Improved Tests for Mediation

Hillier, G.; van Garderen, K.J.; van Giersbergen, N.

DOI
10.47004/wp.cem.2022.0122

Publication date
2021

Document Version
Final published version

cemmap

centre for microdata methods and practice

Improved tests for mediation

Grant Hillier
Kees Jan van Garderen
Noud van Giersbergen

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP01/22

Economic and Social Research Council

Improved tests for mediation

Grant Hillier
CeMMAP and University of Southampton

Kees Jan van Garderen
University of Amsterdam

Noud van Giersbergen
University of Amsterdam

December 2021

Abstract

Testing for a mediation effect is important in many disciplines, but is made difficult - even asymptotically - by the influence of nuisance parameters. Classical tests such as likelihood ratio (LR) and Wald tests have very poor size and power properties in some parts of the parameter space, and many attempts have been made to produce improved tests, with limited success. In this paper we show that augmenting the critical region of the LR test can produce a test with much improved behaviour everywhere. In fact, we first show that there exists a test of this type that is (asymptotically) exact for certain test sizes α, including the common choices $\alpha = .01, .05, .10$. This is evidently an important result, but we also observe that the critical region of this exact test has some undesirable properties. Thus, we then go on to show that there is a very simple class of augmented LR critical regions which provides tests that, while not exact, are very nearly so, and which avoid the issues inherent in the exact test. We suggest an optimal member of this class, and provide the tables needed to implement it. Although motivated by a simple two-equation linear model, the results apply to any model structure that reduces to the same testing problem asymptotically. A short application of the method to an entrepreneurial attitudes study is included for illustration.

1 Introduction

Testing for a mediation effect has important applications in many disciplines, including psychology, accounting, marketing, sociology, epidemiology, and economics.[1] A simple context for the problem - which we will use to motivate the results to follow - is a model of the type

$$y = \tau x + \beta m + u_1, \tag{1}$$
$$m = \theta x + u_2. \tag{2}$$

Here y, x, and m are $n \times 1$ vectors of observed variables, y being the variable of ultimate interest.[2] The vectors u_1, u_2 are unobserved random errors. The variable m is potentially a mediating variable, in that the influence of x on y may be both direct (the term τx), and/or indirect via the term βm if θ is non-zero in the second equation. There is no mediation effect if either $\beta = 0$, or $\theta = 0$, or both, so a test for the absence of a mediation effect is a test of $H_0 : \beta\theta = 0$.

This testing problem is complicated by the fact that, even asymptotically, there is a nuisance parameter present under the null - either β or θ may be non-zero - and this seriously impacts on the properties of most of the tests that have been proposed for the problem - see MacKinnon et al. (2002) for a survey. Typically, the extant tests have very poor size and power behaviour near the origin ($\beta = \theta = 0$). Specifically, the null rejection probability (NRP) of the test can be very much smaller than its nominal size - near zero in fact - and its power and size can be very nearly equal. The bank of standard tests exhibiting this behaviour all reject the null hypothesis when some particular test statistic is large. However, in a recent paper Van Garderen and Van Giersbergen (2021) have shown that both size and power can be improved considerably - particularly near the origin - by using a critical region that cannot be defined in this way, but is simply a subset of a two-dimensional sample space. After reducing the problem by invariance, they consider a critical region (CR) consisting of the likelihood ratio region (CR_{LR}), augmented by an additional region closer to the origin. This additional region is carefully constructed using a piecewise-linear spline, and is optimized in terms of both size and power.

In this paper we employ the same idea - augmenting the CR_{LR} by an additional region. We first show that, for certain test sizes (including the popular choices $\alpha = .01, .05$, and $.10$) an (asymptotically) exact test of this type does exist, and we show how to construct it. However, this exact test has some undesirable characteristics, so we then go on to discuss another augmented CR_{LR} constructed by an alternative, very simple, device. We show that there are tests in this class that have the correct size (i.e., maximum NRP), and identify the best of these in terms of power. This test is very nearly, but not quite, exact, and it does remedy some of the unsatisfactory aspects of the exact test. Tables are provided with the data needed to implement the new test. The new tests are far superior to the LR test in terms of size, and, trivially (because their critical regions are larger), also have greater power. We begin by reducing the dimension of the relevant sample space by sufficiency and invariance.

