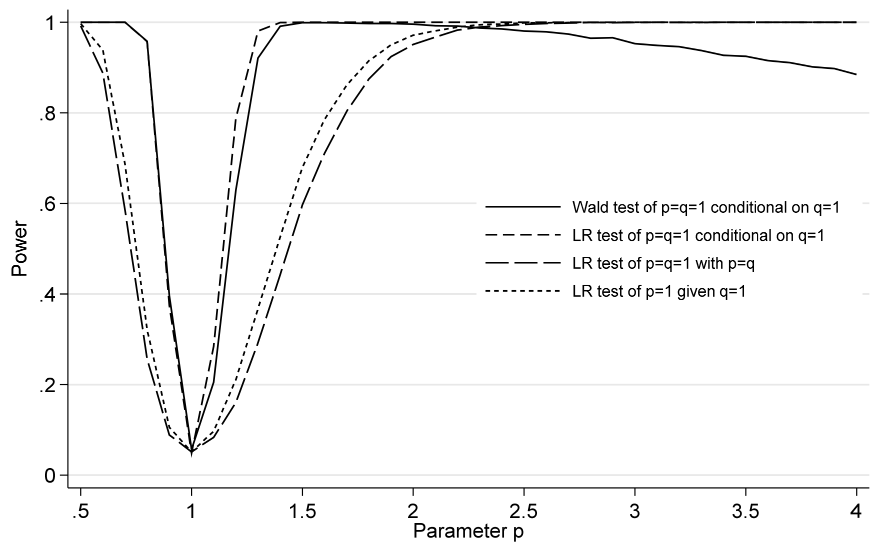


Figure 7 Power comparisons for Wald and LR tests when *q* = 1. The hypotheses are $H_0 : p = q = 1$ and $H_0 : \ln(p) = \ln(q) = 0$ for the Wald and LR tests. The top two curves are conditional on $q = 1$. The lower curve is the LR test for $H_0 : p = q = 1$ given that $p = q$; it does not condition on $q = 1$. The second lowest curve is the power from the LR test of $H_0 : p = 1$ given that $q = 1$. Settings are $0.5 \le p \le 4.0$, $a = 5$, $b = 1$, $N = 4000$ and 10,000 samples.

Results for $0.5 \le p \le 4.0$ *with* $q = 1.75$

In this case the null hypothesis $H_0 : p = q = 1$ is false for all values of *p*. The most desirable outcome is for all powers to be equal to one over the complete range of *p*. We plot the estimated power functions in Figure 8. However, besides the poor power of the Wald test, the most striking outcome in Figure 8 is the decline in power for all tests when *p* is near the fixed value, $q = 1.75$. Conditional on this false value, all tests exhibit nonmonotonic power with respect to an increasing value of *p* beyond one, reaching a minimum at $p = q = 1.75$. Some experimentation suggested this result is a more general one that holds for other settings of *q*.[19] In this regard it is useful to distinguish between the Wald test for $H_0 : p = q = 1$ and the other two tests, the log-metric Wald test for $H_0 : \ln(p) = \ln(q) = 0$, and the LR test. For other false settings of *q*, the power of the Wald test reaches a minimum when $p = q$,

[19] See Section 9.1 of the supplementary material.

increases, then declines again, becoming particularly poor for larger values of q, such as $q = 3$. For $q = 3$, the power of the other tests was one over the complete range of p.

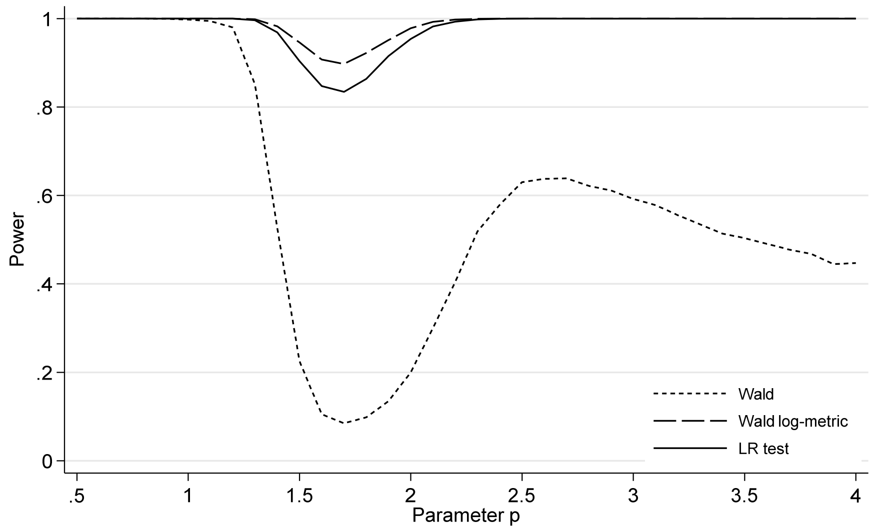


Figure 8 Power comparisons for Wald and LR tests with *q* = 1.75. The hypotheses are $H_0 : p = q = 1$ and $H_0 : \ln(p) = \ln(q) = 0$ for the log-metric Wald test. Settings are $0.5 \leq p \leq 4.0$, $a = 5$, $b = 1$, $q =$ 1.75, $N = 4000$ and 10,000 samples.

7. Concluding remarks

After estimating a GB2 distribution, a decision about whether a special case with fewer parameters is adequate is likely to be routinely made based on a Wald test. Typical output from software programs that report estimates and standard errors facilitate such decisions. We have shown that such tests on the original parameters for the Singh-Maddala, Dagum and Fisk distributions have a high probability of failing to reject an incorrect special case. Overcome this problem by using a likelihood ratio test or a Wald test on the log-transformed parameters. Of these two, we recommend the likelihood ratio test. Power computed using a noncentral chi-square distribution, valid when considering only local alternatives, can be a very poor approximation to actual power. Sometimes it is an understatement, and sometimes an overstatement, of actual power. An extensive comparison of Monte-Carlo estimated distributions with noncentral chi-square distributions, reported in Section S3 of the supplementary material, shows how the asymptotic local alternatives assumption can be a poor one.

References

Bickel, P. J., and Doksum, K. A. (2015), Mathematical Statistics, Basic Ideas and Selected Topics, Volume 1, Second edition, Boca Raton, FL: CRC Press, Boca Raton, Florida.

Butler, R.J., and McDonald, J. B. (1989), “Using Incomplete Moments to Measure Inequality,” Journal of Econometrics, 42, 109-120.

Cao, M., Sun, W., and Kosorok, M. R. (2013), “The Optimal Power Puzzle: Scrutiny of the Momnotone Likelihood Ratio Assumption in Multiple Testing,” Biometrika, 100, 495-502.

Chotikapanich, D., Griffiths, W. E. (2023), “Poverty in Southeast Asia”, Ch.65 in J. Silber, editor, Handbook of Research on Measuring Poverty and Deprivation, Edward Elgar, 696-709.

Chotikapanich, D., Griffiths, W. E., Hajargasht, G., Karunarathne, W., and Rao, D. S. P. (2018), “Using the GB2 Income Distribution”. Econometrics, 6(2), 21, 1-24. https://doi.org/10.3390/econometrics6020021.

Crainiceanu, C. M., and Vogelsang, T. J. (2007), “Nonmonotonic Power for Tests of a Mean Shift in a Time Series,” Journal of Statistical Computation and Simulation, 77, 457-476.

Dagum, C. (1977), “A New Model of Personal Income Distribution: Specification and Estimation,” Economie Appliquée, 30, 413-437.

Fisk, P. R. (1961), “The graduation of income distributions,” Econometrica, 29, 171-185.

Greene, W. H. (2018), Econometric Analysis (8th ed.), Upper Saddle River, NJ: Pearson, Upper Saddle River, New Jersey.

Griffiths, W. E., and Hill, R. C. (2022) “On the Power of the F-test for Hypotheses in a Linear Model,” The American Statistician, 76, 78-84.

Jenkins, S. P. (2009), “Distributionally-Sensitive Inequality Indices and the GB2 Income Distribution,” Review of Income and Wealth, 55, 392-98.

Jenkins, S. P. (2014), “GB2LFIT: Stata module to fit Generalized Beta of the Second Kind distribution by maximum likelihood (log parameter metric)”, Statistical Software Components S457897, Boston College Department of Economics.

Kleiber, C. (2008), “A Guide to Dagum Distributions,” in Chotikapanich, D. (Ed.), Modeling Income Distributions and Lorenz Curves, New York: Springer, New York, 97-118.

Kleiber, C. and Kotz, S. (2003), Statistical Size Distributions in Economics and Actuarial Sciences, New York: John Wiley and & Sons, New York.

McDonald, J. B. (1984), "Some generalized functions for the size distribution of income," Econometrica, 52(3), 647-663.

McDonald, J. B., and M. R. Ransom (2008), "The Generalized Beta Distribution as a Model for the Distribution of Income: Estimation of Related Measures of Inequality," in Chotikapanich, D. (Ed.), Modeling Income Distributions and Lorenz Curves, New York: Springer, New York, 147-166.

McDonald J. B. and J. W. Triplett (2024), "*gintreg*: A generalization of Stata's *intreg* and *stintreg* Commands", mimeo.

McDonald, J. B., and Xu, Y. J. (1995). A generalization of the beta distribution with applications. Journal of Econometrics, 66(1-2), 133-152. [Erratum: Journal of Econometrics, 69, 427-428.]

McDonald, J.B and Y. Xu (1992), "An empirical investigation of the likelihood ratio test when the boundary condition is violated", Communications in Statistics: Simulation and Computation, 21(3), 879-892.

McDonald, J. B., J. Sorensen, J., and P.A. Turley, P. A. (2011), "Skewness and Kurtosis Properties of Income Distribution Models," Review of Income and Wealth, 59, 360-374.

McManus, D. A. (1991), "Who Invented Local Power Analysis," Econometric Theory, 7, 265-268.

Nelson, F. D., and Savin, N. E. (1988), "The Nonmonotonicity of the Power Function of the Wald Test in Nonlinear Models," Department of Economics Working Paper Series No. 88-7, University of Iowa.

Nelson, F. D., and Savin, N. E. (1990), "The Danger of Extrapolating Asymptotic Local Power," Econometrica, 58(4), 977-981.

Rao, D. S. P., Rambaldi, A. N., Hajargasht, G., and Griffiths, W. E. (2023), "The University of Queensland International Comparison Database, UQICD V3.0: User Guide," CEPA Working Papers Series WP09/2022, School of Economics, University of Queensland, Australia.

Ruud, P. A. (2000), An Introduction to Classical Econometric Theory, Oxford: Oxford University Press, Oxford.

Sarabia, J. M., Jordá. V., and Remuzgo, L. (2017), "The Theil Indices in Parametric Families of Income Distributions – A Short Review", Review of Income and Wealth, 63, 867-880.

Savin, N. E., and Würtz, A. M. (1999), "Power of Tests in Binary Response Models," Econometrica, 67, 413-421.

Singh, S. K. and G.S. Maddala, G. S. (1976), A function for the size distribution of incomes," Econometrica, 44, 963-970.

StataCorp (2023), Stata Statistical Software: Release 18, College Station, TX: StataCorp LLC.

Vartia, P. L. I., and Vartia, Y. O. (1980), "Description of the income distribution by the scaled *F* distribution model," in N.A. Klevmarken, N.A. and J.A. Lybeck, J.A. (Eds.), The Statics and Dynamics of Income, Clevedon, U.K., 23-36.

Vogelsang, T. J. (1997), "Wald-type Tests for Detecting Shifts in the Trend Function of a Dynamic Time Series," Econometric Theory, 13, 818-849.

Vogelsang, T. J. (1999), "Sources of Nonmonotonic Power When Testing for a Shift in Mean of a Dynamic Time Series," Journal of Econometrics, 88, 283-299.

Testing Special Cases of the GB2 Distribution

William E. Griffiths
University of Melbourne
R. Carter Hill
Louisiana State University

Supplementary Material

Supplementary material outline

Supplementary Material Figures

Figure S12 The Singh-Maddala distribution, $p = 1$, with $b = 1$ and $a = 5$

Figure S13 The Fisk distribution, $p = q = 1$ over a, with $b = 1$

Figure S14 GB2 distributions used to test for the Dagum distribution, $q = 1$, for values of q given $p = 1.75$, $a = 5$ and $b = 1$

Figure S15 GB2 distributions used to test for the Beta distribution of the second kind, $a = 1$, for values of a given $p = q = 1.75$ and $b = 1$

Figure S16 GB2 distributions used to test for the Singh-Maddala distribution, $p = 1$, for values of p given $q = 1.75$, $a = 5$ and $b = 1$.

Figure S17 GB2 distributions used to test for the Fisk distribution, $p = q = 1$, for values of p and q given $a = 5$ and $b = 1$.

Figure S18 The derivative of the asymptotic variance of the MLE for p.

Figure S19 Wald test hypothetical power deconstruction.

Figure S20 Monte Carlo power testing $q = 1$, and the log-metric version $\ln(q) = 0$.

Figure S21 Monte-Carlo estimated power for the LM, LR and Wald tests for testing a.

Figure S22(a) Contours of a log-likelihood function when $a = 5$, $b = 1$, $p = 1$ and $q = 1.75$
Figure S22(b) Contours of a log-likelihood function when $a = 5$, $b = 1$, $p = 2$ and $q = 1.75$
Figure S22(c) Contours of a log-likelihood function when $a = 5$, $b = 1$, $p = 3$ and $q = 1.75$
Figure S22(d) Contours of a log-likelihood function when $a = 5$, $b = 1$, $p = 4$ and $q = 1.75$

Figure S23(a) Contours of a log-likelihood function when $a = 5$, $b = 1$, $p = 1$ and $q = 1.75$ in log-metric
Figure S23(b) Contours of a log-likelihood function when $a = 5$, $b = 1$, $p = 2$ and $q = 1.75$ in log-metric
Figure S23(c) Contours of a log-likelihood function when $a = 5$, $b = 1$, $p = 3$ and $q = 1.75$ in log-metric
Figure S23(d) Contours of a log-likelihood function when $a = 5$, $b = 1$, $p = 4$ and $q = 1.75$ in log-metric

Figure S24 Conditioning diagnostics**.** The average condition indexes for the gradient matrices $\mathbf{G}$ and $\mathbf{GD}^{-1}$.

Supplementary Material Tables

S1 First and second order derivatives and information matrix

If x is the GB2 random variable, $x \sim \text{GB2}(a,b,p,q)$, it has probability density function

$$f(x)=\frac{ax^{ap-1}}{b^{ap}B(p,q)\left[1+\left(x/b\right)^{a}\right]^{p+q}}=\frac{a\left(x/b\right)^{ap}}{xB(p,q)\left[1+\left(x/b\right)^{a}\right]^{p+q}}$$

Using a transformation of variables, it is straightforward to show that $y=\frac{\left(x/b\right)^{a}}{1+\left(x/b\right)^{a}}$ is a $\text{beta}(p,q)$ random variable with density function

$$f(y)=\frac{1}{B(p,q)}y^{p-1}(1-y)^{q-1}$$

where $B(p,q)=\frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$ is the beta function. We also use the following notation:

$$\psi(z)=\frac{d\ln\Gamma(z)}{dz}=\frac{1}{\Gamma(z)}\frac{d\Gamma(z)}{dz} \qquad \text{(digamma function)}$$

$$\psi'(z)=\frac{d\psi(z)}{dz}=\frac{d^{2}\ln\Gamma(z)}{dz^{2}} \qquad \text{(trigamma function)}$$

The log-likelihood for a single observation $\ln L_i=\ln f\left(x_i\right)$ is

$$\ln L_i=\ln a+ap\ln\left(\frac{x_i}{b}\right)-\ln x_i-\ln B(p,q)-(p+q)\ln\left[1+\left(\frac{x_i}{b}\right)^{a}\right]$$

This expression must be summed over i for the complete log-likelihood; the same holds for the derivatives that follow. It is convenient to write these derivatives in terms of $y_i=\left(x_i/b\right)^{a}\Big/\left[1+\left(x_i/b\right)^{a}\right]$. It should be kept in mind that y_i depends on a and b.

S1.1 First derivatives

$$\frac{\partial\ln L_i}{\partial a}=\frac{1}{a}+\ln\left(\frac{x_i}{b}\right)\left[p-(p+q)y_i\right]$$

$$\frac{\partial\ln L_i}{\partial b}=\frac{a}{b}\left\{(p+q)y_i-p\right\}$$

$$\frac{\partial\ln L_i}{\partial p}=\ln y_i-\psi(p)+\psi(p+q)$$

$$\frac{\partial\ln L_i}{\partial q}=-\psi(q)+\psi(p+q)+\ln\left(1-y_i\right)$$

S1.2 Second derivatives

$$\frac{\partial^{2}\ln L_i}{\partial a^{2}}=-\frac{1}{a^{2}}-(p+q)\ln^{2}\left(\frac{x_i}{b}\right)\left[y_i\left(1-y_i\right)\right]$$

$$\frac{\partial^2 \ln L_i}{\partial a \partial b} = \frac{1}{b}\left\{-p+(p+q)y_i\left[1+\ln\left(\frac{y_i}{1-y_i}\right)-y_i\ln\left(\frac{y_i}{1-y_i}\right)\right]\right\}$$

$$\frac{\partial^2 \ln L_i}{\partial p \partial a} = \ln\left(\frac{x_i}{b}\right)(1-y_i)$$

$$\frac{\partial^2 \ln L_i}{\partial q \partial a} = -y_i \ln\left(\frac{x_i}{b}\right)$$

$$\frac{\partial^2 \ln L_i}{\partial b^2} = -\frac{a}{b^2}\left[a(p+q)y_i(1-y_i)+(p+q)y_i-p\right]$$

$$\frac{\partial^2 \ln L_i}{\partial p \partial b} = \frac{a}{b}(y_i-1)$$

$$\frac{\partial^2 \ln L_i}{\partial q \partial b} = \frac{a}{b}y_i$$

$$\frac{\partial^2 \ln L_i}{\partial p^2} = \psi'(p+q)-\psi'(p)$$

$$\frac{\partial^2 \ln L_i}{\partial q \partial p} = \psi'(p+q)$$

$$\frac{\partial^2 \ln L_i}{\partial q^2} = \psi'(p+q)-\psi'(q)$$

S1.3 Information matrix

$$\Im = \begin{pmatrix} \Im_{aa} & \Im_{ab} & \Im_{ap} & \Im_{aq} \\ \Im_{ab} & \Im_{bb} & \Im_{bp} & \Im_{bq} \\ \Im_{ap} & \Im_{bp} & \Im_{pp} & \Im_{pq} \\ \Im_{aq} & \Im_{bq} & \Im_{pq} & \Im_{qq} \end{pmatrix}$$

where

$$\Im_{aa} = -E\left(\frac{\partial^2 \ln L_i}{\partial a^2}\right)$$

$$= \frac{1}{a^2}\left\{1+\left(\frac{1}{p+q+1}\right)\left[pq\left\{\left[\psi(p)-\psi(q)\right]^2+\psi'(p)+\psi'(q)\right\}+2(q-p)\left[\psi(p)-\psi(q)\right]-2\right]\right\}$$

$$\Im_{ab} = -E\left(\frac{\partial^2 \ln L_i}{\partial a \partial b}\right) = \frac{-pq\left[\psi(p)-\psi(q)\right]+p-q}{b(p+q+1)}$$

$$\Im_{ap} = -E\left(\frac{\partial^2 \ln L_i}{\partial p \partial a}\right) = \frac{1}{a(p+q)}\left\{1-q\left[\psi(p)-\psi(q)\right]\right\}$$

$$\Im_{aq} = -E\left(\frac{\partial^2 \ln L_i}{\partial q \partial a}\right) = \frac{1}{a(p+q)}\left\{1+p\left[\psi(p)-\psi(q)\right]\right\}$$

$$\Im_{bb} = -E\left(\frac{\partial^2 \ln L_i}{\partial b^2}\right) = \frac{a^2 pq}{b^2(p+q+1)}$$

$$\Im_{bp} = -E\left(\frac{\partial^2 \ln L_i}{\partial p \partial b}\right) = \frac{aq}{b(p+q)}$$

$$\Im_{bq} = -E\left(\frac{\partial^2 \ln L_i}{\partial q \partial b}\right) = -\frac{ap}{b(p+q)}$$

$$\Im_{pp} = -E\left(\frac{\partial^2 \ln L_i}{\partial p^2}\right) = \psi'(p) - \psi'(p+q)$$

$$\Im_{pq} = -E\left(\frac{\partial^2 \ln L_i}{\partial q \partial p}\right) = -\psi'(p+q)$$

$$\Im_{qq} = -E\left(\frac{\partial^2 \ln L_i}{\partial q^2}\right) = \psi'(q) - \psi'(p+q)$$

S1.4 Derivations

S1.4.1 Some Basic Results

$$\psi(z+1) = \psi(z) + \frac{1}{z} \qquad\qquad \psi'(z+1) = \psi'(z) - \frac{1}{z^2} \tag{1.1}$$

$$B(p+1,q) = B(p,q)\frac{p}{P+q} \qquad\qquad B(p,q+1) = B(p,q)\frac{q}{p+q} \tag{1.2}$$

$$E(y) = E\left(\frac{(x/b)^a}{1+(x/b)^a}\right) = \frac{p}{p+q} \qquad\qquad E(1-y) = E\left(\frac{1}{1+(x/b)^a}\right) = \frac{q}{p+q} \tag{1.3}$$

$$E\left[y(1-y)\right] = E\left(\frac{(x/b)^a}{\left[1+(x/b)^a\right]^2}\right) = \frac{pq}{(p+q)(p+q+1)} \tag{1.4}$$

$$x = b\left(\frac{y}{1-y}\right)^{1/a} \qquad\qquad \ln\left(\frac{x}{b}\right) = \frac{1}{a}\left[\ln y - \ln(1-y)\right] \tag{1.5}$$

From Gradshteyn and Ryzik (1980, pages 538 and 541),

$$\begin{aligned} E(\ln y) &= \psi(p) - \psi(p+q) \\ E\left[\ln(1-y)\right] &= \psi(q) - \psi(p+q) \\ E\left[\ln y - \ln(1-y)\right] &= \psi(p) - \psi(q) \end{aligned} \tag{1.6}$$

$$E\left(\ln^2 y\right)=\left[\psi(p)-\psi(p+q)\right]^2+\psi'(p)-\psi'(p+q)$$
$$E\left[\ln^2(1-y)\right]=\left[\psi(q)-\psi(p+q)\right]^2+\psi'(q)-\psi'(p+q) \qquad (1.7)$$
$$E\left[\left(\ln y\right)\left(\ln(1-y)\right)\right]=\left[\psi(p)-\psi(p+q)\right]\left[\psi(q)-\psi(p+q)\right]-\psi'(p+q)$$

Using (1.6) and reparametrizing the beta distribution, we can obtain

$$E(y\ln y)=\left(\frac{p}{p+q}\right)\left[\psi(p+1)-\psi(p+q+1)\right]$$
$$E\left[y\ln(1-y)\right]=\left(\frac{p}{p+q}\right)\left[\psi(q)-\psi(p+q+1)\right]$$
$$E\left[y\ln y-y\ln(1-y)\right]=\left(\frac{p}{p+q}\right)\left[\psi(p+1)-\psi(q)\right] \qquad (1.8)$$
$$=\left(\frac{p}{p+q}\right)\left[\psi(p)-\psi(q)\right]+\frac{1}{p+q}$$

$$E\left(y^2\ln y\right)=\left(\frac{p}{p+q}\right)\left(\frac{p+1}{p+q+1}\right)\left[\psi(p+2)-\psi(p+q+2)\right]$$
$$E\left[y^2\ln(1-y)\right]=\left(\frac{p}{p+q}\right)\left(\frac{p+1}{p+q+1}\right)\left[\psi(q)-\psi(p+q+2)\right]$$
$$E\left[y^2\ln y-y^2\ln(1-y)\right]=\left(\frac{p}{p+q}\right)\left(\frac{p+1}{p+q+1}\right)\left[\psi(p+2)-\psi(q)\right] \qquad (1.9)$$
$$=\left(\frac{p}{p+q}\right)\left(\frac{p+1}{p+q+1}\right)\left[\psi(p)-\psi(q)\right]+\frac{2p+1}{\left(p+q\right)\left(p+q+1\right)}$$