For any test with critical region w, and where the distribution of the statistics involved depends on a vector of parameters ψ, we denote the power of the test by $P_w(\psi)$, and the NRP when the null distribution depends on the parameter ψ_0 by $P_w(\psi_0)$. The *size* of the test is as usual defined to be $\sup_{\psi_0} P_w(\psi_0)$. We emphasize that the issue we are concerned

[1] See, for example, Baron and Kenny (1986), Coletti et al. (2005), MacKenzie et al. (1986), Alwin and Hauser (1975), Freedman and Schatzkin (1992), Heckman and Pinto (2015a,b).

[2] The vectors y, x, and m may be the projections of some initial variables onto a subspace associated with a more extensive model - e.g., residuals. This has no bearing on what follows.

with here is not that of finding tests of the correct size - the LR test has this property - but that the usual tests can have NRP and power that are near zero in certain parts of the parameter space.

2 Model, testing problem, invariance

The model is, under Gaussian assumptions:[3]

$$\begin{aligned} y|m &\sim N(x\tau + \beta m, \sigma_{11} I_n), \quad (3) \\ m &\sim N(x\theta, \sigma_{22} I_n). \quad (4) \end{aligned}$$

Here, y, x, and m are $n \times 1$ vectors of observables; y is the interest variable, and m is the so-called mediation variable: the influence of x on y may be both direct (the term $x\tau$), and indirect via m if neither θ nor β is zero. To test for the *absence* of a mediation effect we want to test the composite null hypothesis $H_0 : \theta\beta = 0$, i.e., that either m does not appear in (1), or x does not appear in (2), or both. The variances σ_{11}, σ_{22} are assumed unknown to begin with. A more general version of the model would have x $n \times k$, and m $n \times p$. We deal here only with the case $k = p = 1$, but briefly discuss the more general case in Section 7.

The statistics of interest are the sufficient statistics in the Gaussian model:

$$\begin{pmatrix} \hat{\tau} \\ \hat{\beta} \end{pmatrix} = [(x, m)'(x, m)]^{-1}(x, m)'y, s_{11} = y' M_{x,m} y \quad (5)$$

from equation (1), and

$$\hat{\theta} = (x'x)^{-1}x'm, s_{22} = m' M_x m \quad (6)$$

from equation (2). Here, for any matrix A of full column rank, $M_A = I_n - A(A'A)^{-1}A'$. The distributions of these statistics are, respectively:

$$\begin{pmatrix} \hat{\tau} \\ \hat{\beta} \end{pmatrix} |m \sim N\left(\begin{pmatrix} \tau \\ \beta \end{pmatrix}, \sigma_{11}[(x, m)'(x, m)]^{-1} \right), \quad (7)$$

$s_{11}/\sigma_{11} \sim \chi^2(n-2)$, $\hat{\theta} \sim N(\theta, \sigma_{22}/s_{xx})$, where $s_{xx} = x'x$, and $s_{22}/\sigma_{22} \sim \chi^2(n-1)$. The joint density of the sufficient statistics under Gaussian assumptions may be written down directly from these facts, and is equivalent to the likelihood for $(\tau, \beta, \theta, \sigma_{11}, \sigma_{22})$. However, we will not use the Gaussian joint density directly in this paper.