Using (1.7) and reparametrizing the beta distribution, we can obtain

$$E\left(y\ln^2 y\right)=\frac{p}{p+q}\left\{\left[\psi(p+1)-\psi(p+q+1)\right]^2+\psi'(p+1)-\psi'(p+q+1)\right\}$$
$$E\left[y\ln^2(1-y)\right]=\frac{p}{p+q}\left\{\left[\psi(q)-\psi(p+q+1)\right]^2+\psi'(q)-\psi'(p+q+1)\right\} \qquad (1.10)$$
$$E\left[y\left(\ln y\right)\left(\ln(1-y)\right)\right]=\frac{p}{p+q}\left\{\left[\psi(p+1)-\psi(p+q+1)\right]\left[\psi(q)-\psi(p+q+1)\right]-\psi'(p+q+1)\right\}$$

$$E\left(y^2\ln^2 y\right)=\left(\frac{p}{p+q}\right)\left(\frac{p+1}{p+q+1}\right)\left\{\left[\psi(p+2)-\psi(p+q+2)\right]^2+\psi'(p+2)-\psi'(p+q+2)\right\}$$
$$E\left[y^2\ln^2(1-y)\right]=\left(\frac{p}{p+q}\right)\left(\frac{p+1}{p+q+1}\right)\left\{\left[\psi(q)-\psi(p+q+2)\right]^2+\psi'(q)-\psi'(p+q+2)\right\} \qquad (1.11)$$
$$E\left[y^2\left(\ln y\right)\left(\ln(1-y)\right)\right]=\left(\frac{p}{p+q}\right)\left(\frac{p+1}{p+q+1}\right)\left\{\left[\psi(p+2)-\psi(p+q+2)\right]\left[\psi(q)-\psi(p+q+2)\right]\right.$$
$$\left.-\psi'(p+q+2)\right\}$$

Now consider

$$E\left[y\ln^2\left(\frac{x}{b}\right)\right]=\left(\frac{1}{a^2}\right)E\left\{y\left[\ln y-\ln(1-y)\right]^2\right\}$$

$$E\left[y\ln^2\left(\frac{x}{b}\right)\right]=\left(\frac{1}{a^2}\right)E\left\{y\left[\ln y-\ln(1-y)\right]^2\right\}$$

$$=\left(\frac{1}{a^2}\right)\left\{E\left(y\ln^2 y\right)+E\left[y\ln^2(1-y)\right]-2E\left[y\left(\ln y\right)\left(\ln(1-y)\right)\right]\right\}$$

$$=\left(\frac{1}{a^2}\right)\left(\frac{p}{p+q}\right)\left[\psi^2(p+1)+\psi'(p+1)+\psi^2(q)+\psi'(q)-2\psi(q)\psi(p+1)\right]$$

$$=\left(\frac{1}{a^2}\right)\left(\frac{p}{p+q}\right)\left\{\psi^2(p)+\frac{2}{p}\psi(p)+\frac{1}{p^2}+\psi'(p)-\frac{1}{p^2}+\psi^2(q)+\psi'(q)-2\psi(q)\left[\psi(p)+\frac{1}{p}\right]\right\}$$

$$=\left(\frac{1}{a^2}\right)\left(\frac{p}{p+q}\right)\left\{\left[\psi(p)-\psi(q)\right]^2+\psi'(p)+\psi'(q)+\frac{2}{p}\left[\psi(p)-\psi(q)\right]\right\} \tag{1.12}$$

$$E\left[y^2\ln^2\left(\frac{x}{b}\right)\right]=\left(\frac{1}{a^2}\right)E\left\{y^2\left[\ln y-\ln(1-y)\right]^2\right\}$$

$$=\left(\frac{1}{a^2}\right)\left\{E\left(y^2\ln^2 y\right)+E\left[y^2\ln^2(1-y)\right]-2E\left[y^2\left(\ln y\right)\left(\ln(1-y)\right)\right]\right\}$$

$$=\left(\frac{1}{a^2}\right)\left(\frac{p}{p+q}\right)\left(\frac{p+1}{p+q+1}\right)\left[\psi^2(p+2)+\psi'(p+2)+\psi^2(q)+\psi'(q)-2\psi(q)\psi(p+2)\right]$$

$$=\left(\frac{1}{a^2}\right)\left(\frac{p}{p+q}\right)\left(\frac{p+1}{p+q+1}\right)\left\{\left[\psi(p)+\frac{1}{p+1}+\frac{1}{p}\right]^2+\psi'(p)-\frac{1}{(p+1)^2}-\frac{1}{p^2}+\psi^2(q)\right.$$
$$\left.+\psi'(q)-2\psi(q)\psi(p)-2\psi(q)\left[\frac{1}{p+1}+\frac{1}{p}\right]\right\}$$

$$=\left(\frac{1}{a^2}\right)\left(\frac{p}{p+q}\right)\left(\frac{p+1}{p+q+1}\right)\left\{\psi^2(p)+\frac{2(2p+1)}{p(p+1)}\psi(p)+\psi'(p)+\psi^2(q)+\psi'(q)\right.$$
$$\left.-2\psi(q)\psi(p)-\frac{2(2p+1)}{p(p+1)}\psi(q)+\frac{2}{p(p+1)}\right\}$$

$$=\left(\frac{1}{a^2}\right)\left(\frac{p}{p+q}\right)\left(\frac{p+1}{p+q+1}\right)\left\{\left[\psi(p)-\psi(q)\right]^2+\psi'(p)+\psi'(q)+\frac{2}{p(p+1)}\right.$$
$$\left.+\frac{2(2p+1)}{p(p+1)}\left[\psi(p)-\psi(q)\right]\right\} \tag{1.13}$$

$$\frac{\partial\ln B(p,q)}{\partial p}=\frac{\partial}{\partial p}\left[\ln\Gamma(p)+\ln\Gamma(q)-\ln\Gamma(p+q)\right]=\psi(p)-\psi(p+q)$$

$$\frac{\partial\ln B(p,q)}{\partial q}=\frac{\partial}{\partial q}\left[\ln\Gamma(p)+\ln\Gamma(q)-\ln\Gamma(p+q)\right]=\psi(q)-\psi(p+q)$$

S1.4.2 Derivatives

Consider derivatives of the log-likelihood for a single observation $\ln L_i = \ln f\left(x_i\right)$, with the *i*-subscript dropped.

$$\ln L = \ln a + ap\ln\left(\frac{x}{b}\right) - \ln x - \ln B(p,q) - (p+q)\ln\left[1+\left(\frac{x}{b}\right)^a\right]$$

$$\frac{\partial \ln L}{\partial a} = \frac{1}{a} + p\ln\left(\frac{x}{b}\right) - \frac{(p+q)\left(\frac{x}{b}\right)^a \ln\left(\frac{x}{b}\right)}{1+\left(\frac{x}{b}\right)^a}$$

$$\frac{\partial \ln L}{\partial b} = \frac{apb}{x}\left(-\frac{x}{b^2}\right) + \frac{a(p+q)x^a b^{-a-1}}{1+\left(\frac{x}{b}\right)^a} = \frac{a}{b}\left\{\frac{(p+q)\left(\frac{x}{b}\right)^a}{\left[1+\left(\frac{x}{b}\right)^a\right]} - p\right\}$$

$$= \frac{a}{b}\left\{(p+q)\left[\left(\frac{b}{x}\right)^a + 1\right]^{-1} - p\right\}$$

$$\frac{\partial \ln L}{\partial p} = a\ln\left(\frac{x}{b}\right) - \psi(p) + \psi(p+q) - \ln\left[1+\left(\frac{x}{b}\right)^a\right]$$

$$\frac{\partial \ln L}{\partial q} = -\psi(q) + \psi(p+q) - \ln\left[1+\left(\frac{x}{b}\right)^a\right]$$

$$\frac{\partial^2 \ln L}{\partial a \partial b} = -\frac{p}{b} + \frac{(p+q)\left(\frac{x}{b}\right)^a}{b\left[1+\left(\frac{x}{b}\right)^a\right]} + \frac{a(p+q)}{b}\left\{\frac{\left(\frac{x}{b}\right)^a \ln\left(\frac{x}{b}\right)}{1+\left(\frac{x}{b}\right)^a} + \left(\frac{x}{b}\right)^a\left[-\frac{\left(\frac{x}{b}\right)^a \ln\left(\frac{x}{b}\right)}{\left[1+\left(\frac{x}{b}\right)^a\right]^2}\right]\right\}$$

$$= -\frac{p}{b} + \frac{(p+q)y}{b} + \frac{a(p+q)}{b}\left\{\frac{y}{a}\left[\ln y - \ln(1-y)\right] - \frac{y^2}{a}\left[\ln y - \ln(1-y)\right]\right\}$$

$$-E\left(\frac{\partial^2 \ln L}{\partial a \partial b}\right) = -\frac{(p+q)}{b}\left\{\left(\frac{p}{p+q}\right)\left[\psi(p)-\psi(q)\right] + \frac{1}{p+q}\right.$$

$$\left. - \left(\frac{p}{p+q}\right)\left(\frac{p+1}{p+q+1}\right)\left[\psi(p)-\psi(q)\right] - \frac{2p+1}{(p+q)(p+q+1)}\right\}$$

$$= -\frac{1}{b}\left\{\left[\psi(p)-\psi(q)\right]p\left(1-\frac{p+1}{p+q+1}\right) + 1 - \frac{2p+1}{p+q+1}\right\}$$

$$= \frac{-pq\left[\psi(p)-\psi(q)\right] + p - q}{b(p+q+1)}$$

$$\frac{\partial^2 \ln L}{\partial p \partial a} = \ln\left(\frac{x}{b}\right) - \frac{\left(\frac{x}{b}\right)^a \ln\left(\frac{x}{b}\right)}{1+\left(\frac{x}{b}\right)^a}$$

$$-E\left(\frac{\partial^2 \ln L}{\partial p \partial a}\right) = E\left[y \ln\left(\frac{x}{b}\right)\right] - E\left[\ln\left(\frac{x}{b}\right)\right]$$

$$= \frac{1}{a}\left\{E\left[y \ln y - y\ln(1-y)\right] - E\left[\ln y - \ln(1-y)\right]\right\}$$

$$= \frac{1}{a}\left\{\left(\frac{p}{p+q}\right)\left[\psi(p)-\psi(q)\right] + \frac{1}{p+q} - \left[\psi(p)-\psi(q)\right]\right\}$$

$$= \frac{1}{a(p+q)}\left\{1 - q\left[\psi(p)-\psi(q)\right]\right\}$$

$$\frac{\partial^2 \ln L}{\partial q \partial a} = -\frac{\left(\frac{x}{b}\right)^a \ln\left(\frac{x}{b}\right)}{1+\left(\frac{x}{b}\right)^a}$$

$$-E\left(\frac{\partial^2 \ln L}{\partial q \partial a}\right) = E\left[y \ln\left(\frac{x}{b}\right)\right]$$

$$= \frac{1}{a}\left\{E\left[y \ln y - y\ln(1-y)\right]\right\}$$

$$= \frac{1}{a}\left\{\left(\frac{p}{p+q}\right)\left[\psi(p)-\psi(q)\right] + \frac{1}{p+q}\right\}$$

$$= \frac{1}{a(p+q)}\left\{1 + p\left[\psi(p)-\psi(q)\right]\right\}$$

$$\frac{\partial^2 \ln L}{\partial a^2} = -\frac{1}{a^2} - (p+q)\ln\left(\frac{x}{b}\right)\left\{\frac{1}{1+\left(\frac{x}{b}\right)^a}\left(\frac{x}{b}\right)^a \ln\left(\frac{x}{b}\right) - \left(\frac{x}{b}\right)^a \frac{1}{\left[1+\left(\frac{x}{b}\right)^a\right]^2}\left(\frac{x}{b}\right)^a \ln\left(\frac{x}{b}\right)\right\}$$

$$= -\frac{1}{a^2} - (p+q)\left[y\ln^2\left(\frac{x}{b}\right) - y^2\ln^2\left(\frac{x}{b}\right)\right]$$

From (1.12) and (1.13),

$$-E\left(\frac{\partial^2 \ln L}{\partial a^2}\right)=\frac{1}{a^2}+(p+q)\left\{E\left[y\ln^2\left(\frac{x}{b}\right)\right]-E\left[y^2\ln^2\left(\frac{x}{b}\right)\right]\right\}$$

$$=\frac{1}{a^2}+\left(\frac{p}{a^2}\right)\left\{\left[\psi(p)-\psi(q)\right]^2+\psi'(p)+\psi'(q)+\frac{2}{p}\left[\psi(p)-\psi(q)\right]\right\}$$

$$-\left(\frac{p}{a^2}\right)\left(\frac{p+1}{p+q+1}\right)\left\{\left[\psi(p)-\psi(q)\right]^2+\psi'(p)+\psi'(q)+\frac{2}{p(p+1)}+\frac{2(2p+1)}{p(p+1)}\left[\psi(p)-\psi(q)\right]\right\}$$

$$=\frac{1}{a^2}\left[1+\left(\frac{pq}{p+q+1}\right)\left\{\left[\psi(p)-\psi(q)\right]^2+\psi'(p)+\psi'(q)\right\}+\frac{2(q-p)}{p+q+1}\left[\psi(p)-\psi(q)\right]\right.$$

$$\left.-\frac{2}{p+q+1}\right]$$

$$=\frac{1}{a^2}\left\{1+\left(\frac{1}{p+q+1}\right)\left[pq\left\{\left[\psi(p)-\psi(q)\right]^2+\psi'(p)+\psi'(q)\right\}+2(q-p)\left[\psi(p)-\psi(q)\right]-2\right]\right\}$$

Kleiber and Kotz (2003, p.194) report this result as

$$-E\left(\frac{\partial^2 \ln L}{\partial a^2}\right)=\frac{1}{a^2(p+q+1)}\left\{1+p+q+pq\left[\psi'(p)+\psi'(q)+\left(\psi(q)-\psi(p)+\frac{p-q}{pq}\right)^2-\frac{p^2+q^2}{p^2q^2}\right]\right\}$$

We can establish the equivalence of the two results by noting that

$$pq\left(2\left[\psi(q)-\psi(p)\right]\left(\frac{p-q}{pq}\right)\right)=2(q-p)\left[\psi(p)-\psi(q)\right]$$

and

$$pq\left\{\left[\frac{(p-q)^2}{p^2q^2}\right]-\frac{p^2+q^2}{p^2q^2}\right\}=-2$$

$$\frac{\partial^2 \ln L}{\partial b^2}=\frac{a}{b}\left\{-(p+q)\left[\left(\frac{b}{x}\right)^a+1\right]^{-2}\frac{ab^{a-1}}{x^a}\right\}-\frac{a}{b^2}\left\{(p+q)\left[\left(\frac{b}{x}\right)^a+1\right]^{-1}-p\right\}$$

$$=\frac{a^2}{b^2}\left\{-(p+q)\frac{\left(\frac{x}{b}\right)^a}{\left[\left(\frac{x}{b}\right)^a+1\right]^2}\right\}-\frac{a}{b^2}\left\{(p+q)\frac{\left(\frac{x}{b}\right)^a}{\left(\frac{x}{b}\right)^a+1}-p\right\}$$

$$=-\frac{a^2(p+q)y(1-y)}{b^2}-\frac{a(p+q)y}{b^2}+\frac{ap}{b^2}$$

$$-E\left(\frac{\partial^2 \ln L}{\partial b^2}\right)=\frac{a}{b^2}\left\{a(p+q)E\left[y(1-y)\right]+(p+q)E(y)-p\right\}$$

$$=\frac{a^2 pq}{b^2(p+q+1)}+\frac{ap}{b^2}-\frac{ap}{b^2}=\frac{a^2 pq}{b^2(p+q+1)}$$

$$\frac{\partial^2 \ln L}{\partial p \partial b}=\frac{a}{b}\left\{\frac{\left(\frac{x}{b}\right)^a}{\left[1+\left(\frac{x}{b}\right)^a\right]}-1\right\}=\frac{a}{b}(y-1)$$

$$-E\left(\frac{\partial^2 \ln L}{\partial p \partial b}\right)=\frac{a}{b}\left(1-\frac{p}{p+q}\right)=\frac{aq}{b(p+q)}$$

$$\frac{\partial^2 \ln L}{\partial q \partial b}=\frac{a}{b}\frac{\left(\frac{x}{b}\right)^a}{\left[1+\left(\frac{x}{b}\right)^a\right]}=\frac{a}{b}y$$

$$-E\left(\frac{\partial^2 \ln L}{\partial q \partial b}\right)=-\frac{ap}{b(p+q)}$$

$$-E\left(\frac{\partial^2 \ln L}{\partial p^2}\right)=-\frac{\partial^2 \ln L}{\partial p^2}=\psi'(p)-\psi'(p+q)$$

$$-E\left(\frac{\partial^2 \ln L}{\partial q \partial p}\right)=-\frac{\partial^2 \ln L}{\partial q \partial p}=-\psi'(p+q)$$

$$-E\left(\frac{\partial^2 \ln L}{\partial q^2}\right)=-\frac{\partial^2 \ln L}{\partial q^2}=\psi'(q)-\psi'(p+q)$$

S2 Estimation and Testing

For a single observation x_i, the log-likelihood function for the original parameterization is

$$\ln L_i=\ln a+ap\ln\left(\frac{x_i}{b}\right)-\ln x_i-\ln B(p,q)-(p+q)\ln\left[1+\left(\frac{x_i}{b}\right)^a\right]$$

For the log-metric parameterization where $\alpha=\ln(a)$, $\beta=\ln(b)$, $\gamma=\ln(p)$ and $\delta=\ln(q)$, the log likelihood for a single observation is

$$\ln L_i=\alpha+\exp(\alpha+\gamma)\ln\left(\frac{x_i}{\exp(\beta)}\right)-\ln x_i-\ln B\big(\exp(\gamma),\exp(\delta)\big)$$
$$-\left[\exp(\gamma)+\exp(\delta)\right]\ln\left[1+\left(\frac{x_i}{\exp(\beta)}\right)^{\exp(\alpha)}\right]$$

S2.1 Wald Tests

Let $\hat{\theta} = \left(\hat{a}\,\hat{b}\,\hat{p}\,\hat{q}\right)'$ be the maximum likelihood estimator for $\theta = \left(a\,b\,p\,q\right)'$. To describe a general form of the Wald statistic used for testing the special cases, we first note that

$$\sqrt{N}\left(\hat{\theta} - \theta\right) \xrightarrow{d} N\left(\mathbf{0}, \mathfrak{I}^{-1}\right)$$

where $\xrightarrow{d}$ denotes converges in distribution. It then follows that, for linear functions of the form $\mathbf{R}\theta$,

$$N\left(\hat{\theta} - \theta\right)' \mathbf{R}'\left(\mathbf{R}\mathfrak{I}^{-1}\mathbf{R}'\right)^{-1} \mathbf{R}\left(\hat{\theta} - \theta\right) \xrightarrow{d} \chi^2_{(J)}$$

where J is the row dimension and rank of $\mathbf{R}$. For testing the null hypothesis $H_0 : \mathbf{R}\theta = \mathbf{r}$, the Wald test statistic for testing special cases is defined by suitable choices of $\mathbf{R}$ and $\mathbf{r}$ and by replacing $\mathfrak{I}$ by a consistent estimator $\hat{\mathfrak{I}}$ in the statistic

$$W = N\left(\mathbf{R}\hat{\theta} - \mathbf{r}\right)'\left(\mathbf{R}\mathfrak{I}^{-1}\mathbf{R}'\right)^{-1}\left(\mathbf{R}\hat{\theta} - \mathbf{r}\right)$$

which has a limiting $\chi^2_{(J)}$-distribution when H_0 is true. See, for example, Greene (2018, p.556). Three alternative estimators for the information matrix are common in software: those obtained by replacing unknown parameters by their maximum likelihood estimators in (1) the information matrix, (2) the negative Hessian of the log-likelihood function $-N^{-1}\sum_{i=1}^{N}\partial^2 \ln L_i \big/ \partial\theta\partial\theta'$, and (3) the outer-product of gradients of the log-likelihood function $\mathbf{G}'\mathbf{G} = N^{-1}\sum_{i=1}^{N}\mathbf{g}_i\mathbf{g}_i'$, where $\mathbf{G}' = \left(\mathbf{g}_1 \ldots \mathbf{g}_N\right)$ and $\mathbf{g}_i = \partial \ln L_i \big/ \partial\theta$. The large-sample approximate distribution for $\hat{\theta}$ is $\hat{\theta} \sim N\left(\theta, \mathfrak{I}^{-1}\big/N\right)$. We refer to the covariance matrix of this distribution as the asymptotic variance of $\hat{\theta}$ and abbreviate it as $\operatorname{var}\left(\hat{\theta}\right) = \mathfrak{I}^{-1}\big/N$. In our Monte Carlo experiments, we estimate power of the Wald test statistics using all three covariance matrix estimators. They led to similar results and so only those for the Hessian, the most common estimator used in software, are reported.

Maximum likelihood estimates are invariant under one-to-one transformations of the parameters, but Wald tests are not. One transformation that turns out to be convenient both computationally and because of the power properties of the resulting Wald tests is a log-transformation. That is, instead of considering Wald tests for θ, we consider Wald tests for $\phi = \left(\ln(a)\,\ln(b)\,\ln(p)\,\ln(q)\right)'$. Alternatives to the null hypotheses $H_0 : p = 1$, $H_0 : q = 1$, $H_0 : a = 1$ and $H_0 : p = q = 1$ are $H_0 : \ln(p) = 0$, $H_0 : \ln(q) = 0$, $H_0 : \ln(a) = 0$, and $H_0 : \ln(p) = \ln(q) = 0$. Using the delta method, the limiting distribution for the maximum likelihood estimator $\hat{\phi}$ is $\sqrt{N}\left(\hat{\phi} - \phi\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{D}\mathfrak{I}^{-1}\mathbf{D}\right)$ where

$\mathbf{D} = \text{diag}\left(a^{-1}\, b^{-1}\, p^{-1}\, q^{-1}\right)$. We write the large-sample approximate variance of $\hat{\boldsymbol{\phi}}$ as $\text{var}\left(\hat{\boldsymbol{\phi}}\right) = \mathbf{D}\mathfrak{I}^{-1}\mathbf{D}/N$. The corresponding alternative Wald statistic is

$$W = N\left(\mathbf{R}\hat{\boldsymbol{\phi}}\right)'\left(\mathbf{R}\hat{\mathbf{D}}\hat{\mathfrak{I}}^{-1}\hat{\mathbf{D}}\mathbf{R}'\right)^{-1}\mathbf{R}\hat{\boldsymbol{\phi}}$$

where $\hat{\mathbf{D}} = \text{diag}\left(\hat{a}^{-1}\, \hat{b}^{-1}\, \hat{p}^{-1}\, \hat{q}^{-1}\right)$.

S2.2 Lagrange Multiplier Tests

LM test statistics are based on the limiting distribution of the gradient vector evaluated at the restricted ML estimators, those obtained assuming the null hypothesis is true. The average gradient vector $\mathbf{g} = N^{-1}\sum_{i=1}^{N}\mathbf{g}_i = N^{-1}\sum_{i=1}^{N}\partial \ln L_i/\partial\boldsymbol{\theta}$ has limiting distribution $\sqrt{N}\mathbf{g} \xrightarrow{d} \left(0, \mathfrak{I}\right)$. Because $\mathbf{g} = \mathbf{0}$ when evaluated at the unrestricted ML estimates, when it is evaluated at the restricted ML estimates, large deviations of $\mathbf{g}$ from zero suggest the null hypothesis is false. To describe a general form of an LM test statistic, let $\tilde{\boldsymbol{\theta}}$ denote the restricted ML estimator, let $\tilde{\mathbf{g}}_i = \partial \ln L_i/\partial\boldsymbol{\theta}\big|_{\tilde{\boldsymbol{\theta}}}$ denote the gradient vector for the ith observation evaluated at $\tilde{\boldsymbol{\theta}}$, let the corresponding matrix of all gradients be given by $\tilde{\mathbf{G}}' = \left(\tilde{\mathbf{g}}_1 \ldots \tilde{\mathbf{g}}_N\right)$, and denote their average as $\tilde{\mathbf{g}} = N^{-1}\sum_{i=1}^{N}\tilde{\mathbf{g}}_i$. A general expression for an LM test statistic is

$$\text{LM} = N\,\tilde{\mathbf{g}}'\tilde{\mathfrak{I}}^{-1}\tilde{\mathbf{g}} \tag{2.1}$$

where $\tilde{\mathfrak{I}}$ is an estimate of the information matrix obtained using the restricted estimator $\tilde{\boldsymbol{\theta}}$. Both the negative Hessian and the outer product of the gradients are possibilities for computing an estimate $\tilde{\mathfrak{I}}$. The LM statistic in (10) has a limiting $\chi^2_{(J)}$-distribution when H_0 is true. See, for example, Greene (2018, p.558).

The LM test statistic for testing $H_0: p = 1$ use the restricted estimator $\tilde{\boldsymbol{\theta}}' = \left(\tilde{a}\,\tilde{b}\,1\,\tilde{q}\right)$. Also, $\tilde{\mathbf{g}}' = \left(0\;0\;\tilde{g}_3\;0\right)$ and the LM statistic in equation (2.1) is

$$\text{LM} = N\tilde{g}_3\tilde{\mathfrak{I}}^{pp}$$

From results in Section S1, and using the digamma function $\Psi\left(\bullet\right)$,

$$\tilde{g}_3 = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial \ln L_i}{\partial p}\bigg|_{\tilde{\boldsymbol{\theta}}} = \psi(\tilde{q}+1) - \psi(1) + \frac{1}{N}\sum_{i=1}^{N}\ln\left[\frac{\left(x_i/\tilde{b}\right)^{\tilde{a}}}{1+\left(x_i/\tilde{b}\right)^{\tilde{a}}}\right]$$

The LM test statistic for testing $H_0: q=1$ uses the restricted estimator $\tilde{\boldsymbol{\theta}}' = \left(\tilde{a}\,\tilde{b}\,\tilde{p}\,1\right)$. Also, $\tilde{\mathbf{g}}' = \left(0\,0\,0\,\tilde{g}_4\right)$ and the LM statistic in equation (2.1) is

$$\text{LM} = N\tilde{g}_4\tilde{\Im}^{qq}$$

From results in Section S1,

$$\tilde{g}_4 = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial \ln L_i}{\partial q}\bigg|_{\tilde{\boldsymbol{\theta}}} = \psi(\tilde{p}+1) - \psi(1) - \frac{1}{N}\sum_{i=1}^{N}\ln\left[1+\left(\frac{x_i}{\tilde{b}}\right)^{\tilde{a}}\right]$$

The LM test statistic for testing $H_0: a=1$ uses the restricted estimator $\tilde{\boldsymbol{\theta}}' = \left(1\,\tilde{b}\,\tilde{p}\,\tilde{q}\right)$. Also, $\tilde{\mathbf{g}}' = \left(\tilde{g}_1\;0\;0\;0\right)$ and the LM statistic in equation (2.1) becomes

$$\text{LM} = N\tilde{g}_1\tilde{\Im}^{aa}$$

From results in Section S1,

$$\tilde{g}_1 = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial \ln L_i}{\partial a}\bigg|_{\tilde{\boldsymbol{\theta}}} = 1 + \tilde{p}\frac{1}{N}\sum_{i=1}^{N}\ln\left(\frac{x_i}{\tilde{b}}\right) - (\tilde{p}+\tilde{q})\frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i/\tilde{b}}{1+x_i/\tilde{b}}\right)\ln\left(\frac{x_i}{\tilde{b}}\right)$$

Details of the LM test for the Fisk distribution, $H_0: p=q=1$ are provided in Section S9.2

S3 Distributions of test statistics under the alternative hypothesis

We first consider the limiting distribution of the Wald statistic $W = N\left(\mathbf{R}\hat{\boldsymbol{\theta}}-\mathbf{r}\right)'\left(\mathbf{R}\hat{\Im}^{-1}\mathbf{R}'\right)^{-1}\left(\mathbf{R}\hat{\boldsymbol{\theta}}-\mathbf{r}\right)$ under the local alternative $H_1: \mathbf{R}\boldsymbol{\theta}-\mathbf{r} = \boldsymbol{\delta}/\sqrt{N}$ and then compare some empirical distributions of test statistics with central and noncentral chi-square distributions.

S3.1 Limiting distribution of Wald statistic under a local alternative

Consider testing $H_0: \mathbf{R}\boldsymbol{\theta} = \mathbf{r}$ against the local alternatives $H_1: \mathbf{R}\boldsymbol{\theta}-\mathbf{r} = \boldsymbol{\delta}/\sqrt{N}$. Under these circumstances, $\mathbf{R}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)$ becomes $\mathbf{R}\hat{\boldsymbol{\theta}}-\mathbf{r}-\boldsymbol{\delta}/\sqrt{N}$ and

$$\sqrt{N}\left(\mathbf{R}\hat{\boldsymbol{\theta}}-\mathbf{r}-\boldsymbol{\delta}/\sqrt{N}\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{R}\Im^{-1}\mathbf{R}'\right)$$

or

$$\sqrt{N}\left(\mathbf{R}\Im^{-1}\mathbf{R}'\right)^{-1/2}\left(\mathbf{R}\hat{\boldsymbol{\theta}}-\mathbf{r}\right) - \left(\mathbf{R}\Im^{-1}\mathbf{R}'\right)^{-1/2}\boldsymbol{\delta} \xrightarrow{d} N\left(\mathbf{0},\mathbf{I}\right)$$

where $\left(\mathbf{R}\Im^{-1}\mathbf{R}'\right)^{-1/2}$ is a matrix such that $\left(\mathbf{R}\Im^{-1}\mathbf{R}'\right)^{-1/2}\left(\mathbf{R}\Im^{-1}\mathbf{R}'\right)^{-1/2}=\left(\mathbf{R}\Im^{-1}\mathbf{R}'\right)^{-1}$ and hence $\left(\mathbf{R}\Im^{-1}\mathbf{R}'\right)^{-1/2}\mathbf{R}\Im^{-1}\mathbf{R}'\left(\mathbf{R}\Im^{-1}\mathbf{R}'\right)^{-1/2}=\mathbf{I}$. It follows that

$$\sqrt{N}\left(\mathbf{R}\hat{\Im}^{-1}\mathbf{R}'\right)^{-1/2}\left(\mathbf{R}\hat{\theta}-\mathbf{r}\right)\xrightarrow{d}N\left(\left(\mathbf{R}\Im^{-1}\mathbf{R}'\right)^{-1/2}\delta,\mathbf{I}\right) \tag{3.1}$$

where $\hat{\Im}$ is one of the three possible consistent estimators for $\Im$. From (3.1), the Wald statistic has the following limiting distribution

$$W=N\left(\mathbf{R}\hat{\theta}-\mathbf{r}\right)'\left(\mathbf{R}\hat{\Im}^{-1}\mathbf{R}'\right)^{-1}\left(\mathbf{R}\hat{\theta}-\mathbf{r}\right)\xrightarrow{d}\chi^2_{(J,\lambda)}$$

where $\chi^2_{(J,\lambda)}$ is a noncentral chi-square distribution with noncentrality parameter $\lambda=\delta'\left(\mathbf{R}\Im^{-1}\mathbf{R}'\right)^{-1}\delta$. To obtain an approximate noncentrality parameter for a finite sample size *N*, we replace δ with $\sqrt{N}\left(\mathbf{R}\theta-\mathbf{r}\right)$, leading to the asymptotic local noncentrality parameter

$$\lambda=N\left(\mathbf{R}\theta-\mathbf{r}\right)'\left(\mathbf{R}\Im^{-1}\mathbf{R}'\right)^{-1}\left(\mathbf{R}\theta-\mathbf{r}\right)$$

An asymptotic local power function can be obtained by using the noncentral chi-square distribution to compute $\Pr\left(W>\chi^2_{(J),c}\right)$ where $\chi^2_{(J),c}$ is a chosen critical value from the central chi-square distribution.

S3.2 A comparison of empirical distributions of Wald statistics with chi-square distributions

Writing the alternative hypothesis as $H_1:\mathbf{R}\theta-\mathbf{r}=\delta/\sqrt{N}$ is an artificial special case. In practice violations of the null hypothesis $H_0:\mathbf{R}\theta=\mathbf{r}$ cannot be written as a decreasing function of sample size, and the Wald statistic no longer has a noncentral chi-square distribution. We illustrate this fact in the following figures. We compare empirical distributions of Wald statistics with their noncentral chi-square counterparts that would be relevant under the local alternatives' assumption. We use the inverse of the negative Hessian, $-\left(N^{-1}\sum_{i=1}^{N}\partial^2\ln L_i/\partial\theta\partial\theta'\right)^{-1}$, to estimate $\Im^{pp}$ because it is a common choice in commercial software. The hypothesis we consider is $H_0:p=1$ against the alternative $H_1:p\neq 1$ for $p=(0.75,1,1.25,1.75,2.25)$. Results are presented for two sample sizes $N=10{,}000$ and $N=100{,}000$; the empirical distributions are based on 10,000 samples. For $p=1$, the empirical distributions resemble the central chi-square distribution, as they should. For all other values of *p* there is a vast difference, particularly for values of p further from the null hypothesis. In all cases a = 5, b = 1, and q = 1.75.

Consider Figure S1(a) where three distributions are plotted for $p = 0.75$ and $N = 10{,}000$: the histogram is the Monte-Carlo based distribution; the two curves are noncentral chi-square distributions, one with local noncentrality parameter

$$\lambda = N(\mathbf{R}\boldsymbol{\theta} - \mathbf{r})'\left(\mathbf{R}\mathfrak{I}^{-1}\mathbf{R}'\right)^{-1}(\mathbf{R}\boldsymbol{\theta} - \mathbf{r}) = N(p-1)^2 / \mathfrak{I}^{pp} \cong 13.62$$

and the other with noncentrality parameter equal to 15.0, the Monte Carlo average of the 10,000 Wald statistic values. Neither of the noncentral distributions is a good fit. Denote the 0.95 percentile of the $\chi^2_{(1)}$ distribution as $\chi^2_{(0.95,1)} \cong 3.841$. When the true p = 0.75, and N = 10,000, 89.23% of the Wald statistic values exceed the $\alpha = 0.05$ critical value of the $\chi^2_{(1)}$ distribution.

When p = 1, Figure S1(b), the central chi-square distribution fits the histogram well and the percentage rejections, 4.71%, is close to the nominal test size. As p increases (Figures S1(c), (d) and (e)), the percentage of null hypotheses rejected increases, and the fit of the noncentral chi-square distribution becomes worse. Indeed, for none of the Wald statistic cases does a noncentral chi-square distribution fit the data.

Now consider Figure S2, the cases in which the sample size is N = 100,000. In all cases except when p = 1, we reject the null hypothesis in 100% of the samples. When p = 1 the central chi-square density fits the histogram well, and the 5.04% rejections is close to the nominal size of the test. As the values of p increase to 1.25, 1.75 and 2.25 the noncentral chi-square approximation becomes worse and worse. We show only one figure here because when N = 100,000 the average Wald statistic is so close to the hypothetical noncentrality parameter that there is no visible difference in the graphs. For values of p < 1 the Wald statistic has the largest average test statistic and a larger Monte Carlo power. We do not recommend this test as the first choice.

Figure S1(a) Wald statistic histogram and densities when $p = 0.75$ and $N = 10{,}000$.

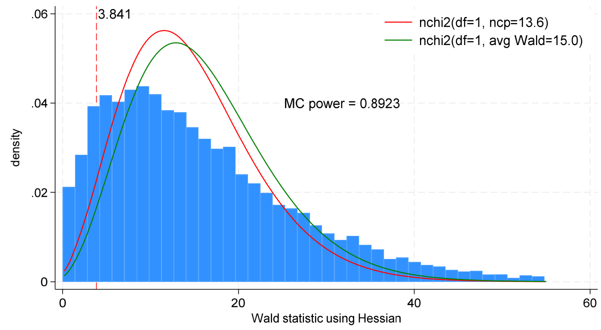


Figure S1(b) Wald statistic histogram and density when $p = 1$ and $N = 10{,}000$.

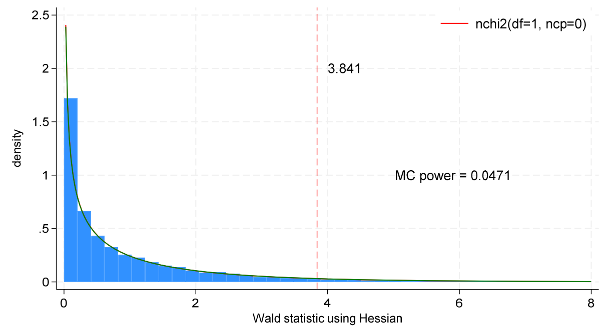


Figure S1(c) Wald statistic histogram and densities when $p = 1.25$ and $N = 10{,}000$.

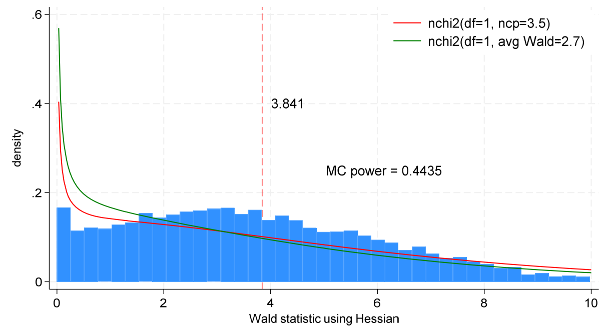


Figure S1(d) Wald statistic histogram and densities when $p = 1.75$ and $N = 10{,}000$.

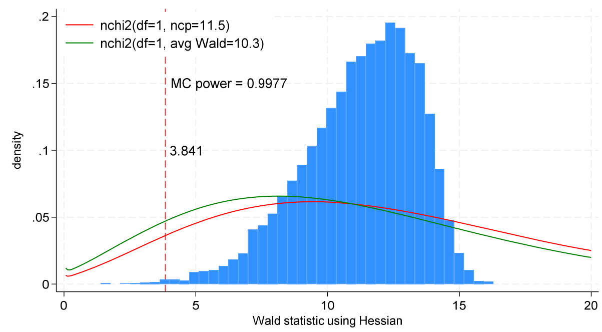
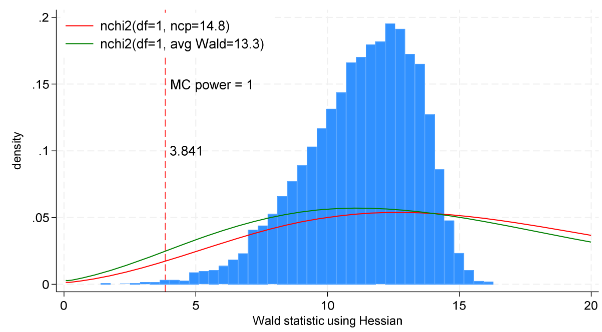


Figure S1(e) Wald statistic histogram and densities when $p = 2.25$ and $N = 10{,}000$.



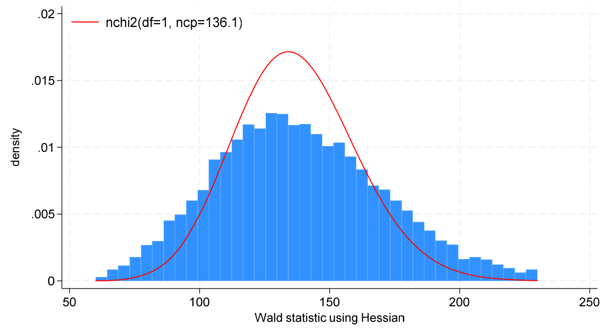


Figure S2(a) Wald statistic histogram and density when $p = 0.75$ and $N = 100{,}000$.

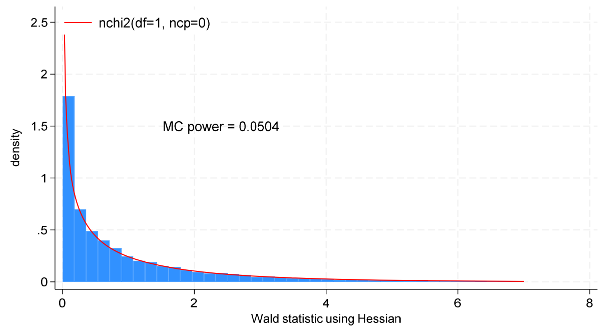


Figure S2(b) Wald statistic histogram and density when $p = 1$ and $N = 100{,}000$.

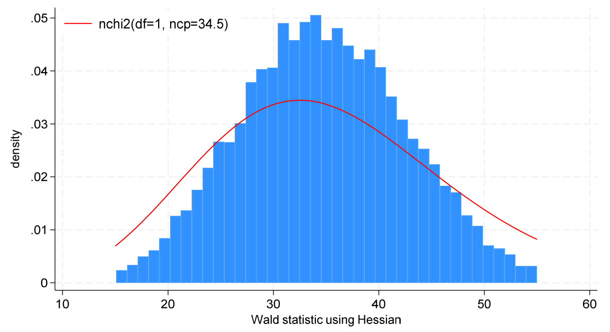


Figure S2(c) Wald statistic histogram and density when $p = 1.25$ and $N = 100{,}000$.

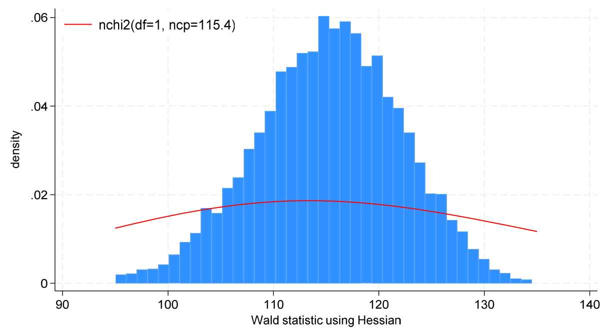


Figure S2(d) Wald statistic histogram and density when $p = 1.75$ and $N = 100{,}000$.

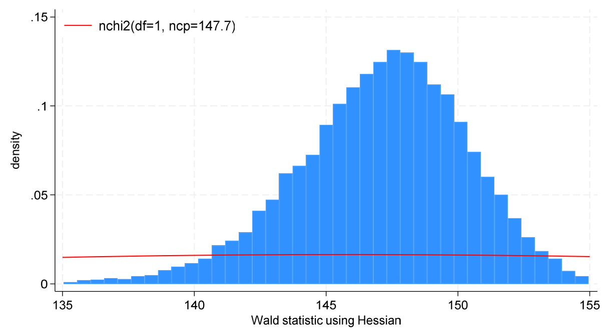


Figure S2(e) Wald statistic histogram and density when $p = 2.25$ and $N = 100{,}000$.

S3.3 A comparison of empirical distributions of Wald log-metric statistics with chi-square distributions

Consider the log-metric case in which we test the null hypothesis $H_0 : \ln(p) = 0$ against the alternative $H_1 : \ln(p) \neq 0$. Here the Wald statistic uses the Hessian with the delta method to obtain the estimator covariance matrix. We display the histograms and noncentral chi-square distributions when $N = 10{,}000$ in Figure S3. When $p = 0.75$, Figure S3(a), the Wald log-metric test rejects the null hypothesis in 86.12% of the 10,000 Monte Carlo samples. This is slightly less than the Wald test power shown in Figure S1(a). We show only the noncentral chi-square density with the hypothetical noncentrality parameter because the average of the Wald log-metric statistic values is virtually identical to the hypothetical value.[1] It is interesting to note that the density fits the data rather well. When $p = 1$, in Figure S3(b), the central chi-square density fits the Monte Carlo data well, and the Monte Carlo power, the percentage of rejections, is closer to the nominal 5% value than for the Wald statistic. Figures S3(c)-S3(e) show the cases for $p = 1.25$, 1.75 and 2.25. The densities using the hypothetical noncentrality parameters and the average test statistic values are very close to each other. The fit is better in the $p = 1.25$ case than in the others. Overall, based on casual observation, the noncentral chi-square distributions fit the empirical results better using the log-metric Wald test compared to the Wald test in Section S2.2.

We show the results for $N = 100{,}000$ in Figure S4. If $p \neq 1$, all the tests using the Wald log-metric test statistic reject the null hypothesis that $\ln(p) = 0$. When $p = 1$ the Wald log-metric test rejects the null hypothesis in 4.91% of the samples and the values are well fit by the central chi-square distribution. Once again, we show only the noncentral chi-square density with the hypothetical noncentrality parameter because the averages of the Wald log-metric statistics are very close to the hypothetical values. It is very interesting that for $p = 0.75$, 1, and 1.25 the data are well fit by the noncentral chi-square distribution.

---

[1] See Table S3 for these details.

Figure S3(a) Wald log-metric statistic histogram and densities when $p = 0.75$ and $N = 10{,}000$.

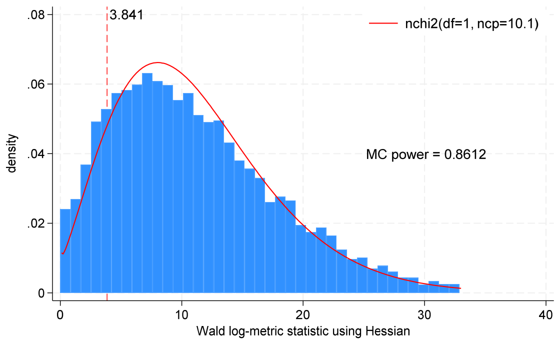


Figure S3(b) Wald log-metric statistic histogram and density when $p = 1$ and $N = 10{,}000$.

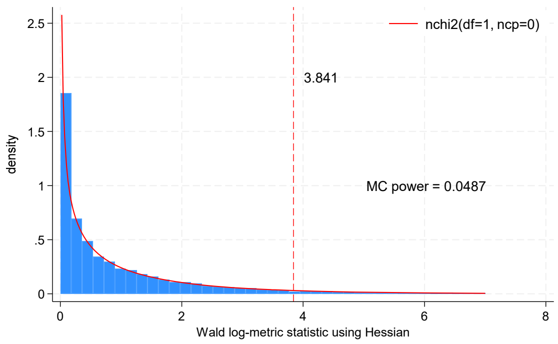


Figure S3(c) Wald log-metric statistic histogram and densities when $p = 1.25$ and $N = 10{,}000$.

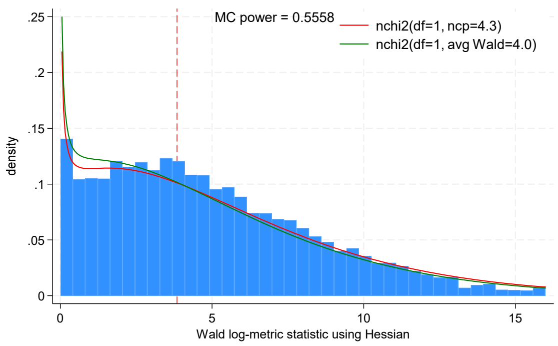


Figure S3(d) Wald log-metric statistic histogram and densities when $p = 1.75$ and $N = 10{,}000$.

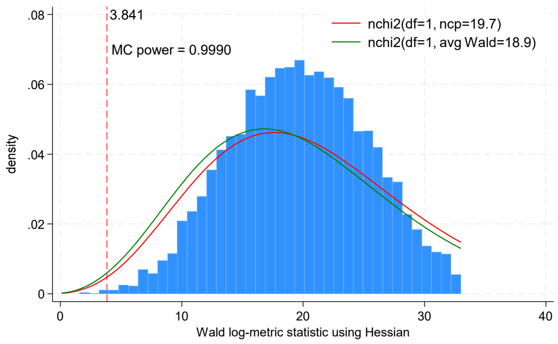
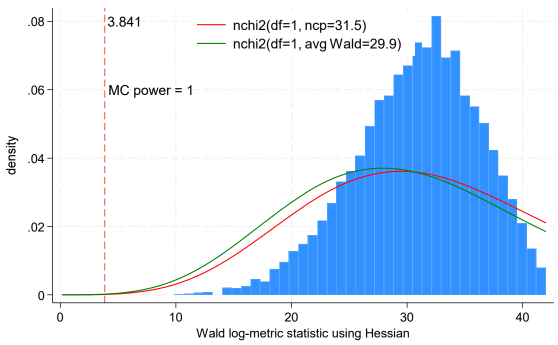
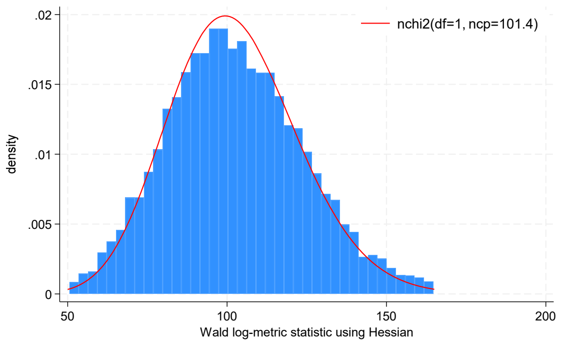
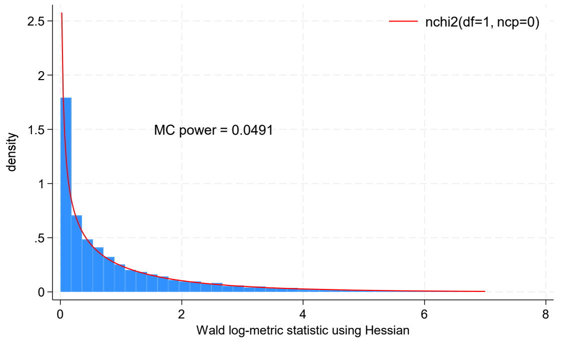
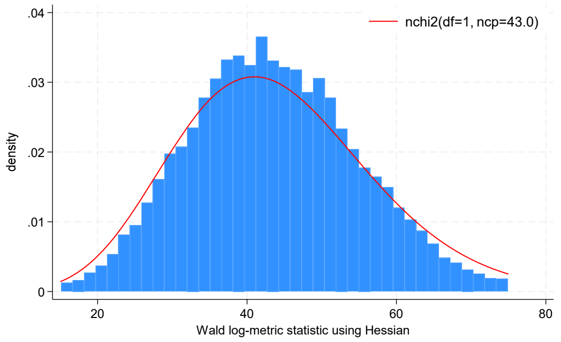
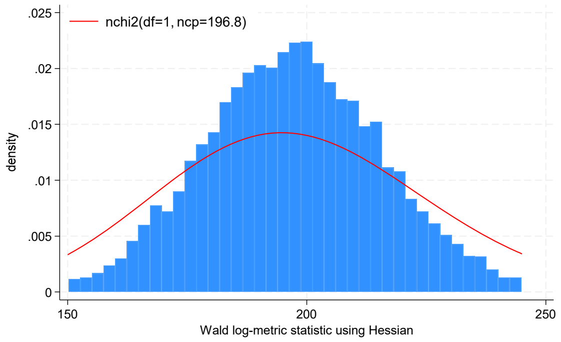
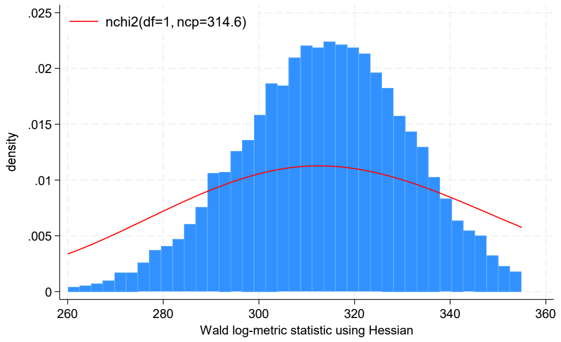


Figure S3(e) Wald log-metric statistic histogram and densities when $p = 2.25$ and $N = 10{,}000$.