Not surprisingly in view of H_0, the testing problem possesses some important invariance properties that reduce the dimension of the relevant statistics - the maximal invariants - to two, rather than five. These are as described in the following:

Theorem 1 *The testing problem is invariant under the group* $\mathbf{K} = \{a_1, a_2, c : a_1, a_2 > 0, c \in \mathbb{R}\}$ *of transformations acting on* $(\hat{\tau}, \hat{\beta}, s_{11}, \theta, s_{22})$ *by*

$$(\hat{\tau}, \hat{\beta}, s_{11}, \hat{\theta}, s_{22}) \to (\sqrt{a_1}\,(\hat{\tau} + c)\,, \sqrt{a_1/a_2}\hat{\beta}, a_1 s_{11}, \sqrt{a_2}\theta, a_2 s_{22}).$$

A sample-space maximal invariant under this group of transformations is

$$T_1 = \hat{\beta}/\sqrt{s_{11}/s_{22}}, T_2 = \hat{\theta}/\sqrt{s_{22}/s_{xx}}. \quad (8)$$

[3]The Gaussian assumption is used - by invoking sufficiency and invariance - to shrink the relevant sample space to that of a two-dimensional statistic, T. However, we will later restrict attention to the asymptotic distribution of the statistic T, which is valid under much more general assumptions.

The induced group of transformations on $(\tau, \beta, \sigma_{11}, \theta, \sigma_{22})$ *is* $(\tau, \beta, \sigma_{11}, \theta, \sigma_{22}) \to (\sqrt{a_1}(\tau + c), \sqrt{a_1/a_2}\beta, a_1\sigma_{11}, \sqrt{a_2}\theta, a_2\sigma_{22})$. *A parameter-space maximal invariant under the induced group is*

$$\mu_1 = \beta/\sqrt{\sigma_{11}/\sigma_{22}}, \mu_2 = \theta/\sqrt{\sigma_{22}/s_{xx}}. \tag{9}$$

The distribution of (T_1, T_2) *depends only on* (μ_1, μ_2).

The proof is straightforward and provided in Appendix A. The usual t-statistics for testing $\beta = 0$ in (1) and $\theta = 0$ in (2), t_1 and t_2, respectively, are simple multiples of $T_1, T_2 : t_1 = \sqrt{n-2}T_1$ and $t_2 = \sqrt{n-1}T_2$. Hence, these are also maximal invariants under the transformations in $\mathbf{K}$.

3 Asymptotic Problem

It is straightforward to show that, under mild conditions, the asymptotic joint distribution of the maximal invariants (t_1, t_2) is given by:

$$\begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \to_d N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, I_2\right). \tag{10}$$

The remainder of the paper will be based on this asymptotic distribution; the exact joint distribution of (t_1, t_2) is tractable, but too complicated to provide a useful basis for inference. Based on this asymptotic distribution, the problem becomes: we observe independent random variables t_1, t_2, with $t_i \sim N(\mu_i, 1), i = 1, 2$, and wish to test the hypothesis $H_0 : \mu_1\mu_2 = 0$. It is clear that this problem is invariant under the group of sign changes $t_i \to -t_i, i = 1, 2$, and under this group of transformations the statistics $f_i = t_i^2, i = 1, 2$, are maximal invariants. These are independent noncentral χ_1^2 variates with noncentrality parameters $\lambda_i = \mu_i^2, i = 1, 2$. The canonical problem thus becomes to test $H_0 : \min\{\lambda_1, \lambda_2\} = 0$ against $H_1 : \lambda_i > 0$ for $i = 1, 2$. This problem is clearly also invariant under the group of permutations of (f_1, f_2), and maximal invariants under this action are (v_1, v_2), with $v_i = f_{(i)}$ the $i - th$ order statistic (so $v_2 \geq v_1 \geq 0$). Thus, we are finally led to focus attention on the pair of order statistics $(v_1, v_2) = (f_{(1)}, f_{(2)})$, which live on the octant $V = \{(v_1, v_2); 0 \leq v_1 \leq v_2 < \infty\}$. The reader should bear in mind, though, that any test (CR) formulated in terms of (v_1, v_2) can equally well be re-expressed in terms of the t-statistics (t_1, t_2).