<table>
<tr><td>Figure S4(a) Wald log-metric statistic histogram and densities when $p = 0.75$ and $N = 100{,}000$.</td><td>
</td></tr>
<tr><td>Figure S4(b) Wald log-metric statistic histogram and density when $p = 1$ and $N = 100{,}000$.</td><td>
</td></tr>
<tr><td>Figure S4(c) Wald log-metric statistic histogram and density when $p = 1.25$ and $N = 100{,}000$.</td><td>
</td></tr>
<tr><td>Figure S4(d) Wald log-metric statistic histogram and density when $p = 1.75$ and $N = 100{,}000$.</td><td>
</td></tr>
<tr><td>Figure S4(e) Wald log-metric statistic histogram and density when $p = 2.25$ and $N = 100{,}000$.</td><td>
</td></tr>
</table>

S3.4 A comparison of empirical distributions of likelihood ratio test statistics with chi-square distributions

The likelihood ratio test is invariant to the log-transformation. We show the histograms and fitted densities for the case with $N = 10{,}000$ observations in Figure S5. In this case the only noncentral chi-square density with the hypothetical noncentrality parameter that fits the data well is when $p = 1$. For other values of p, the noncentral chi-square density with the hypothetical noncentrality parameter does not fit the data at all. However, the noncentral chi-square density with a noncentrality parameter set to the average of the test statistic values fits the data quite well. Referring to Table S1, when $p = 0.75$, the Monte Carlo power of the likelihood ratio test is less than that of both the original and log-metric versions of Wald tests. For values of $p > 1$ the likelihood ratio test has larger Monte Carlo power than the Wald tests.

If $N = 100{,}000$, Figure S6, the same conclusions about the fit of the noncentral chi-square distributions are reached. In all cases other than $p = 1$ the Monte Carlo power is 1.0. The last column of Table S1 shows the Monte Carlo power for the various tests when $p = 1$, which can be compared to the $N = 10{,}000$ case shown in the second column. The 5.03% rejection rate is close to the nominal 5% value. For values of $p > 1$ the likelihood ratio test has a larger average value than the other test statistics. See Tables S3 and S4.

Based on these results we recommend that the likelihood ratio test is a better choice than a Wald test for the null hypothesis $p = 1$ or $\ln(p) = 0$.

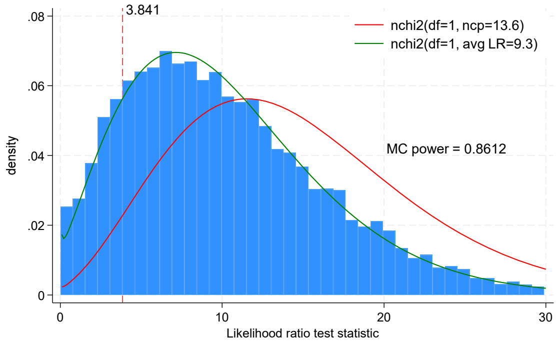
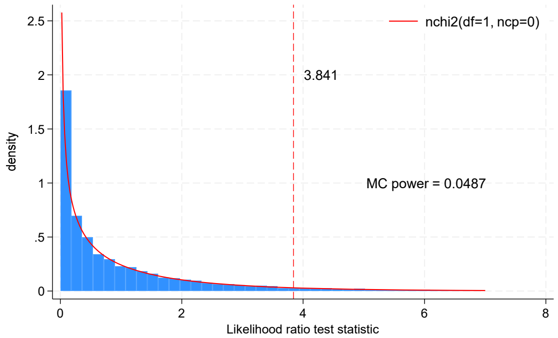
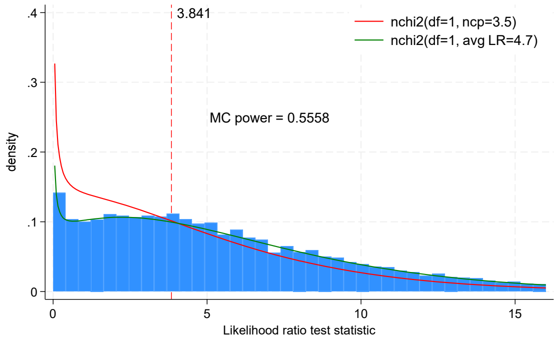
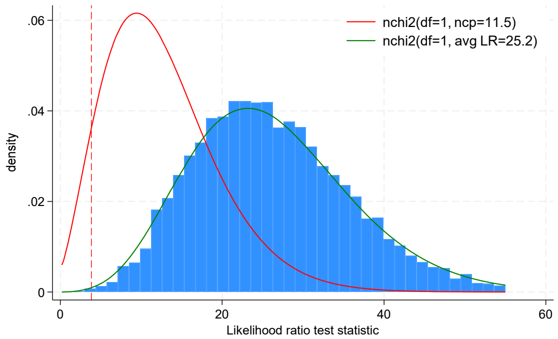
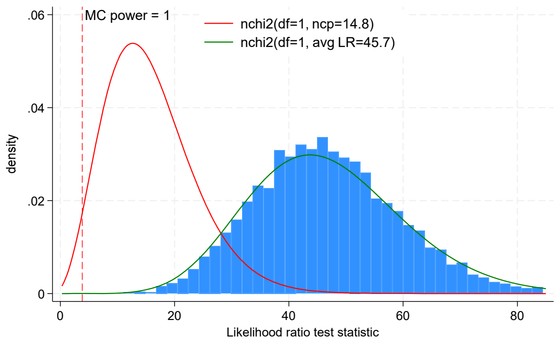
Figure S5(a) Likelihood ratio statistic histogram and densities when $p = 0.75$ and $N = 10{,}000$.	
Figure S5(b) Likelihood ratio statistic histogram and density when $p = 1$ and $N = 10{,}000$.	
Figure S5(c) Likelihood ratio statistic histogram and densities when $p = 1.25$ and $N = 10{,}000$.	
Figure S5(d) Likelihood ratio statistic histogram and densities when $p = 1.75$ and $N = 10{,}000$.	
Figure S5(e) Likelihood ratio statistic histogram and densities when $p = 2.25$ and $N = 10{,}000$.	

Figure S6(a) Likelihood ratio statistic histogram and densities when $p = 0.75$ and $N = 100{,}000$.

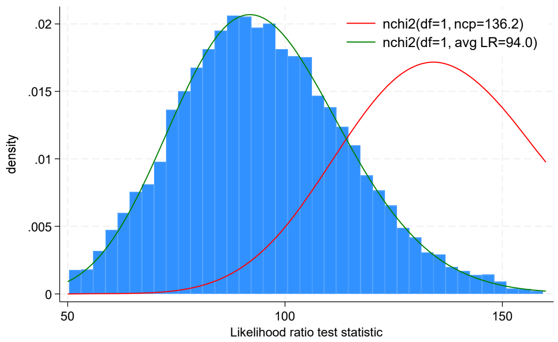


Figure S6(b) Likelihood ratio statistic histogram and density when $p = 1$ and $N = 100{,}000$.

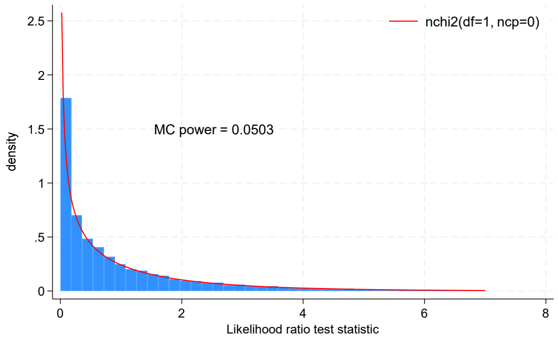


Figure S6(c) Likelihood ratio statistic histogram and densities when $p = 1.25$ and $N = 100{,}000$.

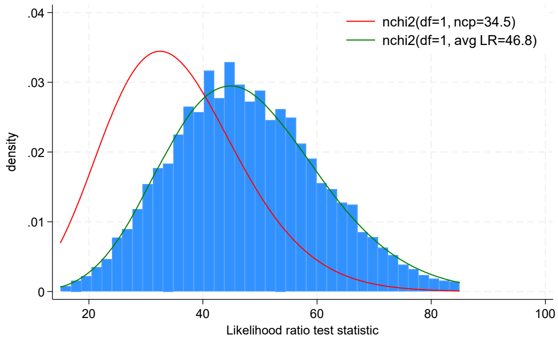


Figure S6(d) Likelihood ratio statistic histogram and densities when $p = 1.75$ and $N = 100{,}000$.

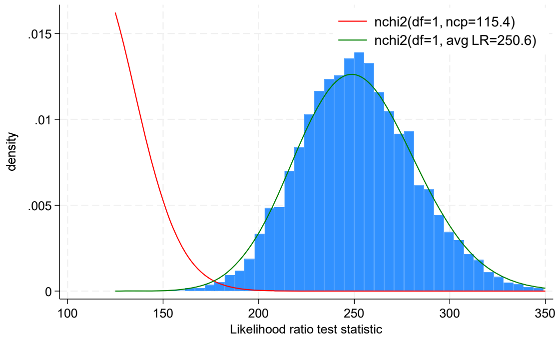
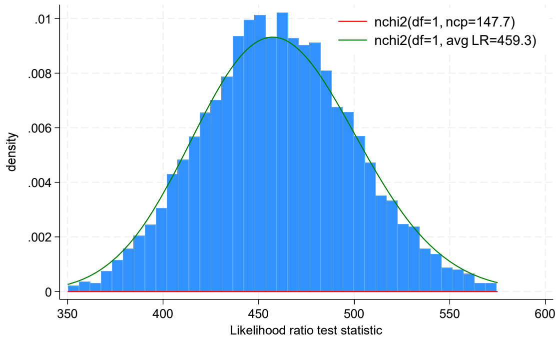


Figure S6(e) Likelihood ratio statistic histogram and densities when $p = 2.25$ and $N = 100{,}000$.



S3.5 A comparison of empirical distributions of Lagrange multiplier test statistics with chi-square distributions

The performance of the Lagrange multiplier test, Figure S7, is like the performance of the likelihood ratio test. The noncentral chi-square density with noncentrality parameter equal to the average of the test statistic fits the data very well. However, the fit using the hypothetical noncentrality parameter is not even close, except when $p = 1$. Refer to Table S1 to compare the Monte Carlo powers of the tests. We do not expect the LM test to be widely used in applications, and the likelihood ratio test is a better alternative.

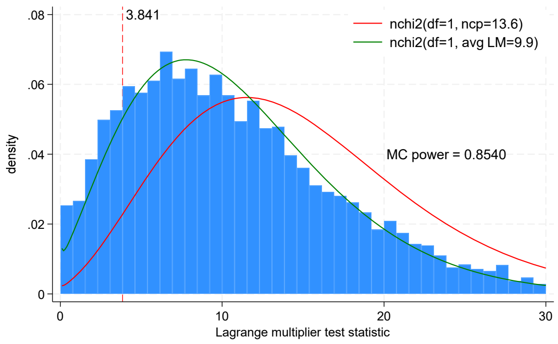


Figure S7(a) Lagrange multiplier statistic histogram and densities when $p = 0.75$ and $N = 10{,}000$.

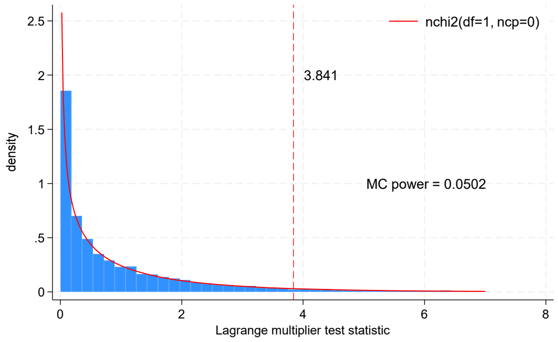


Figure S7(b) Lagrange multiplier statistic histogram and density when $p = 1$ and $N = 10{,}000$.

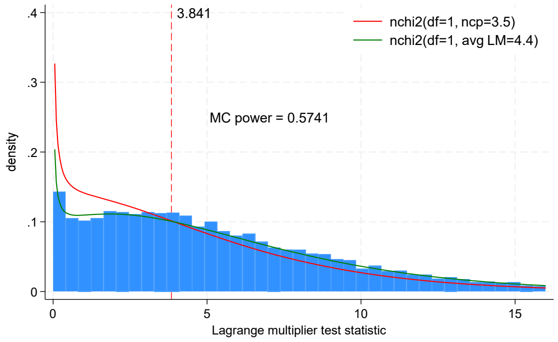


Figure S7(c) Lagrange multiplier statistic histogram and densities when $p = 1.25$ and $N = 10{,}000$.

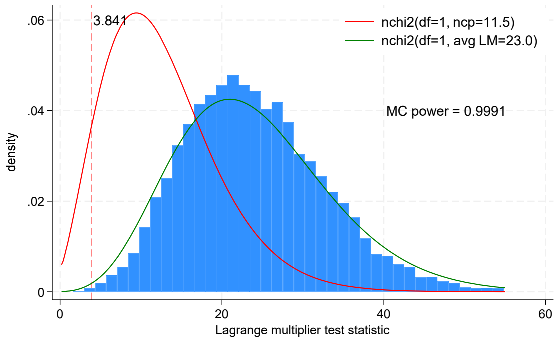


Figure S7(d) Lagrange multiplier statistic histogram and densities when $p = 1.75$ and $N = 10{,}000$.

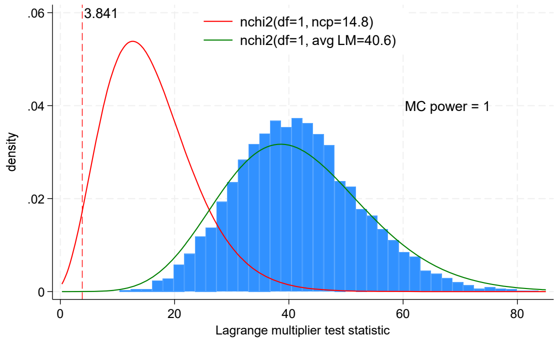
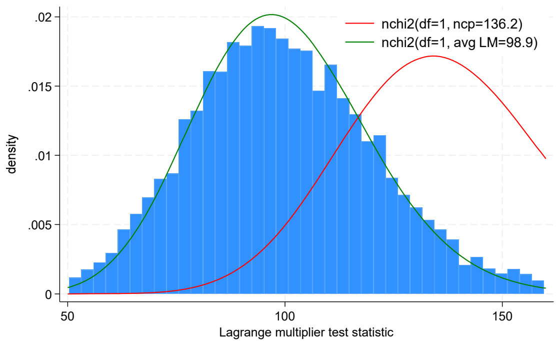
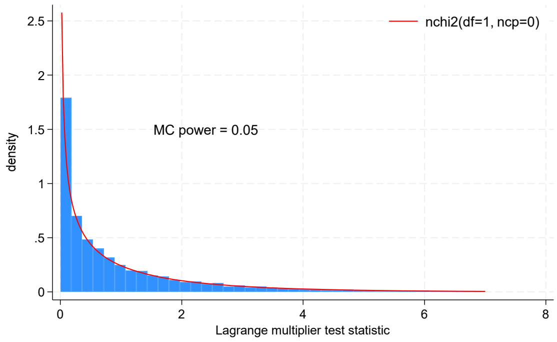
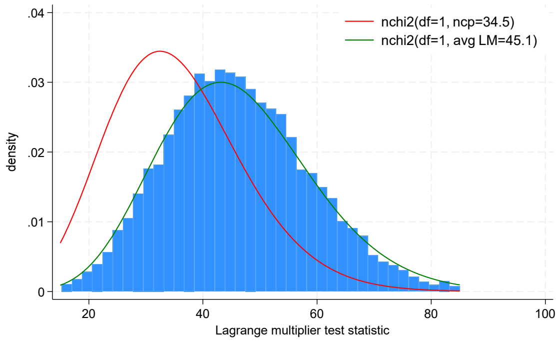
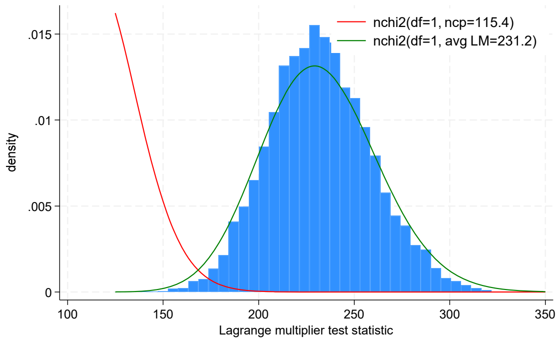
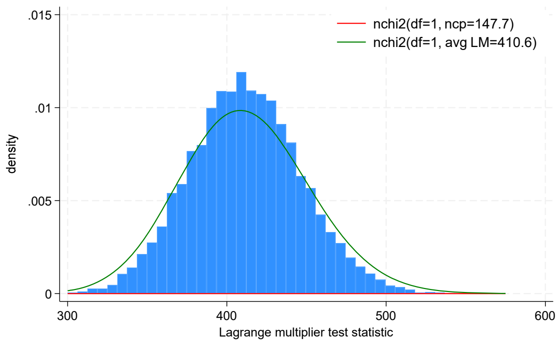


Figure S7(e) Lagrange multiplier statistic histogram and densities when $p = 2.25$ and $N = 10{,}000$.

Figure S8(a) Lagrange multiplier statistic histogram and densities when $p = 0.75$ and $N = 100{,}000$..	
Figure S8(b) Lagrange multiplier statistic histogram and density when $p = 1$ and $N = 100{,}000$.	
Figure S8(c) Lagrange multiplier statistic histogram and densities when $p = 1.25$ and $N = 100{,}000$.	
Figure S8(d) Lagrange multiplier statistic histogram and densities when $p = 1.75$ and $N = 100{,}000$.	
Figure S8(e) Lagrange multiplier statistic histogram and densities when $p = 2.25$ and $N = 100{,}000$.	

S3.6 Some Monte Carlo details

Table S1 Monte Carlo test rejection rates, N = 10,000

p	0.75	1	1.25	1.75	2.25	1
winfo	0.8925	0.0472	0.4458	0.9979	1	0.0503
winfo0	0.9375	0.0597	0.4641	0.9596	0.9922	0.0510
wopg	0.8901	0.0467	0.4417	0.9976	1	0.0503
whess	0.8923	0.0471	0.4435	0.9977	1	0.0504
winfoln	0.8618	0.0496	0.5576	0.999	1	0.0492
wopgln	0.8604	0.0482	0.5541	0.999	1	0.0496
whessln	0.8612	0.0487	0.5558	0.999	1	0.0491
wopglmet	0.8604	0.0482	0.5541	0.999	1	0.0496
lrlmet	0.8494	0.0516	0.5892	0.9993	1	0.0503
lrtest	0.8494	0.0516	0.5892	0.9993	1	0.0503
lmtest	0.8540	0.0502	0.5741	0.9991	1	0.0501
lmtest0	0.8435	0.0504	0.5798	0.9991	1	0.0500
lmopg	0.8308	0.0529	0.6123	0.9993	1	0.0507

Note: The last column contains the rejection rate when p = 1 and N = 100,000. Using N = 100,000, in all the other cases the tests reject the null hypothesis in 100% of the Monte Carlo samples.

Table S2 Test statistic abbreviations

p	Parameter value
winfo	Wald test using information matrix evaluated at the MLE
winfo0	Wald test using information matrix evaluated at the true parameter values
wopg	Wald test using OPG estimator of information matrix evaluated at the MLE
whess	Wald test using the Hessian evaluated at the MLE
winfoln	Wald log-metric test using information matrix evaluated at the MLE
wopgln	Wald log-metric test using OPG estimator evaluated at the MLE
whessln	Wald log-metric test using Hessian estimator evaluated at the MLE
wopglmet	Wald log-metric test using OPG with numerical 2nd derivatives evaluated at the MLE
lrlmet	Likelihood ratio test in the log-metric
lrtest	Likelihood ratio test
lmtest	Lagrange multiplier test using information matrix evaluated at the MLE
lmtest0	Lagrange multiplier test using information matrix with the true parameter values
lmopg	Lagrange multiplier test in the OPG form

Table S3 Noncentrality parameters and hypothetical power, N = 10,000

p	0.75	1	1.25	1.75	2.25
lam_hypo	13.6162	0.0000	3.4514	11.5423	14.7658
hypo_pwr	0.9582	0.0500	0.4594	0.9247	0.9701
lam_lmet	10.1420	0.0000	4.2964	19.6801	31.4608
pwr_lmet	0.8897	0.0500	0.5449	0.9934	0.9999
nc_wald	14.9455	0.0139	2.7431	10.2527	13.2609
nc_lmet	10.1874	-0.0001	4.0152	18.8526	29.9336
nc_lr	9.2879	0.0150	4.7015	25.1968	45.7352
nc_lm	9.9098	0.0111	4.4440	23.0316	40.6084

Table S4 Noncentrality parameters and hypothetical power, $N = 100{,}000$

p	0.75	1	1.25	1.75	2.25
lam_hypo	136.1618	0.0000	34.5140	115.4231	147.6583
hypo_pwr	1.0000	0.0500	1.0000	1.0000	1.0000
lam_lmet	101.4199	0.0000	42.9639	196.8005	314.6081
pwr_lmet	1.0000	0.0500	1.0000	1.0000	1.0000
nc_wald	138.1021	0.0107	33.8417	114.1785	146.1122
nc_lmet	101.8763	0.0097	42.7491	196.1077	312.9074
nc_lr	93.9764	0.0111	46.7847	250.6459	459.2857
nc_lm	98.9031	0.0107	45.1406	231.2009	410.5913

Table S5 Noncentrality and hypothetical power abbreviations

lam_hypo	Hypothetical noncentrality parameter
hypo_pwr	Power calculated using the hypothetical noncentrality parameter
lam_lmet	Hypothetical noncentrality parameter in the log-metric
pwr_lmet	Power calculated using the hypothetical log-metric noncentrality parameter
nc_wald	Average value of the Wald statistic
nc_lmet	Average value of the Wald log-metric statistic
nc_lr	Average value of the likelihood ratio test statistic
nc_lm	Average value of the Lagrange multiplier test statistic

S4 Covariance matrix for $\hat{p}$ and $\hat{q}$

Theorem 1

Let $f_i(p,q), i = 1,2,\ldots,10$ denote the following functions of p and q,

$$f_1 = f_1(p,q) = 1 + \left(\frac{1}{p+q+1}\right)\left[pq\left\{\left[\psi(p)-\psi(q)\right]^2 + \psi'(p) + \psi'(q)\right\} + 2(q-p)\left[\psi(p)-\psi(q)\right] - 2\right]$$

$$f_2 = f_2(p,q) = \frac{-pq\left[\psi(p)-\psi(q)\right] + p - q}{p+q+1}$$

$$f_3 = f_3(p,q) = \frac{1 - q\left[\psi(p)-\psi(q)\right]}{p+q}$$

$$f_4 = f_4(p,q) = \frac{1 + p\left[\psi(p)-\psi(q)\right]}{p+q}$$

$$f_5 = f_5(p,q) = \frac{pq}{p+q+1}$$

$$f_6 = f_6(p,q) = \frac{q}{p+q}$$

$$f_7 = f_7(p,q) = \frac{-p}{p+q}$$

$$f_8 = f_8(p,q) = \psi'(p) - \psi'(p+q)$$

$$f_9 = f_9(p,q) = -\psi'(p+q)$$

$$f_{10} = f_{10}(p,q) = \psi'(q) - \psi'(p+q)$$

so that the information matrix can be written as

$$\mathfrak{I} = \begin{pmatrix} f_1/a^2 & f_2/b & f_3/a & f_4/a \\ f_2/b & a^2 f_5/b^2 & af_6/b & af_7/b \\ f_3/a & af_6/b & f_8 & f_9 \\ f_4/a & af_7/b & f_9 & f_{10} \end{pmatrix}$$

The asymptotic covariance matrix for the maximum likelihood estimator $\sqrt{N}(\hat{p},\hat{q})$ is given by

$$\operatorname{cov}\left(\sqrt{N}(\hat{p},\hat{q})\right) = \left[\begin{pmatrix} f_8 & f_9 \\ f_9 & f_{10} \end{pmatrix} \right.$$

$$\left. - \frac{1}{f_1 f_5 - f_2^2} \begin{pmatrix} f_3^2 f_5 + f_1 f_6^2 - 2f_2 f_3 f_6 & f_3 f_4 f_5 + f_1 f_6 f_7 - f_2 f_3 f_7 - f_2 f_4 f_6 \\ f_3 f_4 f_5 + f_1 f_6 f_7 - f_2 f_3 f_7 - f_2 f_4 f_6 & f_4^2 f_5 + f_1 f_7^2 - 2 f_2 f_4 f_7 \end{pmatrix}\right]^{-1}$$

and hence does not depend on *a* or *b*.