Remark 1 *Although this asymptotic version of the problem has been derived in the context of the model (1) - (2), more general models may also lead to an asymptotic testing problem of this form. That is, our setting is not as restrictive as it appears, and what follows applies to any testing problem that reduces to the form just described asymptotically.*

Thus, the null hypothesis is composite and involves a nuisance parameter $\lambda = \max\{\lambda_1, \lambda_2\}$, the (possibly non-vanishing) noncentrality parameter. It is therefore not obvious how to construct tests (critical regions) whose NRP $P_w(\lambda)$ does not depend on λ. However, we will show below that for all non-negative integers $r \geq 0$ there is in fact a critical region that properly contains the LR critical region, and has exact size $(r+2)^{-1}$ for all λ. In particular, tests of exact size $\alpha = .01, \alpha = .05$, and $\alpha = .10$ exist, and correspond to the choices $r = 98$, $r = 18$, and $r = 8$ respectively.[4] Trivially, these tests have power functions uniformly above that of

[4]Tests of any size $\alpha \in [0, 1]$ can be constructed by randomization from the exact results given, but we do not discuss randomized tests here.

the LR test. The construction of these exact tests resembles those mentioned by Lehmann (1952), p. 542, and later by Nomakuchi and Sakata (1987), p. 492. See also Berger (1989) for related constructions. As in those earlier examples, however, the critical regions corresponding to these exact tests have some undesirable features. Thus, after presenting these results, we go on to discuss a class of tests which, although not exact, are close to being so, and which avoid some of the undesirable aspects of the exact test.

The noncentral χ^2_κ density for a variate f with noncentrality λ can be expressed in several ways, one of which is as a Poisson mixture:

$$g_\kappa(f;\lambda) = \exp\left\{-\frac{1}{2}\lambda\right\}\sum_{j=0}^{\infty}\frac{(\lambda/2)^j}{j!}g_{\kappa+2j}(f),$$

where $g_\kappa(f) = [2^{\frac{\kappa}{2}}\Gamma(\frac{\kappa}{2})]^{-1}\exp\{-\frac{1}{2}f\}f^{\frac{\kappa}{2}-1}$ denotes the χ^2_κ density function, and we write $g_1(f)$ simply as $g(f)$. The corresponding CDFs are denoted by $G_\kappa(\cdot;\lambda)$, and $G_\kappa(\cdot)$ in the central case, the subscript being omitted when $\kappa = 1$. For $\alpha \in [0,1]$, we define z_α by $G(z_\alpha) = 1-\alpha$.[5]

As already remarked, there is no difficulty finding a test whose NRP is bounded above by a known number - the LR test has that property. But, the NRP and power of the LR test, and other standard tests, can in fact be extremely small - size when the nuisance parameter is small, power when both λ_1, λ_2 are small. The popular Wald test is uniformly very much worse than the LR test in both respects. There is clearly an incentive to seek a test whose NRP is closer to the nominal size for all values of the nuisance parameter, and has better power. This is the motivation for what follows.

3.1 Likelihood Ratio and Wald tests

The likelihood ratio (LR) test is derived by minimising $(t_1-\mu_1)^2+(t_2-\mu_2)^2$ subject to the constraint $\mu_1\mu_2 = 0$. It is straightforward to show that this results in the following critical region in the space of the order statistics (v_1, v_2) : reject H_0 when

$$LR = \min\{f_1, f_2\} = v_1 \tag{11}$$

is large. As usual, the LR test embodies all invariance properties of the testing problem. We denote the critical region for the LR test of nominal size α, i.e., the set $\{z_\alpha < v_1 < v_2, v_2 > z_\alpha\}$, by CR_{LR}.

The Wald test, which is certainly not guaranteed to be invariant, in this case is. This test rejects H_0 when

$$W = \frac{f_1 f_2}{f_1+f_2} = \frac{v_1 v_2}{v_1+v_2} \tag{12}$$

is large, $W > z_\alpha$. The Sobel test (Sobel (1982)) is based on $\sqrt{W}$, and seems to be the test favored by practitioners. It is easy to see that the inequality $W \geq z_\alpha$ implies that $v_1 > z_\alpha$,

[5] The distribution $g(\cdot)$ is not a member of the exponential family, but does have monotone likelihood ratio. The cdf $G_\kappa(z,\lambda)$, for z fixed, is monotone decreasing in both κ and λ when the other is fixed. For these and other properties, see Ghosh (1973). In view of the monotone likelihood ratio property, the Karlin-Rubin Theorem says that, for the problem of testing $\lambda = 0$, the best critical region is of the form $f > z$, and has size α if z is chosen so that $\int_{f>z} g(f)df = \alpha$. This test is unbiased, similar, and has power tending to one as $\lambda \to \infty$. Thus, the problem of testing $H_0 : \lambda = 0$ for a single noncentral χ^2_1 variate is straightforward. Our problem involving a pair of such variates is more complicated because there is a nuisance parameter present under the null.

so the critical region of the Wald test is a subset of that of the LR test. Hence, both the NRP and power of the LR test always exceed that of the W test. We will see shortly that the difference in NRP can be substantial, and that the W test has very poor properties indeed, a fact that has long been appreciated.

3.2 Maximal Invariant Distributions

From equation (6) in Vaughan and Venables (1972)) the joint density of the order statistics for $(v_1, v_2) \in V$ is as given in part (i) of the following theorem, which also gives complete details of the distribution of the order statistics: [6]

Theorem 2 *(i) The joint density of the order statistics on the region* $V = \{(v_1, v_2); 0 \leq v_1 \leq v_2 < \infty\}$ *is given by*

$$pdf(v_1, v_2|\lambda_1, \lambda_2) = [g(v_1; \lambda_1)g(v_2; \lambda_2) + g(v_2; \lambda_1)g(v_1; \lambda_2)]\,. \tag{13}$$

(ii) When either $\lambda_1 = 0$ *or* $\lambda_2 = 0$ *the null density is, for* $(v_1, v_2) \in V$,

$$pdf(v_1, v_2|\lambda) = g(v_1)g(v_2; \lambda) + g(v_2)g(v_1; \lambda); \tag{14}$$

Here λ *is the non-zero member of the pair* (λ_1, λ_2), *or zero when both vanish.*

(iii) The marginal density of the smaller order statistic v_1, *i.e., the LR statistic, is given by*

$$pdf(v_1|\lambda_1, \lambda_2) = g(v_1; \lambda_1)\,[1 - G(v_1; \lambda_2)] + g(v_1; \lambda_2)\,[1 - G(v_1; \lambda_1)]\,, v_1 \geq 0, \tag{15}$$

with corresponding CDF

$$H(v_1; \lambda_1, \lambda_2) = G(v_1; \lambda_1) + G(v_1; \lambda_2) - G(v_1; \lambda_1)G(v_1; \lambda_2). \tag{16}$$

In the null case, when one non-centrality parameter vanishes, these reduce to

$$pdf(v_1|\lambda) = g(v_1)[1 - G(v_1; \lambda)] + g(v_1; \lambda)[1 - G(v_1)], \tag{17}$$

and

$$H(v; \lambda) = G(v) + G(v; \lambda) - G(v)G(v; \lambda), \tag{18}$$

respectively. Finally,

$$1 - H(v; 0) = 1 - 2G(v) + G(v)^2 = (1 - G(v))^2. \tag{19}$$

Proof. Part (i) is direct from Vaughan and Venables (1972). Part (iii) is simply the fact that, on integrating over $v_1 < v_2 < \infty$, we have

$$pdf(v_1|\lambda_1, \lambda_2) = g(v_1; \lambda_1) \int_{v>v_1} g(v; \lambda_2)dv + g(v_1; \lambda_2) \int_{v>v_1} g(v; \lambda_1)dv. \tag{20}$$

The results for the null case are specializations of this. It is trivial to check that the derivatives of $H(v_1; \lambda_1, \lambda_2)$ and $H(v; \lambda)$ yield the densities given in (15) and (17). ■

[6] Since the order statistics are maximal invariants under the action of the symmetric group S_2 on (f_1, f_2), the main result in part (i) can also be obtained by invoking Stein's method of obtaining the density of the maximal invariant by averaging the joint density over the group.