Theorem 1 Proof

Let $\eta = (a,b)'$ and $\theta = (p,q)'$, and let the information matrix for a single observation be partitioned as

$$\mathfrak{I} = \begin{pmatrix} \mathfrak{I}_{aa} & \mathfrak{I}_{ab} & \mathfrak{I}_{ap} & \mathfrak{I}_{aq} \\ \mathfrak{I}_{ab} & \mathfrak{I}_{bb} & \mathfrak{I}_{bp} & \mathfrak{I}_{bq} \\ \mathfrak{I}_{ap} & \mathfrak{I}_{bp} & \mathfrak{I}_{pp} & \mathfrak{I}_{pq} \\ \mathfrak{I}_{aq} & \mathfrak{I}_{bq} & \mathfrak{I}_{pq} & \mathfrak{I}_{qq} \end{pmatrix} = \begin{pmatrix} \mathfrak{I}_{\eta\eta} & \mathfrak{I}_{\eta\omega} \\ \mathfrak{I}_{\omega\eta} & \mathfrak{I}_{\omega\omega} \end{pmatrix}$$

Then,

$$\operatorname{cov}\left(\sqrt{N}(\hat{p},\hat{q})\right) = \left(\mathfrak{I}_{\omega\omega} - \mathfrak{I}_{\omega\eta}\mathfrak{I}_{\eta\eta}^{-1}\mathfrak{I}_{\eta\omega}\right)^{-1}$$

where

$$\mathfrak{I}_{\eta\eta}^{-1} = \begin{pmatrix} \frac{1}{a^2} f_1 & \frac{1}{b} f_2 \\ \frac{1}{b} f_2 & \frac{a^2}{b^2} f_5 \end{pmatrix}^{-1} = \frac{b^2}{f_1 f_5 - f_2^2} \begin{pmatrix} \frac{a^2}{b^2} f_5 & -\frac{1}{b} f_2 \\ -\frac{1}{b} f_2 & \frac{1}{a^2} f_1 \end{pmatrix}$$

$$\mathfrak{I}_{\eta\eta}^{-1}\mathfrak{I}_{\eta\omega} = \frac{b^2}{f_1 f_5 - f_2^2}\begin{pmatrix} \frac{a^2}{b^2} f_5 & -\frac{1}{b} f_2 \\ -\frac{1}{b} f_2 & \frac{1}{a^2} f_1 \end{pmatrix}\begin{pmatrix} \frac{1}{a} f_3 & \frac{1}{a} f_4 \\ \frac{a}{b} f_6 & \frac{a}{b} f_7 \end{pmatrix}$$

$$= \frac{b^2}{f_1 f_5 - f_2^2}\begin{pmatrix} \frac{a}{b^2}\left(f_3 f_5 - f_2 f_6\right) & \frac{a}{b^2}\left(f_5 f_4 - f_2 f_7\right) \\ \frac{1}{ab}\left(f_1 f_6 - f_2 f_3\right) & \frac{1}{ab}\left(f_1 f_7 - f_2 f_4\right) \end{pmatrix}$$

$$\mathfrak{I}_{\omega\eta}\mathfrak{I}_{\eta\eta}^{-1}\mathfrak{I}_{\eta\omega} = \frac{b^2}{f_1 f_5 - f_2^2}\begin{pmatrix} \frac{1}{a} f_3 & \frac{a}{b} f_6 \\ \frac{1}{a} f_4 & \frac{a}{b} f_7 \end{pmatrix}\begin{pmatrix} \frac{a}{b^2}\left(f_3 f_5 - f_2 f_6\right) & \frac{a}{b^2}\left(f_5 f_4 - f_2 f_7\right) \\ \frac{1}{ab}\left(f_1 f_6 - f_2 f_3\right) & \frac{1}{ab}\left(f_1 f_7 - f_2 f_4\right) \end{pmatrix}$$

$$= \frac{1}{f_1 f_5 - f_2^2}\begin{pmatrix} f_3\left(f_3 f_5 - f_2 f_6\right) + f_6\left(f_1 f_6 - f_2 f_3\right) & f_3\left(f_5 f_4 - f_2 f_7\right) + f_6\left(f_1 f_7 - f_2 f_4\right) \\ f_4\left(f_3 f_5 - f_2 f_6\right) + f_7\left(f_1 f_6 - f_2 f_3\right) & f_4\left(f_5 f_4 - f_2 f_7\right) + f_7\left(f_1 f_7 - f_2 f_4\right) \end{pmatrix}$$

and

$$\operatorname{cov}\left(\sqrt{N}(\hat{p},\hat{q})\right) = \left(\mathfrak{I}_{\omega\omega} - \mathfrak{I}_{\omega\eta}\mathfrak{I}_{\eta\eta}^{-1}\mathfrak{I}_{\eta\omega}\right)^{-1}$$

$$= \left[\begin{pmatrix} f_8 & f_9 \\ f_9 & f_{10} \end{pmatrix} - \frac{1}{f_1 f_5 - f_2^2}\begin{pmatrix} f_3^2 f_5 + f_1 f_6^2 - 2 f_2 f_3 f_6 & f_3 f_4 f_5 + f_1 f_6 f_7 - f_2 f_3 f_7 - f_2 f_4 f_6 \\ f_3 f_4 f_5 + f_1 f_6 f_7 - f_2 f_3 f_7 - f_2 f_4 f_6 & f_4^2 f_5 + f_1 f_7^2 - 2 f_2 f_4 f_7 \end{pmatrix}\right]^{-1} \tag{4.1}$$

which does not depend on *a* or *b*.

<u>Theorem 2</u>[2]

If $X \sim GB2(a,b,p,q)$ then $\frac{1}{X} \sim GB2\left(a,\frac{1}{b},q,p\right)$ and vice versa.

Proof

$$f(x) = \frac{a\left(x/b\right)^{ap}}{xB(p,q)\left[1+\left(x/b\right)^{a}\right]^{p+q}}$$

Let $y = \frac{1}{x}$, then using a change in variables

[2] We are indebted to a referee who pointed out that Kleiber (1996) has shown that $X \sim SM(a,b,q) \Leftrightarrow X^{-1} \sim Dagum\left(a,b^{-1},q\right)$.

$$f(y) = f(x)\left|\frac{dx}{dy}\right| = \frac{a\left(1/by\right)^{ap}}{\frac{1}{y}B(p,q)\left[1+\left(1/by\right)^{a}\right]^{p+q}}\frac{1}{y^2}$$

$$= \frac{a\left(b^*/y\right)^{ap}\left(b^*/y\right)^{-ap-aq}}{yB(p,q)\left(b^*/y\right)^{-a(p+q)}\left[1+\left(b^*/y\right)^{a}\right]^{p+q}}$$

where $b^* = 1/b$. Then,

$$f(y) = \frac{a\left(y/b^*\right)^{aq}}{yB(p,q)\left\{\left(b^*/y\right)^{-a}\left[1+\left(b^*/y\right)^{a}\right]\right\}^{p+q}}$$

$$= \frac{a\left(y/b^*\right)^{aq}}{yB(p,q)\left[1+\left(y/b^*\right)^{a}\right]^{p+q}}$$

$$= GB2(a,b^*,q,p)$$

Corollary

Because the asymptotic covariance matrix for the maximum likelihood estimators p and q does not depend on a and b, the symmetry between $GB2(a,b,p,q)$ and $GB2(a,b^*,q,p)$ implies

$$\text{var}\left(\sqrt{N}\hat{p} \mid p = c_1, q = c_2\right) = \text{var}\left(\sqrt{N}\hat{q} \mid p = c_2, q = c_1\right)$$

It follows that the properties of Wald tests for p when $p = c_1$ and $q = c_2$ will be the same as those for tests for q when $p = c_2$ and $q = c_1$. To confirm this result, in the Tables S6-S9 we report values for $\lambda_{SM} = N\left(p-1\right)^2/\Im^{pp}$, $\lambda_{SM}^{\log} = N\left[p\ln(p)\right]^2/\Im^{pp}$, $\lambda_D = N\left(q-1\right)^2/\Im^{qq}$ and $\lambda_D^{\log} = N\left[q\ln(q)\right]^2/\Im^{qq}$. These quantities are the noncentrality parameters of the noncentral chi-square distributions that would be valid under asymptotic local approximations. They are the Wald test statistics with estimates replaced by their corresponding true parameters. The Wald test symmetry between p and q is clear from the tables.

Table S6 Values of $\lambda_{SM} = N(p-1)^2 / \Im^{pp}$ for $H_0: p=1$ for selected values of *p* and *q*

		Value of *p*					
		0.5	1	1.5	2	2.5	3
Value of *q*	0.5	46.409	0.000	2.896	4.761	5.174	4.969
	1	55.309	0.000	3.343	5.602	6.220	6.098
	1.5	56.895	0.000	3.311	5.560	6.212	6.139
	2	56.721	0.000	3.190	5.336	5.960	5.903
	2.5	56.113	0.000	3.065	5.097	5.678	5.618
	3	55.442	0.000	2.955	4.884	5.420	5.351

Table S7 Values of $\lambda_{D} = N(q-1)^2 / \Im^{qq}$ for $H_0: q=1$ for selected values of *p* and *q*

		Value of *p*					
		0.5	1	1.5	2	2.5	3
Value of *q*	0.5	46.409	55.309	56.895	56.721	56.113	55.442
	1	0.000	0.000	0.000	0.000	0.000	0.000
	1.5	2.896	3.343	3.311	3.190	3.065	2.955
	2	4.761	5.602	5.560	5.336	5.097	4.884
	2.5	5.174	6.220	6.212	5.960	5.678	5.420
	3	4.969	6.098	6.139	5.903	5.618	5.351

Table S8 Values of $\lambda_{SM}^{\log} = N[p\ln(p)]^2 / \Im^{pp}$ for $H_0: \ln(p)=0$ for selected values of *p* and *q*

		Value of *p*					
		0.5	1	1.5	2	2.5	3
Value of *q*	0.5	22.297	0.000	4.285	9.150	12.068	13.494
	1	26.573	0.000	4.946	10.766	14.506	16.561
	1.5	27.336	0.000	4.899	10.685	14.487	16.672
	2	27.252	0.000	4.720	10.254	13.901	16.029
	2.5	26.960	0.000	4.535	9.796	13.243	15.257
	3	26.637	0.000	4.373	9.386	12.640	14.531

Table S9 Values of $\lambda_{D}^{\log} = N[q\ln(q)]^2 / \Im^{qq}$ for $H_0: \ln(q)=0$ for selected values of *p* and *q*

		Value of *p*					
		0.5	1	1.5	2	2.5	3
Value of *q*	0.5	22.297	26.573	27.336	27.252	26.960	26.637
	1	0.000	0.000	0.000	0.000	0.000	0.000
	1.5	4.285	4.946	4.899	4.720	4.535	4.373
	2	9.150	10.766	10.685	10.254	9.796	9.386
	2.5	12.068	14.506	14.487	13.901	13.243	12.640
	3	13.494	16.561	16.672	16.029	15.257	14.531

S5 GB2 density functions

The GB2 distribution is quite flexible and can take a wide range of shapes. A thorough look at GB2 and some related distributions is provided by Kleiber and Kotz (2003, 183-234). The following representations are inspired by their illustrations. The four parameters *a*, *b*, *p* and *q* are positive. The parameter *b* adjusts the scale of the distribution. The other parameters affect the shape of the distribution. If $f(x)$ is the GB2 distribution, the mode occurs at $x = b\left(\frac{ap-1}{aq+1}\right)^{1/a}$, if $ap > 1$.[3] The *k*th moment of the distribution, for $-ap < k < aq$, are

$$E\left(X^k\right) = \frac{b^k B\left(p + k/a, q - k/a\right)}{B\left(p, q\right)}$$

where $B\left(p, q\right)$ is the beta function.[4] Various relationships between the special cases of the GB2 distribution are shown in Figure 4 of Kleiber and Kotz (2003, p. 188). When using the GB2 distribution to study income distributions, it is relatively common to obtain parameter estimates where higher-order moments of the distribution do not exist. Some of our Monte Carlo parameter settings reflect this fact. Examples are plotted in Section 5.2. In the next section, S5.1, we also plot examples which have finite fourth moments.

S5.1 GB2 distributions with fourth moments

In this first section we illustrate GB2 distributions with finite fourth moments. In the next section, S5.2, we consider the GB2 distributions used in our Monte Carlo simulations.

S5.1.1 The Dagum distribution

The Dagum distribution results when $q = 1$. In Figure S9 we plot this distribution for different values of *p*, with $a = 5$ and $b = 1$. The moments all increase as *p* increases, as we show in Table S10.

[3] Kleiber and Kotz (2003, p.188)
[4] Ibid.

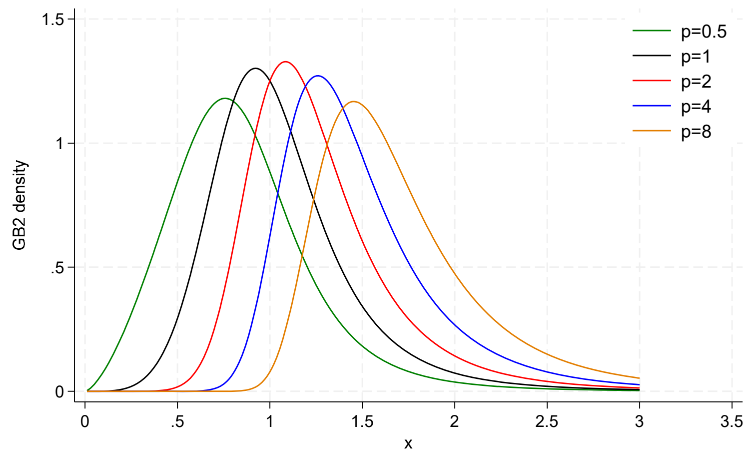


Figure S9 The Dagum distribution, q = 1, for values of p given a = 5 and b = 1.

Table S10 Moments of the Dagum distribution as p increases

p	mean	mode	variance	skewness	kurtosis
0.5	0.853	0.758	0.171	16.854	79.606
1	1.069	0.922	0.179	26.252	133.999
2	1.283	1.084	0.204	34.321	184.257
4	1.505	1.259	0.250	39.467	217.579
8	1.747	1.454	0.319	42.318	236.389

S5.1.2 The beta distribution of the second kind

If a = 1 we obtain a beta distribution of the second kind. First, we examine cases in which q increases. In Figure S10 we plot this distribution for different values of q, with p = 5 and b = 1. In Table S11 we show that as q increases the mean and variance of the distribution becomes smaller, and the skewness greater. The values of q are chosen such that the moments exist.

Figure S10 The beta-2 distribution, a = 1, with b = 1 and p = 5.

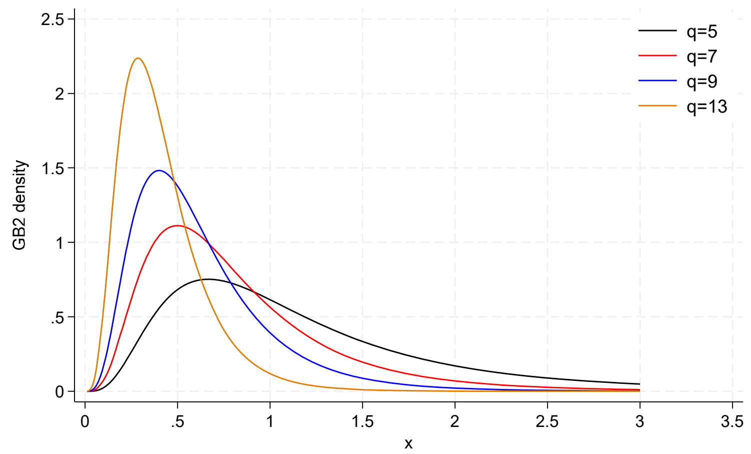


Table S11 Moments of the beta-2 distribution as q increases

q	mean	mode	variance	skewness	kurtosis
5	1.250	0.667	0.937	9.639	79.644
7	0.833	0.500	0.306	10.361	49.983
9	0.625	0.400	0.145	11.309	47.504
13	0.417	0.286	0.054	12.798	49.109

Second, we examine cases in which p increases. In Figure S11 we plot this distribution for different values of p, with q = 5 and b = 1. In Table S12 we show that as p increases the mean and variance of the distribution becomes larger, and the skewness greater.

Figure S11 The beta-2 distribution, a = 1, with b = 1 and q = 5.

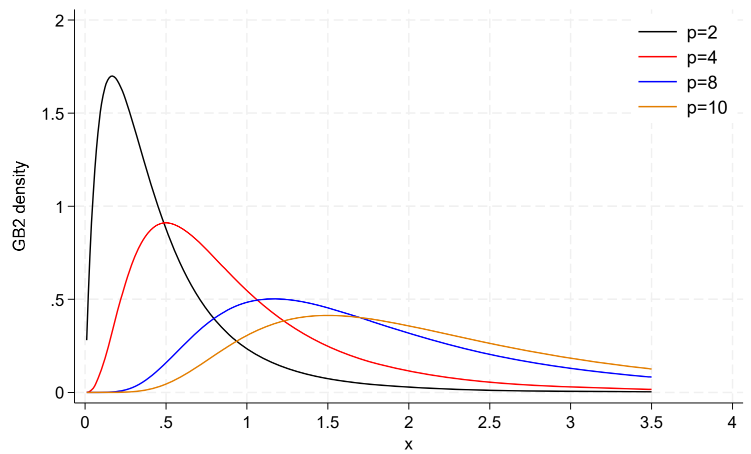


Table S12 Moments of the beta-2 distribution as p increases

p	mean	mode	variance	skewness	kurtosis
2	0.500	0.167	0.250	8.000	80.000
4	1.000	0.500	0.667	9.186	78.750
8	2.000	1.167	2.000	10.607	82.500
10	2.500	1.500	2.917	11.042	84.049

S5.1.3 The Singh-Maddala distribution

We obtain the Singh-Maddala distribution when $p = 1$. We plot the distribution over several values of q in Figure S12, using $a = 5$, and $b = 1$ so that fourth moments exist. We observe that as q increases the probability mass shifts left. In Table S13 we see that as q increases the measures of central tendency fall, but skewness and kurtosis increase.

Figure S12 The Singh-Maddala distribution, $p = 1$, with $b = 1$ and $a = 5$.

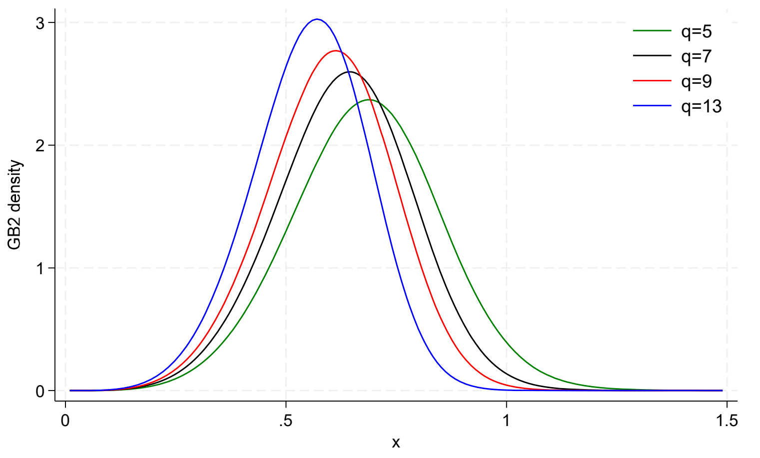


Table S13 Moments of the Singh-Maddala distribution as q increases

q	mean	mode	variance	skewness	kurtosis
5	0.682	0.688	0.029	76.432	358.051
7	0.633	0.644	0.024	81.787	389.332
9	0.600	0.614	0.021	84.851	407.560
13	0.555	0.571	0.017	88.221	427.866

S5.1.4 The Fisk distribution

When $p = 1$ and $q = 1$ we obtain the Fisk distribution. In Figure S13 we plot the distribution over several values of a, with $b = 1$. As a increases the probability mass location changes little, as shown in Table S14, but the variance decreases, and skewness and kurtosis increase.

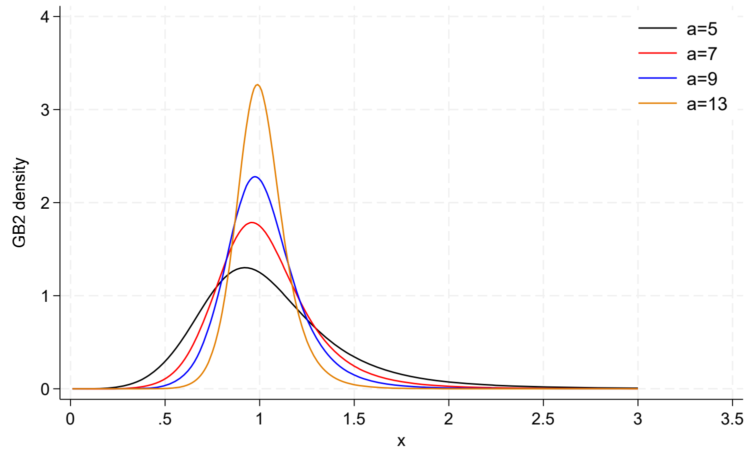
Figure S13 The Fisk distribution, p = q = 1 over a, with b = 1.	

Table S14 Moments of the Fisk distribution as a increases

a	mean	mode	variance	skewness	kurtosis
5	1.069	0.922	0.179	26.252	133.999
7	1.034	0.960	0.078	63.229	301.596
9	1.021	0.976	0.044	128.920	716.770
13	1.010	0.988	0.020	377.301	2843.238

S5.2 GB2 distributions used in the Monte Carlo experiments

In the Monte Carlo experiments we use the choices of the underlying parameter values based on values reported in the literature. In this section we examine the shapes of the resulting GB2 densities and their moments.

S5.2.1 GB2 distributions used to test for the Dagum distribution

To test for the Dagum distribution, $H_0 : q = 1, H_1 : q \neq 1$, we consider 36 values of q in the interval [0.5, 4] in increments of 0.1, setting a = 5, b = 1 and p = 1.75. In Figure S14 we illustrate the shapes of the distributions with q = 0.5, 1, 2.5 and 4. As q increases the center of the distribution shifts leftward, as shown in Figure S14 and Table S15, and the variance decreases. When q = 0.5 the third and fourth moments do not exist. However, for q > 0.5, increasing q increases skewness and kurtosis.

Figure S14 GB2 distributions used to test for the Dagum distribution, $q = 1$, for values of q given $p = 1.75$, $a = 5$ and $b = 1$.

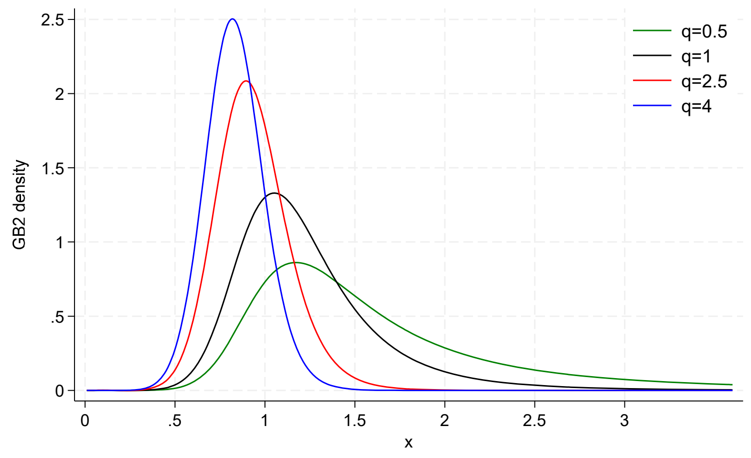


Table S15 Moments of the Dagum distributions in Figure S14

q	mean	mode	variance	skewness	kurtosis
0.5	1.800	1.172	3.028	NA	NA
1	1.241	1.053	0.198	32.990	175.787
2.5	0.936	0.895	0.043	104.517	542.412
4	0.834	0.819	0.027	142.738	801.700

S5.2.2 GB2 distributions used to test for the beta distribution of the second kind

To test for the beta distribution of the second kind, $H_0 : a = 1, H_1 : a \neq 1$, we examine 36 values of a in the interval [0.5, 4] in increments of 0.1, setting $b = 1$ and $p = q = 1.75$. In Figure S15 we illustrate the shapes of the distributions with $a = 0.5$, 1, 2.5 and 4. As a increases the probability mass of the distribution shifts rightward. The moments in Table S16 show that for $a = 0.5$ no finite moments exist, and when $a = 1$ the second and higher moments do not exist.