Remark 2 *Observe that*

$$(\partial/\partial v_1)[G(v_1;\lambda_1)G(v_1;\lambda_2)] = [g(v_1;\lambda_1)G(v_1;\lambda_2) + g(v_1;\lambda_2)G(v_1;\lambda_1)]. \tag{21}$$

This and similar identities are useful in what follows. For example, to verify that the joint density integrates to one:

$$\begin{aligned}\int_{0<v_1<v_2}\int_{v_2>0} pdf(v_1,v_2|\lambda_1,\lambda_2)dv_2dv_1 &= \int_{v_2>0}[G(v_2;\lambda_1)g(v_2;\lambda_2)+G(v_2;\lambda_2)g(v_2;\lambda_1)]\,dv_2\\ &= \int_{v_2>0}(\partial/\partial v_2)[G(v_2;\lambda_1)G(v_2;\lambda_2)]dv_2\\ &= [G(v_2;\lambda_1)G(v_2;\lambda_2)]_0^\infty = 1.\end{aligned}$$

3.3 Properties of the LR test

Theorem 2 provides a very direct description of the properties of the LR test:

Corollary 1 *The NRP* $P_{CR_{LR}}(\lambda)$ *of the LR test is given by*

$$P_{CR_{LR}}(\lambda) = \Pr\{v_1 > z_\alpha|\lambda\} = \alpha[1-G(z_\alpha;\lambda)], \tag{22}$$

and its power function by

$$P_{CR_{LR}}(\lambda_1,\lambda_2) = [1-G(z_\alpha;\lambda_1)][1-G(z_\alpha;\lambda_2)] = \Pr\{\chi_1^2(\lambda_1)>z_\alpha\}\Pr\{\chi_1^2(\lambda_2)>z_\alpha\}. \tag{23}$$

Note that the power of the LR test, $P_{CR_{LR}}(\lambda_1,\lambda_2)$, will always be greater than its *NRP*, $P_{CR_{LR}}(\lambda)$, since, for given (λ_1,λ_2),

$$\begin{aligned}\Pr\{\chi_1^2(\lambda_1) &> z_\alpha\}\Pr\{\chi_1^2(\lambda_2)>z_\alpha\} > \min\left[\alpha\Pr\{\chi_1^2(\lambda_1)>z_\alpha\},\alpha\Pr\{\chi_1^2(\lambda_2)>z_\alpha\}\right]\\ &= P_{CR_{LR}}(\lambda).\end{aligned} \tag{24}$$

Next, for any fixed $z>0$, $1-G(z;\lambda)$ is an increasing function of λ, tending to one as $\lambda\to\infty$. Therefore,

Corollary 2 *For the LR test of nominal size* α, *and all* $\lambda \geq 0$,

$$\alpha^2 \leq P_{CR_{LR}}(\lambda) \leq \alpha. \tag{25}$$

Thus, the LR test has size α, but, for small λ the NRP of the LR test will be near α^2, hence very near zero, and only approaches the nominal size α as $\lambda\to\infty$. The analogous results for the Wald (Sobel) test are given in the next subsection.

The tests we consider below are constructed by augmenting the LR critical region, and it is clear from the expression for $P_{CR_{LR}}(\lambda)$ above that the region added should have null content either exactly equal to $\alpha G(z_\alpha;\lambda)$ for all λ, rendering the test exact, or have this property approximately. Both exact and approximate augmented LR tests will be constructed below.