Figure S15 GB2 distributions used to test for the beta distribution of the second kind, a = 1, for values of a given p = q = 1.75 and b = 1.

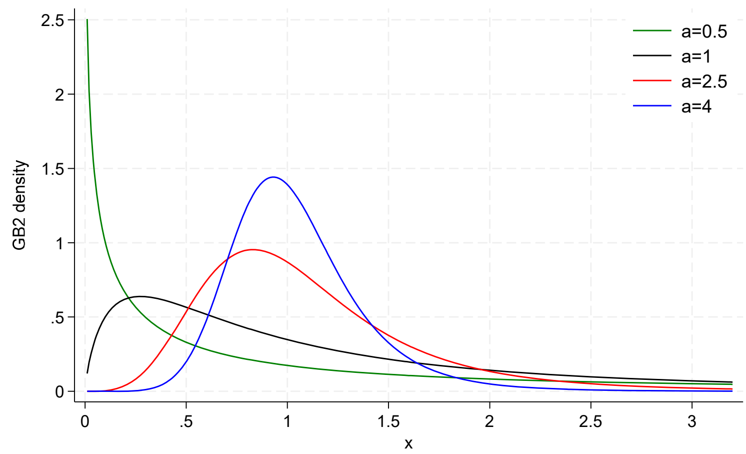


Table S16 Moments of the beta-2 distributions in Figure S15

a	mean	mode	variance	skewness	kurtosis
0.5	NA	NA	NA	NA	NA
1	0.593	0.273	NA	NA	NA
2.5	0.288	0.830	0.345	4.589	44.504
4	0.267	0.931	0.238	3.447	10.478

S5.2.3 GB2 distributions used to test for the Singh-Maddala distribution

To test for the Singh-Maddala distribution, $H_0 : p = 1, H_1 : p \neq 1$, we consider 36 values of p in the interval [0.5, 4] in increments of 0.1, setting a =5, b = 1 and q = 1.75. In Figure S16 we illustrate the shapes of the distributions with p = 0.5, 1, 2.5 and 4. As p increases the center of the distribution shifts rightward, as shown in Figure S16 and Table S17. The variance decreases slightly and skewness and kurtosis increase.

Figure S16 GB2 distributions used to test for the Singh-Maddala distribution, p = 1, for values of p given q = 1.75, a = 5 and b = 1.

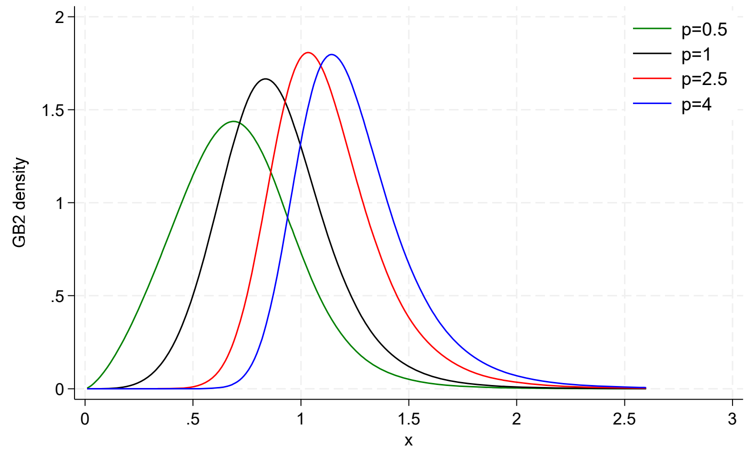


Table S17 Moments of the Singh-Maddala distributions in Figure S16

p	mean	mode	variance	skewness	kurtosis
0.5	0.708	0.688	0.083	22.819	82.636
1	0.888	0.837	0.072	47.181	202.945
2.5	1.124	1.034	0.070	90.781	463.433
4	1.250	1.143	0.075	110.736	596.779

S5.2.4 GB2 distributions used to test for the Fisk distribution

To test for the Fisk distribution, $H_0 : p = q = 1$, we consider 36 values of $p = q$ in the interval [0.5, 4] in increments of 0.1, setting a =5, b =1. In Figure S17 we illustrate the shapes of the distributions with p = 0.5, 1, 2.5 and 4. As p and q increase the mean of the distribution shifts leftward and the mode rightward, as shown in Figure S17 and Table S18. When $p = q = 0.5$ the second and higher moments do not exist. The variance decreases slightly and skewness and kurtosis increase. For larger values the variance gets smaller, and skewness and kurtosis fall and then increase.

Figure S17 GB2 distributions used to test for the Fisk distribution, $p = q = 1$, for values of p and q given $a = 5$ and $b = 1$.

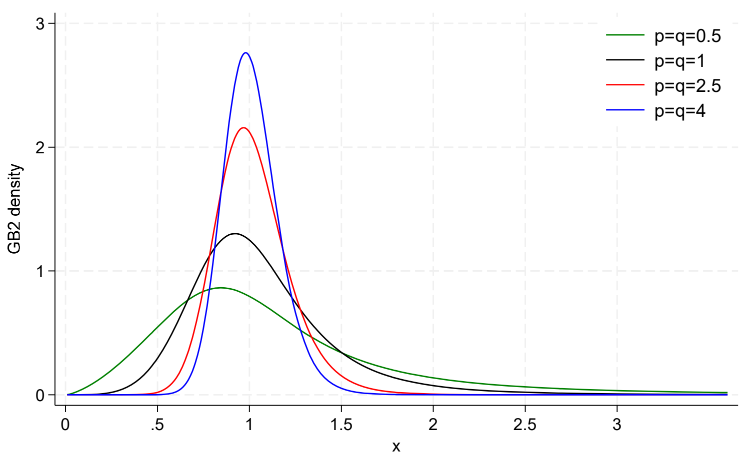


Table S18 Moments of the Fisk distributions in Figure S17

p=q	mean	mode	variance	skewness	kurtosis
0.5	3.883	0.844	NA	NA	NA
1	1.069	0.922	0.179	26.252	133.999
2.5	0.075	0.968	0.074	4.371	18.532
4	0.007	0.980	0.007	12.376	155.687

S6 Further analysis of power nonmonotonicity of Wald tests.

The powers of the Wald tests for $H_0: p=1$ and $H_0: \ln(p)=0$ do not increase monotonically for $p>1$. This property can be explained by the behavior of the "noncentrality parameters" $\lambda_{SM}=(p-1)^2/\operatorname{var}(\hat{p})$ and $\lambda_{SM}^{\log}=[p\ln(p)]^2/\operatorname{var}(\hat{p})$. Both the numerators and denominator of λ_{SM} and $\lambda_{SM}^{\log}$ increase as p increases. The denominator increases at an increasing rate, as illustrated in Figure S18 where we plot the numerical derivative of the asymptotic variance, $d\left[\operatorname{var}(\hat{p})\right]/dp = N^{-1}\Im^{pp}$.

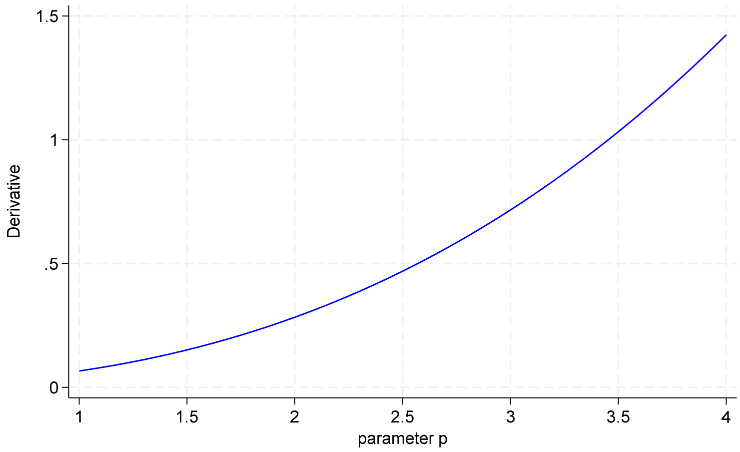


Figure S18 The derivative of the asymptotic variance of the MLE for *p*. The derivative $d\left[\operatorname{var}(\hat{p})\right]/dp$ is computed using numerical derivatives for $1\le p\le 4$, with $N=4000$ and $q=1.75$.

The parameter λ_{SM} becomes nonmonotonic because its denominator increases at an increasing rate whereas its numerator increases linearly, $d\left[(p-1)^2\right]/dp=2(p-1)$. To find the point at which λ_{SM} starts to decline, consider

$$\frac{d(\lambda_{SM})}{dp}=\left[\frac{\operatorname{var}(\hat{p})d\left[(p-1)^2\right]/dp-(p-1)^2 d\left[\operatorname{var}(\hat{p})\right]/dp}{\operatorname{var}^2(\hat{p})}\right]=0$$

The maximum value for λ_{SM} is reached when

$$\frac{d\left[(p-1)^2\right]/dp}{(p-1)^2}=\frac{d\left[\operatorname{var}(\hat{p})\right]/dp}{\operatorname{var}(\hat{p})}$$

That is, the proportional rate of change of the numerator equals the proportional rate of change of the denominator. The asymptotic local approximation suggests power is increasing when the rate of change in the numerator is greater than the rate of change in the denominator,

$$\frac{d\left[(p-1)^2\right]/dp}{(p-1)^2} > \frac{d\left[\operatorname{var}(\hat{p})\right]/dp}{\operatorname{var}(\hat{p})}$$

and decreasing when the rate of change in the numerator is less than the rate of change in the denominator,

$$\frac{d\left[(p-1)^2\right]/dp}{(p-1)^2} < \frac{d\left[\operatorname{var}(\hat{p})\right]/dp}{\operatorname{var}(\hat{p})}$$

In Figure S19(a), we plot both $\left\{d\left[(p-1)^2\right]/dp\right\}/(p-1)^2$ and $\left\{d\left[\operatorname{var}(\hat{p})\right]/dp\right\}/\operatorname{var}(\hat{p})$. The point at which they are equal and hence the point at which the hypothetical power starts to decline is 2.663. In Figure 1 we observe that, although larger than the hypothetical power, the Monte Carlo estimated power starts to decline at approximately the same point.

We can repeat the same exercise for the Wald test in the log metric, that is, $H_0: \ln(p) = 0$, with $\lambda_{SM}^{\log} = [p\ln(p)]^2/\operatorname{var}(\hat{p})$. Comparing $\left\{d\left[[p\ln(p)]^2\right]/dp\right\}/[p\ln(p)]^2$ with $\left\{d\left[\operatorname{var}(\hat{p})\right]/dp\right\}/\operatorname{var}(\hat{p})$ we find that hypothetical power is increasing when

$$\frac{d\left[[p\ln(p)]^2\right]/dp}{[p\ln(p)]^2} > \frac{d\left[\operatorname{var}(\hat{p})\right]/dp}{\operatorname{var}(\hat{p})}$$

and decreasing when

$$\frac{d\left[[p\ln(p)]^2\right]/dp}{[p\ln(p)]^2} < \frac{d\left[\operatorname{var}(\hat{p})\right]/dp}{\operatorname{var}(\hat{p})}$$

The two rates of change are equal when $p = 4.356$ [5]. This result is illustrated in Figure S19(b). The larger value of p at which the two curves intersect and a small difference between the two curves after this point helps explain the power results in Figure 1 in the main paper.

---

[5] The crossing values were found by minimizing the absolute difference of the change in the numerator and denominator.

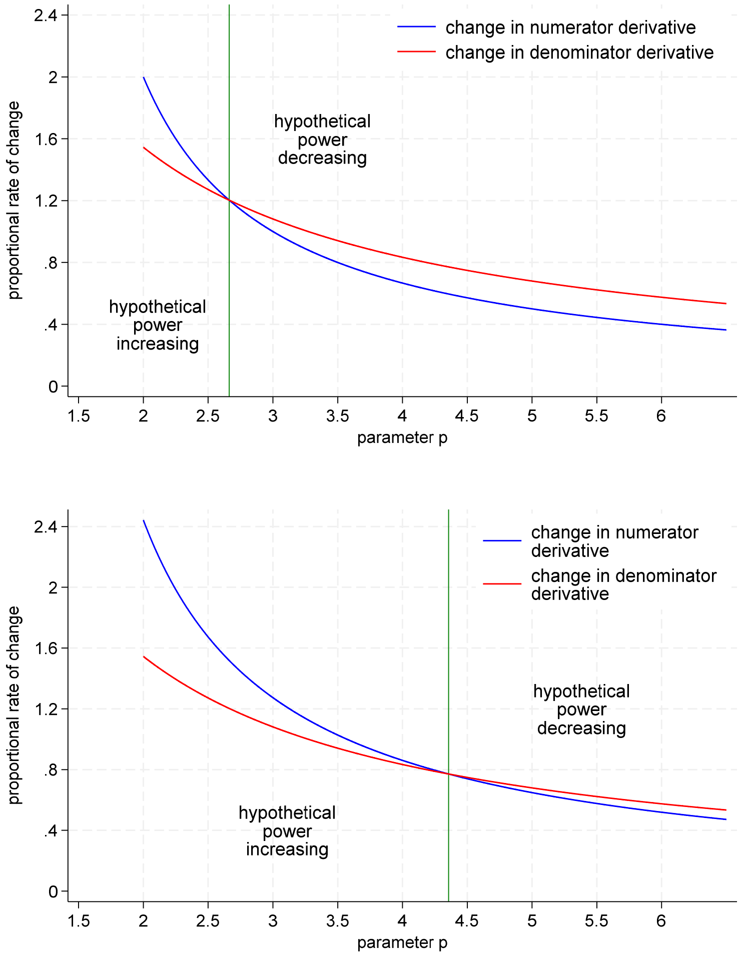


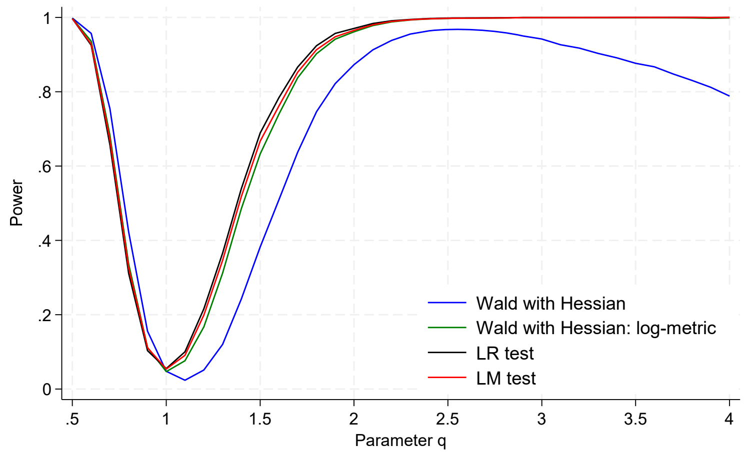
Figure S19 Wald test hypothetical power deconstruction. In the upper panel (a), the curves denoted by change in the numerator and change in the denominator are plots of $\left\{d\left[(p-1)^2\right]/dp\right\}/(p-1)^2$ and $\left\{d\left[\mathrm{var}(\hat{p})\right]/dp\right\}/\mathrm{var}(\hat{p})$, respectively, over the range $2 \le p \le 6$. They relate to testing $H_0: p=1$. In the lower panel (b), the change in the numerator is $\left\{d\left[[p\ln(p)]^2\right]/dp\right\}/[p\ln(p)]^2$; the change in the denominator is the same as that in the upper panel, namely $\left[d\left[\mathrm{var}(\hat{p})\right]/dp\right]/\mathrm{var}(\hat{p})$. These curves relate to testing $H_0: \ln(p)=0$. For $H_0: p=1$, the parameter λ_{SM} increases as p increases up to the point 2.663, and then declines. For $H_0: \ln(p)=0$, the decline in $\lambda_{SM}^{\log}$ occurs at the larger value of 4.356. Other settings were $q=1.75$ and $N=4000$.

S7 Monte Carlo experiments for testing *q*

If $q = 1$ we obtain the Dagum distribution. To estimate the finite sample power profiles of the four tests for testing $q=1$, and $\ln(q)=0$, we carried out a Monte Carlo experiment with 10,000 samples, a sample size of $N=4000$, the settings $a=5, b=1$, $p=1.75$, and values of q from 0.5 to 4.0 in increments of 0.1. We use Wald chi-square tests using the test statistics

$W = N\left(\mathbf{R}\hat{\boldsymbol{\theta}} - \mathbf{r}\right)'\left(\mathbf{R}\hat{\mathfrak{I}}^{-1}\mathbf{R}'\right)^{-1}\left(\mathbf{R}\hat{\boldsymbol{\theta}} - \mathbf{r}\right)$ and $W = N\left(\mathbf{R}\hat{\boldsymbol{\phi}}\right)'\left(\mathbf{R}\hat{\mathbf{D}}\hat{\mathfrak{I}}^{-1}\hat{\mathbf{D}}\mathbf{R}'\right)^{-1}\mathbf{R}\hat{\boldsymbol{\phi}}$ for the null hypotheses $H_0 : q = 1$ and $H_0 : \ln(q) = 0$, respectively. The Wald test statistics using estimates of the information matrix, the negative Hessian of the log-likelihood function, and the outer-product of gradients (OPG) of the log-likelihood function yield almost identical Monte Carlo power over the range of values we consider. The likelihood ratio test, LR, and the Lagrange multiplier test, LM, and the Lagrange multiplier test using the OPG have almost identical power. We summarize these results in Figure S20. The Wald test of $H_0 : q = 1$ using the Hessian has slightly higher power when $q < 1$, and obvious lower power when $q > 1$. Given the performance of the other tests in these Monte Carlo results we do not recommend using the Wald test of $H_0 : q = 1$.

Figure S20 Monte Carlo power testing $q = 1$, and the log-metric version $\ln(q) = 0$. We use 10,000 Monte Carlo samples with sample size N = 4000. The other parameter values are a = 5, b =1 and p = 1.75.



We report the numerical values in Table S20, preceded by an abbreviation legend in Table S19

Table S19 Testing q Monte Carlo legend

q	parameter of interest
winfo	Wald stat using information matrix evaluated at MLE
wopg	Wald stat using OPG with analytic gradient
whess	Wald stat using analytic Hessian
winfoln	Wald stat using cov_info and delta method
wopgln	Wald stat using cov_opg and delta method
whessln	Wald stat using cov_hess and delta method
wopglmet	Wald stat using numerical OPG cov mat
wopglmet	Wald stat using numerical OPG cov mat in log-metric
lrtest	likelhood ratio test
lmtest	LM test using analytic gradient
lmopg	LM test OPG version using analytic gradient

Table S20 Testing q: Wald test Monte Carlo power

q	winfo	wopg	whess	winfoln	wopgln	whessln	wopglmet	lrtest	lmtest	lmopg
0.5	0.9986	0.9985	0.9986	0.9983	0.9979	0.9978	0.9979	0.9971	0.9967	0.9951
0.6	0.9588	0.9563	0.9571	0.9383	0.9340	0.9362	0.9340	0.9251	0.9284	0.9103
0.7	0.7562	0.7525	0.7547	0.6874	0.6796	0.6825	0.6796	0.6570	0.6701	0.6205
0.8	0.4245	0.4200	0.4229	0.3388	0.3326	0.3363	0.3326	0.3115	0.3253	0.2819
0.9	0.1564	0.1550	0.1557	0.1135	0.1120	0.1118	0.1120	0.1033	0.1110	0.0924
1	0.0478	0.0472	0.0476	0.0481	0.0450	0.0466	0.0450	0.0548	0.0535	0.0592
1.1	0.0263	0.0222	0.0237	0.0785	0.0745	0.0760	0.0745	0.1000	0.0908	0.1160
1.2	0.0518	0.0467	0.0512	0.1706	0.1631	0.1670	0.1631	0.2157	0.1990	0.2422
1.3	0.1234	0.1130	0.1199	0.3146	0.3056	0.3120	0.3056	0.3656	0.3464	0.4012
1.4	0.2467	0.2254	0.2422	0.4876	0.4781	0.4850	0.4781	0.5391	0.5180	0.5738
1.5	0.3859	0.3593	0.3824	0.6390	0.6282	0.6328	0.6282	0.6890	0.6672	0.7181
1.6	0.5162	0.4848	0.5093	0.7454	0.7351	0.7400	0.7351	0.7842	0.7618	0.8097
1.7	0.6480	0.6146	0.6370	0.8393	0.8340	0.8377	0.8340	0.8664	0.8525	0.8851
1.8	0.7590	0.7279	0.7462	0.9060	0.8998	0.9025	0.8998	0.9242	0.9132	0.9368
1.9	0.8307	0.8041	0.8219	0.9441	0.9397	0.9417	0.9397	0.9570	0.9481	0.9638
2	0.8802	0.8561	0.8727	0.9626	0.9602	0.9615	0.9602	0.9703	0.9653	0.9753
2.1	0.9221	0.8988	0.9130	0.9809	0.9767	0.9784	0.9767	0.9840	0.9804	0.9886
2.2	0.9466	0.9267	0.9384	0.9885	0.9873	0.9883	0.9873	0.9913	0.9893	0.9929
2.3	0.9623	0.9426	0.9555	0.9935	0.9920	0.9932	0.9920	0.9947	0.9938	0.9958
2.4	0.9698	0.9531	0.9641	0.9965	0.9962	0.9963	0.9962	0.9975	0.9964	0.9980
2.5	0.9734	0.9566	0.9678	0.9978	0.9973	0.9977	0.9973	0.9981	0.9979	0.9984
2.6	0.9726	0.9522	0.9676	0.9982	0.9976	0.9978	0.9976	0.9987	0.9981	0.9989
2.7	0.9713	0.9469	0.9643	0.9986	0.9979	0.9980	0.9979	0.9992	0.9981	0.9997
2.8	0.9664	0.9394	0.9599	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995
2.9	0.9577	0.9296	0.9501	0.9997	0.9996	0.9997	0.9996	0.9997	0.9997	0.9998
3	0.9537	0.9182	0.9422	0.9998	0.9996	0.9998	0.9996	0.9999	0.9999	0.9999
3.1	0.9371	0.9025	0.9266	0.9997	0.9997	0.9997	0.9997	0.9999	0.9999	1.0000
3.2	0.9266	0.8891	0.9175	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.3	0.9162	0.8730	0.9034	0.9992	0.9992	0.9992	0.9992	0.9997	0.9997	0.9998
3.4	0.9022	0.8580	0.8912	0.9995	0.9993	0.9994	0.9993	1.0000	1.0000	1.0000
3.5	0.8890	0.8400	0.8766	0.9997	0.9996	0.9997	0.9996	1.0000	1.0000	1.0000
3.6	0.8782	0.8291	0.8674	0.9997	0.9995	0.9996	0.9995	1.0000	1.0000	1.0000
3.7	0.8635	0.8087	0.8478	0.9990	0.9988	0.9989	0.9988	1.0000	1.0000	1.0000
3.8	0.8428	0.7910	0.8307	0.9988	0.9984	0.9987	0.9984	1.0000	1.0000	1.0000
3.9	0.8226	0.7710	0.8121	0.9981	0.9980	0.9981	0.9980	0.9999	0.9999	0.9999
4	0.8034	0.7445	0.7886	0.9990	0.9988	0.9990	0.9988	1.0000	1.0000	1.0000

S8 Monte Carlo experiments for testing *a*

The Wald test statistic for testing for the B2 distribution using the null hypothesis $H_0 : a = 1$ is given by $W_{B2} = N(\hat{a}-1)^2 / \hat{\Im}^{aa}$ where $\hat{\Im}^{aa}$ is an estimate of $\Im^{aa}$. The alternative Wald statistic using $H_0 : \ln(a) = 0$ is given by

$$W_{B2}^{\log} = \frac{N\hat{\alpha}^2}{\hat{\Im}^{aa} / \hat{a}^2} = \frac{N[\hat{a}\ln(\hat{a})]^2}{\hat{\Im}^{aa}}$$

The LM and LR test statistics for testing $H_0 : a = 1$ use the restricted estimator $\tilde{\boldsymbol{\theta}}' = (1\, \tilde{b}\, \tilde{p}\, \tilde{q})$. Also, $\tilde{\mathbf{g}}' = (\tilde{g}_1\ 0\ 0\ 0)$ and the LM statistic becomes

$$\mathrm{LM} = N\tilde{g}_1\tilde{\Im}^{aa}$$

From results in Section S1

$$\tilde{g}_1 = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial \ln L_i}{\partial a}\bigg|_{\tilde{\theta}} = 1 + \tilde{p}\frac{1}{N}\sum_{i=1}^{N}\ln\left(\frac{x_i}{\tilde{b}}\right) - (\tilde{p}+\tilde{q})\frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i/\tilde{b}}{1+x_i/\tilde{b}}\right)\ln\left(\frac{x_i}{\tilde{b}}\right)$$

8.1 Power results for testing for a B2 distribution $(a = 1)$

Using a sample size of 4000, and 10,000 samples, we estimated the power functions for testing for a B2 distribution. Parameter settings $b = 1$, $p = 1.75$ and $q = 1.75$ were used and a was varied from 0.5 to 4 in increments of 0.1. Examples of the density functions implied by these settings are illustrated in Section S5 for selected values of a. The power functions for the LR, LM and two Wald tests are plotted in Figure S21. In contrast to the tests for p and q, there is very little difference between power functions, all tests perform almost equally well, and the power of the Wald tests increases monotonically.

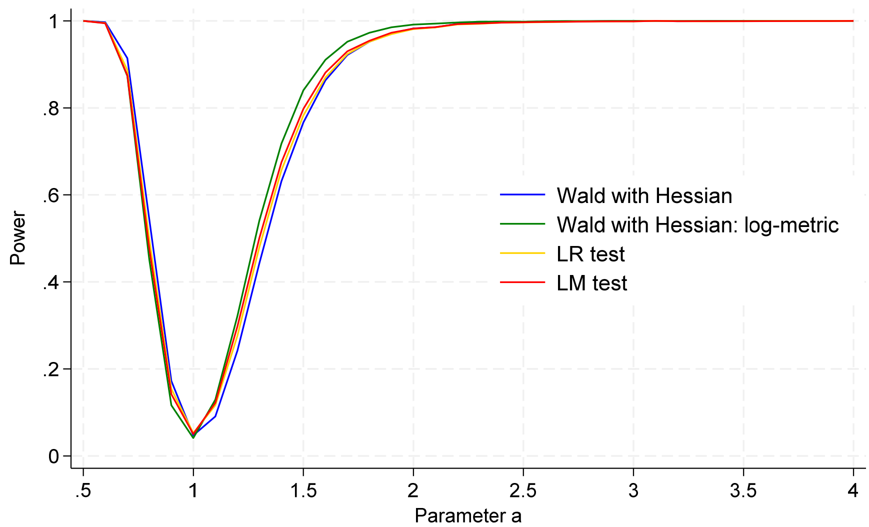


Figure S21 Monte-Carlo estimated power for the LM, LR and Wald tests for testing *a*. Power testing $H_0 : a = 1$ and $H_0 : \ln(a) = 0$, with $N = 4000, b = 1$, $p = 1.75$ and $q = 1.75$, and using 10,000 samples.

S9 Testing for the Fisk Distribution

The Wald and LM tests for the joint null hypotheses $H_0: p = q = 1$ and $H_0: \ln(p) = \ln(q) = 0$, relevant when testing whether the GB2 distribution reduces to the Fisk distribution, reveal several interesting features.

S9.1 Wald tests

Using the notation and partitioning introduced in the proof to Theorem 1, namely, $\eta = (a\, b)'$ and $\omega = (p\, q)'$ and

$$\mathfrak{I} = \begin{pmatrix} \mathfrak{I}_{\eta\eta} & \mathfrak{I}_{\eta\omega} \\ \mathfrak{I}_{\omega\eta} & \mathfrak{I}_{\omega\omega} \end{pmatrix}$$

the Wald test statistic for $H_0: p = 1, q = 1$ can be written as

$$W_F = N\left(\hat{p}-1 \quad \hat{q}-1\right)\left(\hat{\mathfrak{I}}_{\omega\omega} - \hat{\mathfrak{I}}_{\omega\eta}\hat{\mathfrak{I}}_{\eta\eta}^{-1}\hat{\mathfrak{I}}_{\eta\omega}\right)\begin{pmatrix} \hat{p}-1 \\ \hat{q}-1 \end{pmatrix}$$

Replacing the estimates with the true parameters yields

$$\lambda\left(W_F\right) = N\left(p-1 \quad q-1\right)\left(\mathfrak{I}_{\omega\omega} - \mathfrak{I}_{\omega\eta}\mathfrak{I}_{\eta\eta}^{-1}\mathfrak{I}_{\eta\omega}\right)\begin{pmatrix} p-1 \\ q-1 \end{pmatrix} \tag{9.1}$$

which is the noncentrality parameter of the noncentral chi-square distribution with two degrees of freedom, valid under an asymptotic local approximation.

The corresponding quantities for the test for $H_0: \ln(p) = \ln(q) = 0$ are

$$W_F^{\log} = N\left(\hat{p}\ln\hat{p} \quad \hat{q}\ln\hat{q}\right)\left(\hat{\mathfrak{I}}_{\omega\omega} - \hat{\mathfrak{I}}_{\omega\eta}\hat{\mathfrak{I}}_{\eta\eta}^{-1}\hat{\mathfrak{I}}_{\eta\omega}\right)\begin{pmatrix} \hat{p}\ln\hat{p} \\ \hat{q}\ln\hat{q} \end{pmatrix}$$

and

$$\lambda\left(W_F^{\log}\right) = N\left(p\ln p \quad q\ln q\right)\left(\mathfrak{I}_{\omega\omega} - \mathfrak{I}_{\omega\eta}\mathfrak{I}_{\eta\eta}^{-1}\mathfrak{I}_{\eta\omega}\right)\begin{pmatrix} p\ln p \\ q\ln q \end{pmatrix} \tag{9.2}$$

For a sample size of $N = 4000$, Tables S21 and S22 contain the noncentrality parameters and powers respectively for testing $H_0: p = 1, q = 1$ under the asymptotic local approximation assumption. Values are provided for *p* and *q* from 0.75 to 4.25, in increments of 0.25. Tables S23 and S24 contain the corresponding ones for testing $H_0: \ln(p) = \ln(q) = 0$. In line with Theorem 2, we note that the values are symmetric, interchanging *p* and *q* does not change the results. Thus, examining changes in power as *p* changes for a fixed value of *q*, say *c*, is identical to examining changes in power as *q* changes with *p* fixed at *c*.

Table S21 Noncentrality parameters for Wald test $H_0 : p = q = 1$ under the asymptotic local approximation assumption

	q														
p	0.75	1.00	1.25	1.50	1.75	2.00	2.25	2.50	2.75	3.00	3.25	3.50	3.75	4.00	4.25
0.75	5.41	40.29	84.47	112.56	127.41	133.81	135.17	133.63	130.47	126.48	122.12	117.65	113.25	109.00	104.95
1.00	40.29	0.00	9.99	23.27	32.61	38.01	40.58	41.31	40.88	39.76	38.25	36.54	34.74	32.95	31.19
1.25	84.47	9.99	1.47	5.21	9.85	13.18	15.12	16.00	16.15	15.83	15.23	14.46	13.62	12.74	11.87
1.50	112.56	23.27	5.21	3.44	5.03	6.76	7.96	8.59	8.79	8.69	8.39	7.99	7.53	7.04	6.56
1.75	127.41	32.61	9.85	5.03	4.78	5.47	6.11	6.48	6.61	6.54	6.36	6.10	5.81	5.51	5.21
2.00	133.81	38.01	13.18	6.76	5.47	5.52	5.79	5.97	6.01	5.94	5.79	5.60	5.38	5.17	4.96
2.25	135.17	40.58	15.12	7.96	6.11	5.79	5.83	5.88	5.86	5.77	5.64	5.47	5.30	5.13	4.97
2.50	133.63	41.31	16.00	8.59	6.48	5.97	5.88	5.86	5.80	5.70	5.57	5.41	5.26	5.11	4.98
2.75	130.47	40.88	16.15	8.79	6.61	6.01	5.86	5.80	5.73	5.62	5.48	5.34	5.19	5.05	4.93
3.00	126.48	39.76	15.83	8.69	6.54	5.94	5.77	5.70	5.62	5.50	5.37	5.23	5.08	4.95	4.84
3.25	122.12	38.25	15.23	8.39	6.36	5.79	5.64	5.57	5.48	5.37	5.23	5.09	4.95	4.82	4.70
3.50	117.65	36.54	14.46	7.99	6.10	5.60	5.47	5.41	5.34	5.23	5.09	4.95	4.80	4.67	4.55
3.75	113.25	34.74	13.62	7.53	5.81	5.38	5.30	5.26	5.19	5.08	4.95	4.80	4.66	4.52	4.40
4.00	109.00	32.95	12.74	7.04	5.51	5.17	5.13	5.11	5.05	4.95	4.82	4.67	4.52	4.38	4.25
4.25	104.95	31.19	11.87	6.56	5.21	4.96	4.97	4.98	4.93	4.84	4.70	4.55	4.40	4.25	4.11

Table S22 Hypothetical power of Wald test for $H_0 : p = q = 1$

	q														
p	0.75	1.00	1.25	1.50	1.75	2.00	2.25	2.50	2.75	3.00	3.25	3.50	3.75	4.00	4.25
0.75	0.5378	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1.00	1.0000	0.0500	0.8152	0.9942	0.9997	0.9999	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9998	0.9997	0.9995
1.25	1.0000	0.8152	0.1756	0.5212	0.8090	0.9119	0.9457	0.9567	0.9584	0.9548	0.9472	0.9359	0.9209	0.9021	0.8797
1.50	1.0000	0.9942	0.5212	0.3629	0.5058	0.6394	0.7152	0.7500	0.7602	0.7550	0.7395	0.7169	0.6893	0.6583	0.6252
1.75	1.0000	0.9997	0.8090	0.5058	0.4852	0.5428	0.5923	0.6196	0.6284	0.6240	0.6107	0.5917	0.5694	0.5454	0.5211
2.00	1.0000	0.9999	0.9119	0.6394	0.5428	0.5466	0.5676	0.5814	0.5848	0.5794	0.5679	0.5527	0.5355	0.5178	0.5005
2.25	1.0000	1.0000	0.9457	0.7152	0.5923	0.5676	0.5706	0.5747	0.5734	0.5666	0.5558	0.5426	0.5284	0.5143	0.5011
2.50	1.0000	1.0000	0.9567	0.7500	0.6196	0.5814	0.5747	0.5731	0.5688	0.5608	0.5501	0.5379	0.5252	0.5130	0.5018
2.75	1.0000	1.0000	0.9584	0.7602	0.6284	0.5848	0.5734	0.5688	0.5628	0.5542	0.5434	0.5316	0.5196	0.5082	0.4980
3.00	1.0000	1.0000	0.9548	0.7550	0.6240	0.5794	0.5666	0.5608	0.5542	0.5452	0.5343	0.5225	0.5107	0.4996	0.4897
3.25	1.0000	0.9999	0.9472	0.7395	0.6107	0.5679	0.5558	0.5501	0.5434	0.5343	0.5233	0.5114	0.4995	0.4882	0.4782
3.50	1.0000	0.9999	0.9359	0.7169	0.5917	0.5527	0.5426	0.5379	0.5316	0.5225	0.5114	0.4992	0.4870	0.4754	0.4650
3.75	1.0000	0.9998	0.9209	0.6893	0.5694	0.5355	0.5284	0.5252	0.5196	0.5107	0.4995	0.4870	0.4744	0.4622	0.4513
4.00	1.0000	0.9997	0.9021	0.6583	0.5454	0.5178	0.5143	0.5130	0.5082	0.4996	0.4882	0.4754	0.4622	0.4495	0.4378
4.25	1.0000	0.9995	0.8797	0.6252	0.5211	0.5005	0.5011	0.5018	0.4980	0.4897	0.4782	0.4650	0.4513	0.4378	0.4253

Table S23 Noncentrality parameters for Wald test $H_0 : \ln(p) = \ln(q) = 0$ under the asymptotic local approximation assumption

	q														
p	0.75	1.00	1.25	1.50	1.75	2.00	2.25	2.50	2.75	3.00	3.25	3.50	3.75	4.00	4.25
0.75	4.03	30.01	75.59	116.23	147.34	169.61	184.84	194.74	200.71	203.81	204.83	204.35	202.79	200.48	197.63
1.00	30.01	0.00	12.44	34.44	55.59	73.05	86.47	96.34	103.31	107.97	110.86	112.39	112.87	112.57	111.68
1.25	75.59	12.44	1.83	8.86	20.64	32.42	42.59	50.77	57.07	61.73	65.02	67.22	68.53	69.15	69.22
1.50	116.23	34.44	8.86	5.08	9.40	16.17	23.07	29.20	34.28	38.31	41.37	43.61	45.15	46.12	46.63
1.75	147.34	55.59	20.64	9.40	8.16	10.92	14.99	19.18	22.97	26.17	28.75	30.75	32.23	33.26	33.93
2.00	169.61	73.05	32.42	16.17	10.92	10.61	12.40	14.93	17.53	19.90	21.93	23.58	24.87	25.82	26.49
2.25	184.84	86.47	42.59	23.07	14.99	12.40	12.41	13.58	15.16	16.79	18.28	19.56	20.61	21.42	22.01
2.50	194.74	96.34	50.77	29.20	19.18	14.93	13.58	13.66	14.40	15.39	16.40	17.33	18.13	18.77	19.26
2.75	200.71	103.31	57.07	34.28	22.97	17.53	15.16	14.40	14.47	14.92	15.52	16.14	16.70	17.18	17.55
3.00	203.81	107.97	61.73	38.31	26.17	19.90	16.79	15.39	14.92	14.94	15.20	15.55	15.90	16.22	16.48
3.25	204.83	110.86	65.02	41.37	28.75	21.93	18.28	16.40	15.52	15.20	15.17	15.29	15.46	15.64	15.79
3.50	204.35	112.39	67.22	43.61	30.75	23.58	19.56	17.33	16.14	15.55	15.29	15.22	15.24	15.30	15.35
3.75	202.79	112.87	68.53	45.15	32.23	24.87	20.61	18.13	16.70	15.90	15.46	15.24	15.13	15.09	15.06
4.00	200.48	112.57	69.15	46.12	33.26	25.82	21.42	18.77	17.18	16.22	15.64	15.30	15.09	14.95	14.86
4.25	197.63	111.68	69.22	46.63	33.93	26.49	22.01	19.26	17.55	16.48	15.79	15.35	15.06	14.86	14.71

Table S24 Hypothetical power of Wald test for $H_0: \ln(p) = \ln(q) = 0$

	q														
p	0.75	1.00	1.25	1.50	1.75	2.00	2.25	2.50	2.75	3.00	3.25	3.50	3.75	4.00	4.25
0.75	0.4183	0.9992	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1.00	0.9992	0.0500	0.8947	0.9998	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1.25	1.0000	0.8947	0.2094	0.7638	0.9876	0.9996	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1.50	1.0000	0.9998	0.7638	0.5106	0.7896	0.9586	0.9938	0.9990	0.9998	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000
1.75	1.0000	1.0000	0.9876	0.7896	0.7265	0.8498	0.9439	0.9815	0.9936	0.9975	0.9989	0.9994	0.9996	0.9997	0.9998
2.00	1.0000	1.0000	0.9996	0.9586	0.8498	0.8389	0.8938	0.9430	0.9710	0.9848	0.9914	0.9947	0.9963	0.9972	0.9977
2.25	1.0000	1.0000	1.0000	0.9938	0.9439	0.8938	0.8940	0.9201	0.9462	0.9647	0.9763	0.9833	0.9875	0.9901	0.9916
2.50	1.0000	1.0000	1.0000	0.9990	0.9815	0.9430	0.9201	0.9217	0.9349	0.9493	0.9610	0.9695	0.9753	0.9793	0.9819
2.75	1.0000	1.0000	1.0000	0.9998	0.9936	0.9710	0.9462	0.9349	0.9360	0.9429	0.9510	0.9582	0.9640	0.9682	0.9712
3.00	1.0000	1.0000	1.0000	0.9999	0.9975	0.9848	0.9647	0.9493	0.9429	0.9432	0.9468	0.9513	0.9555	0.9591	0.9617
3.25	1.0000	1.0000	1.0000	1.0000	0.9989	0.9914	0.9763	0.9610	0.9510	0.9468	0.9464	0.9480	0.9503	0.9525	0.9543
3.50	1.0000	1.0000	1.0000	1.0000	0.9994	0.9947	0.9833	0.9695	0.9582	0.9513	0.9480	0.9471	0.9473	0.9481	0.9488
3.75	1.0000	1.0000	1.0000	1.0000	0.9996	0.9963	0.9875	0.9753	0.9640	0.9555	0.9503	0.9473	0.9459	0.9452	0.9449
4.00	1.0000	1.0000	1.0000	1.0000	0.9997	0.9972	0.9901	0.9793	0.9682	0.9591	0.9525	0.9481	0.9452	0.9434	0.9420
4.25	1.0000	1.0000	1.0000	1.0000	0.9998	0.9977	0.9916	0.9819	0.9712	0.9617	0.9543	0.9488	0.9449	0.9420	0.9397

Comparing the hypothetical power in Tables S22 and S24 with the results reported in Figures 6, 7 and 8 in the body of the paper gives us an indication of whether the asymptotic local approximation is a good one and provides some insights into the power behavior reported in Figures 6, 7 and 8. In those figures power was considered under three scenarios.

(i) Both *p* and *q* are changed while maintaining $p = q$.

(ii) Only *p* was varied while *q* was set at its hypothesized value $q = 1$.

(iii) Only *p* was varied while *q* was set at $q = 1.75$.

Consider **scenario (i)** which relates to the diagonal elements in Tables S22 and S24. For the Wald test for $H_0 : p = q = 1$ with $p = q > 1$, the hypothetical power gradually increases, reaching a maximum of 0.57 when $p = q = 2.5$, and then gradually declines reaching a value of 0.43 when $p = q = 4.25$. The actual power of the test displayed in Figure 6 is much worse. It declines at first, such that the power is less than the 0.05 significance level, then it increases and declines in a similar way to the hypothetical power, but its values are much less. Its maximum power, reached at $p = q = 2.5$, is only 0.2. This outcome contrasts with that from the test for the Singh-Maddala distribution where hypothetical power understated the actual power of the Wald test for $p > 1.5$. Examining the diagonal elements of Table S24 where hypothetical power for $H_0 : \ln(p) = \ln(q) = 0$ is displayed, we find the power function behaves as one would like a power function to behave, but it understates the actual power given in Figure 6 of the paper.

Consider **scenario (ii)** where the null hypothesis is either $H_0 : p = q = 1$ or $H_0 : \ln(p) = \ln(q) = 0$ and we set $q = 1$, making the null hypothesis partially true for all values of *p*. The relevant values for hypothetical power are those in the third columns of Tables S22 and S24. In both cases these powers are very good, becoming approximately 1 for $p = 0.75$ and $p \geq 0.75$. The results in Figure 7 in the body of the paper suggest that actual power is even greater than hypothetical power.

In **scenario (iii)** where $q = 1.75$ for all values of *p*, the values for hypothetical power are those in the sixth columns of Tables S22 and S24. For testing $H_0 : p = q = 1$, this power declines as *p* increases reaching a minimum of 0.49 when $p = q = 1.75$, then increases to a maximum of 0.63 when $p = 2.75$, and then declines again. The behavior of actual power displayed in Figure 8 in the paper is similar, but the minimum power, reached at $p=q=1.75$, is much worse at approximately 0.1. The same phenomenon is observed when testing $H_0:\ln(p)=\ln(q)=0$, but the initial decline is more moderate, reaching a minimum power of 0.73, and, at least for the values of *p* considered, there is no second decline. An initial decline in actual power is also observed in Figure 7, but its minimum of 0.9 is even greater than the hypothetical power minimum.

S9.2 LM test

An interesting feature of the LM test for $H_0:p=1,q=1$ is that the covariance matrix for $(\hat{p},\hat{q})$, evaluated at the restricted estimates does not depend on any of the parameters. It is a result that follows from Theorem 1. Since the covariance matrix for $(\hat{p},\hat{q})$ does not depend on *a* or *b*, evaluating it at the hypothesized values $p=q=1$ means that it will not depend on any of the parameters. The LM statistic is given in the following theorem. This theorem has practical implications, because this LM statistic only requires the restricted estimates of *a* and *b* and the data. The LM test performs well as we have documented.

Theorem 3

The LM statistic for testing $H_0:p=1,q=1$ is given by

$$\text{LM}=N\left[95.95958\left(\tilde{g}_3^2+\tilde{g}_4^2\right)+178.11978\tilde{g}_3\tilde{g}_4\right]$$

where

$$\tilde{g}_3=1+\frac{1}{N}\sum_{i=1}^{N}\ln\left(\frac{\left(x_i/\tilde{b}\right)^{\tilde{a}}}{1+\left(x_i/\tilde{b}\right)^{\tilde{a}}}\right)\qquad\qquad \tilde{g}_4=1-\frac{1}{N}\sum_{i=1}^{N}\ln\left[1+\left(x_i/\tilde{b}\right)^{\tilde{a}}\right]$$

and $\tilde{a}$ and $\tilde{b}$ are the ML estimates for *a* and *b* assuming $p=q=1$.

Theorem 3 Proof

If $\ln L_i$ is the log-likelihood for a single observation, $\tilde{\mathbf{g}}_i=\partial\ln L_i/\partial\boldsymbol{\theta}\big|_{\tilde{\boldsymbol{\theta}}}$ is the gradient vector for the *i*th observation evaluated at $\tilde{\boldsymbol{\theta}}=\left(\tilde{a}\,\tilde{b}\,1\,1\right)'$ and $\tilde{\mathbf{g}}=N^{-1}\sum_{i=1}^{N}\tilde{\mathbf{g}}_i$, the LM test statistic for $H_0:p=1,q=1$ is (Ruud, 2000, p.385)

$$\text{LM}=N\begin{pmatrix}\tilde{g}_3 & \tilde{g}_4\end{pmatrix}\left(\tilde{\mathfrak{I}}_{\omega\omega}-\tilde{\mathfrak{I}}_{\omega\eta}\tilde{\mathfrak{I}}_{\eta\eta}^{-1}\tilde{\mathfrak{I}}_{\eta\omega}\right)^{-1}\begin{pmatrix}\tilde{g}_3\\ \tilde{g}_4\end{pmatrix}$$

Note that

$$\frac{\partial \ln L_i}{\partial p} = \ln y_i - \psi(p) + \psi(p+q) = \ln\left(\frac{(x_i/b)^a}{1+(x_i/b)^a}\right) - \psi(p) + \psi(p+q)$$

$$\frac{\partial \ln L_i}{\partial q} = \ln(1-y_i) - \psi(q) + \psi(p+q) = -\ln\left[1+(x_i/b)^a\right] - \psi(q) + \psi(p+q)$$

Using the result $\psi(z+1) = \psi(z) + 1/z$, we have

$$\tilde{g}_3 = \frac{1}{N}\sum_{i=1}^{N}\ln\left(\frac{(x_i/\tilde{b})^{\tilde{a}}}{1+(x_i/\tilde{b})^{\tilde{a}}}\right) - \psi(1) + \psi(2) = \frac{1}{N}\sum_{i=1}^{N}\ln\left(\frac{(x_i/\tilde{b})^{\tilde{a}}}{1+(x_i/\tilde{b})^{\tilde{a}}}\right) - \psi(1) + \psi(1) + 1$$

$$= \frac{1}{N}\sum_{i=1}^{N}\ln\left(\frac{(x_i/\tilde{b})^{\tilde{a}}}{1+(x_i/\tilde{b})^{\tilde{a}}}\right) + 1$$

$$\tilde{g}_4 = -\frac{1}{N}\sum_{i=1}^{N}\ln\left[1+(x_i/\tilde{b})^{\tilde{a}}\right] - \psi(1) + \psi(2) = 1 - \frac{1}{N}\sum_{i=1}^{N}\ln\left[1+(x_i/\tilde{b})^{\tilde{a}}\right]$$