3.4 Properties of the Wald Test

The CDF of the Wald statistic W is given in the next Proposition:

Proposition 1 *The CDF of the Wald statistic W is given by*

$$\begin{aligned}\Pr\{W &\leq z;\lambda_1,\lambda_2\} = G(2z;\lambda_1)G(2z;\lambda_2)\\ &+\int_{v_2>2z}\left[g(v_2;\lambda_2)G\left(\frac{zv_2}{v_2-z};\lambda_1\right)+g(v_2;\lambda_1)G\left(\frac{zv_2}{v_2-z};\lambda_2\right)\right]dv_2.\end{aligned} \tag{26}$$

The NRP of the Wald test with critical value z_α, $P_W(\lambda)$, is therefore:[7]

$$\begin{aligned}P_W(\lambda) &= 1-G(2z_\alpha)G(2z_\alpha;\lambda)\\ &-\int_{v_2>2z_\alpha}\left[g(v_2)G\left(\frac{z_\alpha v_2}{v_2-z_\alpha};\lambda\right)+g(v_2;\lambda)G\left(\frac{z_\alpha v_2}{v_2-z_\alpha}\right)\right]dv_2.\end{aligned} \tag{27}$$

A simple application of integration by parts applied to the null density shows that, as $\lambda\to\infty$, $P_W(\lambda)\to\alpha$. But, as remarked earlier, $P_W(\lambda)$ can be very much smaller than $P_{CR_{LR}}(\lambda)$ when λ is small. For instance, for λ near zero the NRP of the Wald test at nominal size .05 is near .00009, while that of the LR test is near $(.05)^2=.0025$. The two power functions behave similarly. Since it is uniformly inferior to the LR test we will not discuss the Wald test further.

3.5 Partition of the sample space

It is convenient at this point to partition the region V into three disjoint regions determined by a scalar $z>0$:

$$A_1 = \{v_2>z, z<v_1<v_2\}, \tag{28}$$
$$A_2 = \{v_2>z, 0<v_1<z\}, \tag{29}$$
$$A_3 = \{v_2<z, 0<v_1<v_2\}. \tag{30}$$

The first of these, A_1, is the level-α CR_{LR} when $z=z_\alpha$. It is not difficult to obtain the following under the null: for any fixed $z>0$,

$$P_{A_1}(\lambda) = [1-G(z)][1-G(z;\lambda)], \tag{31}$$
$$P_{A_2}(\lambda) = G(z)[1-G(z;\lambda)]+G(z;\lambda)[1-G(z)], \tag{32}$$
$$P_{A_3}(\lambda) = G(z)G(z;\lambda), \tag{33}$$

so that

$$P_{A_2\cup A_3}(\lambda) = G(z)+G(z;\lambda)-G(z)G(z;\lambda). \tag{34}$$

In particular, for the choice $z=z_\alpha$, $P_{A_1}(\lambda)=\alpha[1-G(z_\alpha;\lambda)]$, as we have just seen, $P_{A_2}(\lambda)=1-\alpha-(1-2\alpha)G(z_a;\lambda)$, and $P_{A_3}(\lambda)=(1-\alpha)G(z_\alpha;\lambda)$, so that $P_{A_2\cup A_3}(\lambda)=1-\alpha+\alpha G(z_\alpha;\lambda)$. Here and in what follows the regions A_1,A_2,A_3 will be assumed to be defined by $z=z_\alpha$. It follows that:

Proposition 2 *(i) $P_{A_1\cup A_3}(\lambda)=\alpha+(1-2\alpha)G(z_\alpha;\lambda)$, which varies with λ unless $\alpha=.5$. (ii) For $\alpha=.5$, the region $A_1\cup A_3$ has size α for all λ.*

Remark 3 *Part (ii) shows that there does exist an exact test of size $\alpha=.5$, namely, the test with CR $A_1\cup A_3$. We will generalize this property shortly, and show that exact tests of size $\alpha=(r+2)^{-1}$ exist for all integers $r\geq 0$. The case just mentioned is the case $r=0$.*

[7] We let W denote the Wald statistic itself, and the critical region it defines: $W>z_\alpha$.