Also, using $\psi'(z+1) = \psi'(z) - 1/z^2$, and $\psi'(1) = 1.6449340655$, we have

$$\tilde{\mathfrak{I}}_{\omega\omega} = \begin{pmatrix} \psi'(p) - \psi'(p+q) & -\psi'(p+q) \\ -\psi'(p+q) & \psi'(q) - \psi'(p+q) \end{pmatrix}\Bigg|_{p=1,\,q=1} = \begin{pmatrix} \psi'(1) - [\psi'(1) - 1] & -[\psi'(1) - 1] \\ -[\psi'(1) - 1] & \psi'(1) - [\psi'(1) - 1] \end{pmatrix}$$

$$= \begin{pmatrix} 1 & -0.6449340655 \\ -0.6449340655 & 1 \end{pmatrix}$$

$$\tilde{\mathfrak{I}}_{\omega\eta} = -E\begin{pmatrix} \dfrac{\partial^2 \ln L_i}{\partial p \partial a} & \dfrac{\partial^2 \ln L_i}{\partial p \partial b} \\ \dfrac{\partial^2 \ln L_i}{\partial q \partial a} & \dfrac{\partial^2 \ln L_i}{\partial q \partial b} \end{pmatrix}\Bigg|_{\tilde{\theta}} = \begin{pmatrix} \dfrac{1}{2\tilde{a}} & \dfrac{\tilde{a}}{2\tilde{b}} \\ \dfrac{1}{2\tilde{a}} & -\dfrac{\tilde{a}}{2\tilde{b}} \end{pmatrix}$$

$$\tilde{\mathfrak{I}}_{\eta\eta} = -E\begin{pmatrix} \dfrac{\partial^2 \ln L_i}{\partial a^2} & \dfrac{\partial^2 \ln L_i}{\partial a \partial b} \\ \dfrac{\partial^2 \ln L_i}{\partial b \partial a} & \dfrac{\partial^2 \ln L_i}{\partial b^2} \end{pmatrix}\Bigg|_{\tilde{\theta}} = \begin{pmatrix} \dfrac{1+2\psi'(1)}{3\tilde{a}^2} & 0 \\ 0 & \dfrac{\tilde{a}^2}{3\tilde{b}^2} \end{pmatrix}$$

$$\left(\tilde{\mathfrak{I}}_{\eta\eta}\right)^{-1} = \begin{pmatrix} \dfrac{3\tilde{a}^2}{1+2\psi'(1)} & 0 \\ 0 & \dfrac{3\tilde{b}^2}{\tilde{a}^2} \end{pmatrix}$$

$$\tilde{\mathfrak{I}}_{\omega\eta}\tilde{\mathfrak{I}}_{\eta\eta}^{-1}\tilde{\mathfrak{I}}_{\eta\omega} = \begin{pmatrix} \frac{1}{2\tilde{a}} & \frac{\tilde{a}}{2\tilde{b}} \\ \frac{1}{2\tilde{a}} & -\frac{\tilde{a}}{2\tilde{b}} \end{pmatrix} \begin{pmatrix} \frac{3\tilde{a}^2}{1+2\psi'(1)} & 0 \\ 0 & \frac{3\tilde{b}^2}{\tilde{a}^2} \end{pmatrix} \begin{pmatrix} \frac{1}{2\tilde{a}} & \frac{1}{2\tilde{a}} \\ \frac{\tilde{a}}{2\tilde{b}} & -\frac{\tilde{a}}{2\tilde{b}} \end{pmatrix} = \begin{pmatrix} \frac{3\tilde{a}}{2\left(1+2\psi'(1)\right)} & \frac{3\tilde{b}}{2\tilde{a}} \\ \frac{3\tilde{a}}{2\left(1+2\psi'(1)\right)} & \frac{-3\tilde{b}}{2\tilde{a}} \end{pmatrix} \begin{pmatrix} \frac{1}{2\tilde{a}} & \frac{1}{2\tilde{a}} \\ \frac{\tilde{a}}{2\tilde{b}} & -\frac{\tilde{a}}{2\tilde{b}} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{3\left(1+\psi'(1)\right)}{2\left(1+2\psi'(1)\right)} & \frac{-3\psi'(1)}{2\left(1+2\psi'(1)\right)} \\ \frac{-3\psi'(1)}{2\left(1+2\psi'(1)\right)} & \frac{3\left(1+\psi'(1)\right)}{2\left(1+2\psi'(1)\right)} \end{pmatrix} = \frac{3}{2}\begin{pmatrix} 0.6165537 & -0.3834463 \\ -0.3834463 & 0.6165537 \end{pmatrix}$$

$$\tilde{\mathfrak{I}}_{\omega\omega} - \tilde{\mathfrak{I}}_{\omega\eta}\tilde{\mathfrak{I}}_{\eta\eta}^{-1}\tilde{\mathfrak{I}}_{\eta\omega} = \begin{pmatrix} 1 & -0.644934 \\ -0.644934 & 1 \end{pmatrix} - \frac{3}{2}\begin{pmatrix} 0.6165537 & -0.3834463 \\ -0.3834463 & 0.6165537 \end{pmatrix}$$

$$= \begin{pmatrix} 0.07516945 & -0.06976461 \\ -0.06976461 & 0.07516945 \end{pmatrix}$$

and

$$\left(\tilde{\mathfrak{I}}_{\omega\omega} - \tilde{\mathfrak{I}}_{\omega\eta}\tilde{\mathfrak{I}}_{\eta\eta}^{-1}\tilde{\mathfrak{I}}_{\eta\omega}\right)^{-1} = \begin{pmatrix} 0.07516945 & -0.06976461 \\ -0.06976461 & 0.07516945 \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} 95.95958 & 89.05989 \\ 89.05989 & 95.95958 \end{pmatrix}$$

Thus,

$$\mathrm{LM} = N\begin{pmatrix} \tilde{g}_3 & \tilde{g}_4 \end{pmatrix} \begin{pmatrix} 95.95958 & 89.05989 \\ 89.05989 & 95.95958 \end{pmatrix} \begin{pmatrix} \tilde{g}_3 \\ \tilde{g}_4 \end{pmatrix}$$

$$= N\left[95.95958\left(\tilde{g}_3^2 + \tilde{g}_4^2\right) + 178.11978\tilde{g}_3\tilde{g}_4\right]$$

S.10 The Monte Carlo experiments and numerical issues

Maximum likelihood estimation and testing of the GB2 model presents challenges even though there are only four parameters. Estimation and hypothesis testing uses the analytical expressions for the derivatives of the log-likelihood function and the information matrix given in Supplement Section S1.3. We use Gauss 24[6] for calculations and use the Gauss module OPTMUM 3.1 to obtain maximum likelihood estimates. This module minimizes an objective function, which in our case is the negative of

[6] GAUSS Version 24, Aptech Systems, Inc. (2023). www.aptech.com. Estimates match those provided by the Stata module GB2LFIT by Jenkins (2007).

the log-likelihood function. We use 10,000 Monte Carlo samples with N = 4000 observations. To generate samples of data from the GB2 distribution, we use the representation in Cummins et al. (1990, p. 259). If X_1 and X_2 are independent gamma random variables, with scale parameters one, and shape parameters p and q respectively, then

$$X = b\left(X_1/X_2\right)^{1/a} \sim GB2(a,b,p,q)$$

As noted in Section S5, the four parameters a, b, p, and q are positive. The parameter b adjusts the scale of the distribution. The other parameters affect the shape of the distribution. If $f(x)$ is the GB2 distribution, the mode occurs at $x = b\left(\dfrac{ap-1}{aq+1}\right)^{1/a}$, if $ap > 1$.[7] The kth moments of the distribution, for $-ap < k < aq$, are

$$E\left(X^k\right) = \frac{b^k B\left(p + k/a, q - k/a\right)}{B\left(p,q\right)}$$

where $B\left(p,q\right)$ is the beta function.[8]

OPTMUM offers several choices of algorithm direction and step length. We use the quasi-Newton method BFGS (Broyden-Fletcher-Goldfarb-Shanno). Press, et al. (2007, 521) say "…. It has become generally recognized that, empirically, the BFGS scheme is superior…." Interestingly, step length choice matters a great deal. After extensive trial and error, we chose a two-phase method. In the first phase we employ a method due to Brent (2002, Chapter 5). Press, et al. (2007, 496) conclude that "*Brent's method* is up to the task in all particulars." The particulars include a "…sure-but-slow technique, like golden section search, but that switches over to [an inverse parabolic interpolation] when the function allows." The second phase, which begins after 20 iterations, is "step selection by backtracking" [Dennis and Schnabel (1996, 126-129)] which is a quadratic/cubic interpolation method.

We employed the log-metric transformed (negative) log-likelihood as our objective function, estimating $\boldsymbol{\phi} = \left(\ln(a)\ \ln(b)\ \ln(p)\ \ln(q)\right)'$. This change dramatically reduced failures to converge

---

[7] Kleiber and Kotz (2003, p.188)

[8] Ibid.

relative to estimating the untransformed parameters. For example, when estimating the model for 10,000 samples over 36 values of p = 0.5 to 4, we rejected, and redrew, seven samples, or 0.0019%. The rejections were primarily due to failures computing the log-likelihood function, which includes the natural logarithm of the beta function, $\ln B(p,q) = \ln\Gamma(p) + \ln\Gamma(q) - \ln\Gamma(p+q)$. We also rejected a few samples in which the Hessian, information matrix or outer product of the gradient matrix (OPG) was not positive definite at the MLE.

Figure S18 illustrates the increasing estimator variance for larger parameter values. The increasing variance of $\hat{p}$ as p increases is a consequence of the information matrix, Hessian, and outer product of the gradients, becoming more poorly conditioned. This has computational implications as well as its impact on the power of the test. Following Press, et al. (2007, 69), we define a condition index to be the ratio of the largest singular value to the smallest singular value.

Luenberger and Ye (2021, 259-261) discuss the steepest descent method for a quadratic function of the form $f(\mathbf{x}) = 0.5\mathbf{x}'\mathbf{Q}\mathbf{x} - \mathbf{x}'\mathbf{b}$, where **Q** is positive definite, and **b** is a vector of constants. They note that "…the rate of convergence depends on the ratio of the lengths of the axes of the elliptical contours of the…" objective function. For a quadratic function, these lengths are proportional to the reciprocal of the eigenvalues. They state (p. 261) "…that the convergence rate actually depends only on the ratio…of the largest to the smallest eigenvalue" of **Q**. Luenberger and Ye define the condition number using the eigenvalue ratio, but of course the ratio of the singular values gives the same information. Their lesson concerning quadratic functions carries over to "Newton-like" algorithms. We show illustrations in Figures S22(a)-S22(d) that demonstrate these concerns. As p increases the ill-conditioning becomes worse, and as Luenberger and Ye describe, "eccentricity of the ellipsoids" is more pronounced, making convergence slower. In Figure S23(a)-S23(d) are the contours in the log-metric. These are visibly less "eccentric" than those in Figures S22(a)-S22(d), again providing evidence for estimation in the log-metric.

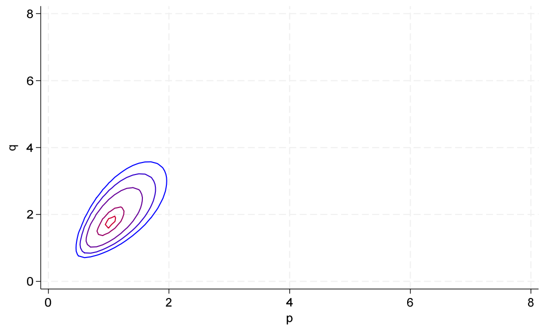
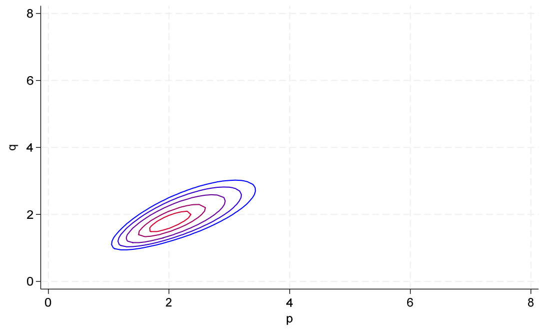
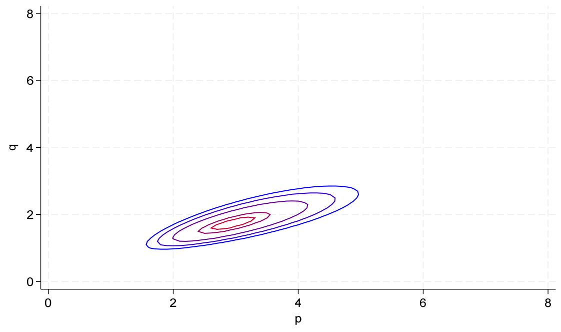
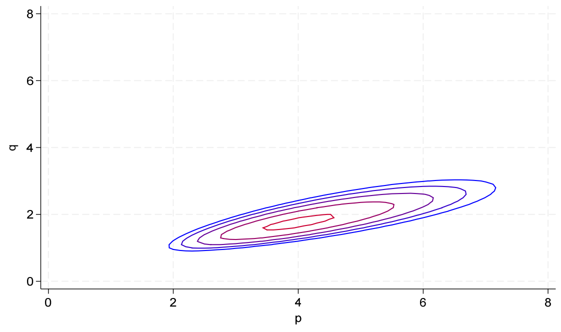
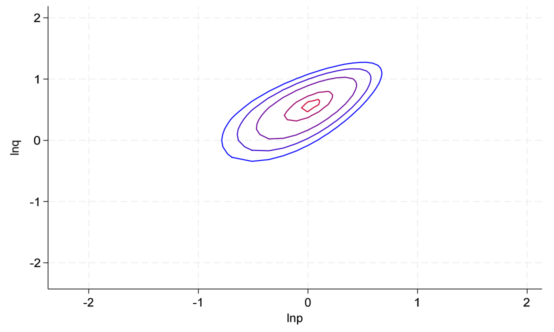
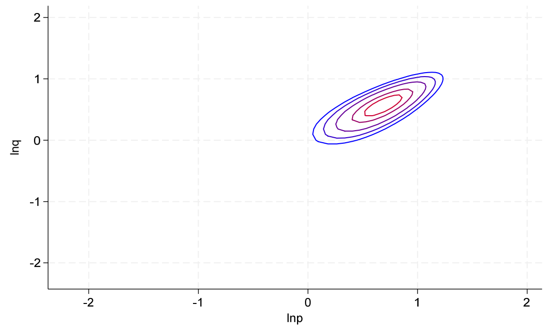
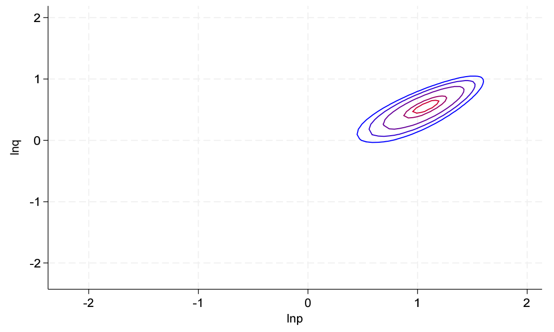
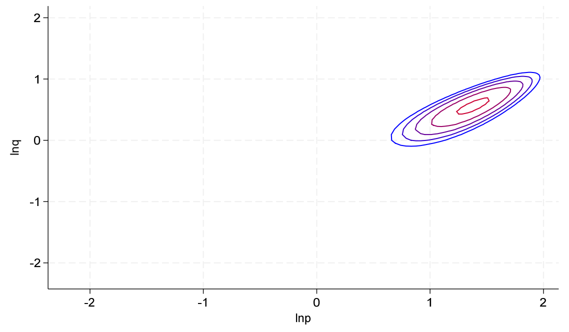
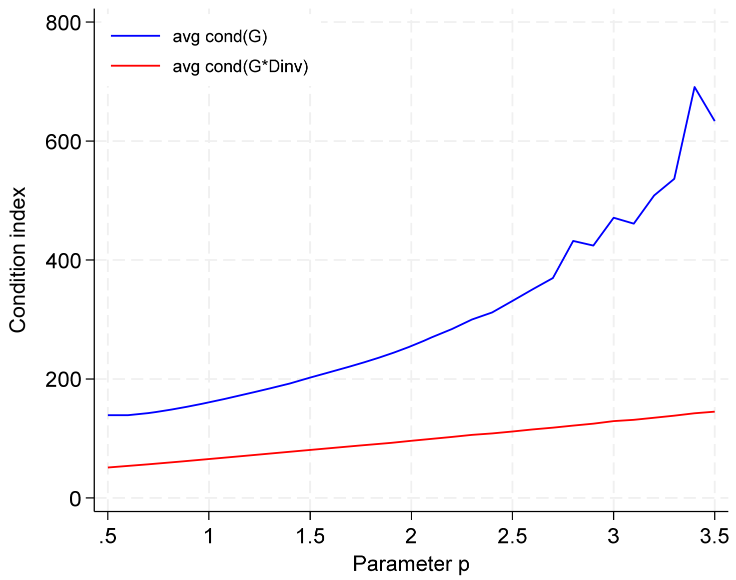
Figure S22(a) Contours of a log-likelihood function when $a = 5$, $b = 1$, $p = 1$ and $q = 1.75$.	
Figure S22(b) Contours of a log-likelihood function when $a = 5$, $b = 1$, $p = 2$ and $q = 1.75$.	
Figure S22(c) Contours of a log-likelihood function when $a = 5$, $b = 1$, $p = 3$ and $q = 1.75$.	
Figure S22(d) Contours of a log-likelihood function when $a = 5$, $b = 1$, $p = 4$ and $q = 1.75$.	

Figure S23(a) Contours of a log-likelihood function when $a = 5$, $b = 1$, $p = 1$ and $q = 1.75$ in log-metric.	
Figure S23(b) Contours of a log-likelihood function when $a = 5$, $b = 1$, $p = 2$ and $q = 1.75$ in log-metric.	
Figure S23(c) Contours of a log-likelihood function when $a = 5$, $b = 1$, $p = 3$ and $q = 1.75$ in log-metric.	
Figure S23(d) Contours of a log-likelihood function when $a = 5$, $b = 1$, $p = 4$ and $q = 1.75$ in log-metric.	

In maximum likelihood estimation of the parameters of the GB2 distribution, the condition indexes are large in magnitude using either the original or log-transformed data.[9] In Figure S24 we plot the average condition numbers, over the 10,000 samples, for the gradient matrix $\mathbf{G}$ and the log-metric transformation $\mathbf{GD}^{-1}$, where $\mathbf{D} = \text{diag}\left(a^{-1}\; b^{-1}\; p^{-1}\; q^{-1}\right)$. The log-transformation has much lower average condition indexes.[10] This is documented in Table S25. There is considerable variation in the condition indexes from sample to sample. The first three columns report the mean, standard deviation and median of the condition index of $\mathbf{G}$. Note that when $p = 3.6$ and $p = 3.7$ there are huge jumps in the mean and standard deviation, indicating one or more unusual samples. The median shows no such jump. The remaining columns report the ratio of the condition indexes of $\mathbf{GD}^{-1}$ relative to those of $\mathbf{G}$. The condition indexes of $\mathbf{GD}^{-1}$ vary between 28% to 41% of the condition indexes of $\mathbf{G}$. The transformation $\mathbf{GD}^{-1}$ multiplies each column in $\mathbf{G}$ by a constant, a transformation that will not change the correlations between the columns, but it does increase the variability of the columns, improving the condition indexes in the log-transformed data relative to the untransformed data. The relatively poor condition index for the estimator of (a, b, p, q) explains why Jenkins (2014) found estimation of the log-transformed parameters more reliable.

Figure S24 Conditioning diagnostics**.** The average condition indexes for the gradient matrices $\mathbf{G}$ and $\mathbf{GD}^{-1}$. These are based on 10,000 Monte Carlo samples. The settings are $N = 4000$, $b = 1$, $a = 5$, and $q = 1.75$.



[9] The ideal condition index is one. Beyond that, larger is worse. Luenberger and Ye (2021, pp. 260-261) define a convergence ratio that they interpret in the quadratic case for the steepest descent algorithm. For the general case, there is no scale measuring severity, like that provided by Belsley, Kuh and Welsch (1980) for the linear regression model.

[10] The singular values of $\mathbf{G}$ are the square roots of the singular values of $\mathbf{G'G}$, which are the eigenvalues of $\mathbf{G'G}$. The eigenvalues of $\left(\mathbf{G'G}\right)^{-1}$ are the reciprocals of the eigenvalues of $\mathbf{G'G}$. Thus, the singular values of $\mathbf{G}$ and $\mathbf{GD}^{-1}$ are directly related to the estimator variances as $\left(\mathbf{G'G}\right)^{-1}$ is an estimator of the information matrix. The same results hold when dissecting the Hessian or the information matrix.

Table S25 Condition indexes of $\mathbf{G}$ and $\mathbf{GD}^{-1}$

	Condition index of $\mathbf{G}$			Condition index of $\mathbf{GD}^{-1}$ ratio		
p	Mean	Std	Median	Mean ratio	Std ratio	Median ratio
0.5	139.12	16.65	136.87	0.3673	0.3698	0.3653
0.6	139.09	14.45	137.09	0.3869	0.4223	0.3846
0.7	142.69	13.65	141.00	0.3965	0.4614	0.3936
0.8	147.48	13.26	145.94	0.4027	0.4876	0.3993
0.9	154.02	14.83	152.11	0.4050	0.4743	0.4023
1	160.68	15.02	158.64	0.4067	0.4989	0.4037
1.1	168.24	16.63	165.71	0.4076	0.4931	0.4055
1.2	176.14	18.76	172.83	0.4071	0.4743	0.4053
1.3	184.06	22.86	179.99	0.4042	0.4248	0.4034
1.4	193.22	30.43	187.62	0.4023	0.3560	0.4037
1.5	201.93	29.24	195.12	0.3998	0.3880	0.4030
1.6	211.21	36.18	202.80	0.3959	0.3456	0.4001
1.7	220.77	38.40	211.21	0.3922	0.3374	0.3989
1.8	232.30	45.57	220.14	0.3875	0.3142	0.3966
1.9	243.29	56.15	228.43	0.3818	0.2762	0.3931
2	256.34	65.15	237.51	0.3754	0.2598	0.3906
2.1	270.30	115.59	248.14	0.3682	0.1629	0.3863
2.2	284.49	117.82	257.57	0.3613	0.1697	0.3828
2.3	298.15	106.81	268.03	0.3545	0.1966	0.3780
2.4	314.54	134.26	279.91	0.3464	0.1675	0.3727
2.5	330.90	164.06	289.91	0.3379	0.1493	0.3681
2.6	360.90	930.89	301.95	0.3194	0.0339	0.3622
2.7	370.10	419.07	313.54	0.3189	0.0678	0.3572
2.8	389.83	261.64	325.39	0.3114	0.1137	0.3523
2.9	414.00	313.15	340.91	0.3016	0.1029	0.3455
3	461.29	1911.92	355.93	0.2788	0.0210	0.3395
3.1	464.35	380.20	371.45	0.2835	0.0936	0.3325
3.2	519.63	1069.17	385.50	0.2619	0.0410	0.3272
3.3	541.73	1557.40	400.76	0.2554	0.0291	0.3218
3.4	579.35	1756.84	418.66	0.2450	0.0273	0.3151
3.5	780.54	11195.63	436.29	0.1876	0.0067	0.3083
3.6	23082.97	2241785.37	450.74	0.0067	0.0003	0.3043
3.7	178708.21	14076740.90	469.35	0.0010	0.0001	0.2971
3.8	814.97	4869.58	487.46	0.1919	0.0146	0.2921
3.9	871.78	7967.55	505.67	0.1832	0.0094	0.2870
4	10617.25	888621.31	527.55	0.0159	0.0004	0.2805

References

Belsley, D. A., Kuh, E., Welsch, R. E. (1980), Regression Diagnostics: Identifying Influential Data and Sources of Collinearity, New York: John Wiley & Sons, Inc.

Brent, R. P. (2002), Algorithms for Minimization Without Derivatives, Mineola, New York: Dover Publications, Inc., Kindle Edition.

Cummins, J. D., Dionne, G., McDonald, J. B. and Pritchett, B. M. (1990), "Applications of the GB2 family of distributions in modeling insurance loss processes," Insurance: Mathematics and Economics, 9, 257-272, North-Holland.

Dennis, Jr., J. E., and Schnabel, R. B. (1996), Numerical Methods for Unconstrained Optimization and Nonlinear Equations, Classics in applied mathematics, Society for Industrial and Applied Mathematics.

GAUSS Version 24 [Computer software]. Aptech Systems, Inc. (2023). Retrieved from www.aptech.com.

Gradshteyn, I. S. and Ryzik, I. M. (1980), Tables of Integrals, Series and Products, Corrected and Enlarged Edition prepared by Alan Jeffery, San Diego, California: Academic Press.

Greene, W. H. (2018), Econometric Analysis (8th ed.), Upper Saddle River, NJ: Pearson, Upper Saddle River, New Jersey.

Jenkins, S. P. (2007), "GB2FIT: Stata module to fit Generalized Beta of the Second Kind distribution by maximum likelihood," Statistical Software Components S456823, Boston College Department of Economics, revised 17 Jul 2012.

Jenkins, S. P. (2014), "GB2LFIT: Stata module to fit Generalized Beta of the Second Kind distribution by maximum likelihood (log parameter metric)," Statistical Software Components S457897, Boston College Department of Economics.

Kleiber, C. (1996), "Dagum versus Singh-Maddala Income Distributions", *Economics Letters*, 53, 265-268.Kleiber, C. and Kotz, S. (2003), Statistical Size Distributions in Economics and Actuarial Sciences, Hoboken, New Jersey: John Wiley & Sons, Inc.

Luenberger, David G. and Ye, Y. (2021), Linear and Nonlinear Programming (5th ed.), Springer.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2007), Numerical Recipes: The Art of Scientific Computing (3rd ed.), Cambridge: Cambridge University Press.

Ruud, P. A. (2000), An Introduction to Classical Econometric Theory, Oxford: Oxford University Press.