4 An exact test

The poor size and power properties of the classical tests motivates the search for more satisfactory tests. Specifically, we would hope to be able to construct tests whose size is exact, or nearly so, for all λ, and whose power improves on that of the LR test, in particular. In this section we shall show that an exact test does indeed exist for certain choices of α, and is easily constructed. We confine attention to tests whose critical regions properly contain that of the LR test. That is, if H_0 is rejected by the LR test it must also be rejected by the new test (but not vice versa).

As we have already observed, $P_{A_1 \cup A_3}(\lambda) = \alpha + (1-2\alpha)G(z_\alpha;\lambda)$, so that there is an exact test of size $\alpha = 1/2$. That is, there is a choice of α, $\alpha = .5$, such that $P_{A_1 \cup A_3}(\lambda) = \alpha$ for all $\lambda \geq 0$. We shall show that for each $\alpha = (r+2)^{-1}$, with r a non-negative integer, there is likewise an exact test of size α. The simple case $A_1 \cup A_3$ is the case $r = 0, \alpha = 1/2$.

To illustrate the general result, consider first choosing a single value $z_1 < z_\alpha$, and using this to define the two triangular regions, disjoint subsets of A_3, $A_{30} = \{0 < v_1 < v_2, 0 < v_2 < z_1\}$ and $A_{31} = \{z_1 < v_1 < v_2, z_1 < v_2 < z_\alpha\}$. These have null combined probability content

$$(1-\alpha-G(z_1))G(z_\alpha;\lambda) - (1-\alpha-2G(z_1))G(z_1;\lambda),$$

which differs from the target value $\alpha G(z_\alpha;\lambda)$ by

$$(1-2\alpha-G(z_1))G(z_\alpha;\lambda) - (1-\alpha-2G(z_1))G(z_1;\lambda). \tag{35}$$

Taking $z_2 = z_\alpha$, so that $\alpha = 1-G(z_2)$, we can choose the pair (z_1, z_2) so that the coefficients of the two non-central distribution functions both vanish, yielding a test of size $1-G(z_2)$ for all λ. This requirement produces two linear equations, $2G(z_2)-G(z_1) = 1$, and $G(z_2)-2G(z_1) = 0$, with unique solution $G(z_1) = 1/3, G(z_2) = 2/3$, so that $\alpha = 1-G(z_2) = 1/3$. This is the case $r = 1$, $\alpha = 1/3$.

Generalizing this construction, we may prove:

Theorem 3 *For each integer $r \geq 0$ there exist unique numbers $z_1 < z_2 < ... < z_r < z_\alpha$ such that the critical region $CR_{LR} \cup w_r(z)$, where $w_r(z) = A_3 \backslash A_r(z)$, with*

$$A_r(z) = A_r(z_1, ..., z_r) = \bigcup_{i=1}^{r} \{0 < v_1 < z_i, z_i < v_2 < z_{i+1}\}, \tag{36}$$

where $z_{r+1} = z_\alpha$, has null rejection probability $\alpha = (r+2)^{-1}$ for all $\lambda \geq 0$. These numbers z_i are the solutions to the identities

$$G(z_i) = \frac{i}{r+2}, i = 1, ..., r+1. \tag{37}$$

In particular, $z_{r+1} = z_\alpha$, so that $\alpha = (r+2)^{-1}$.

When the z_i are chosen in this optimal fashion we denote the augmenting region simply by w_r.[8] For the common case $\alpha = .05$ we require $r = 18$ points to construct a region that has content $\alpha = .05$ for all λ, and similar constructions are available for $\alpha = .1$ ($r = 8$) and $\alpha = .01$ ($r = 98$).

[8]We believe that, within the class of tests whose critical regions consist of CR_{LR}, augmented by a subset of A_3 bounded below by a weakly increasing function $v_1 = h(v_2)$, the tests described above are the *only* exact tests. A proof of this is lacking at the time of writing.

Example 1 *In the case* $\alpha = .05$, $r = 18$, *the critical region is defined by* $z_{.05} = 3.841$, *together with the following 18 values* z_i :

$i =$	1	2	3	4	5	6	7	8	9
z_i	.004	.016	.036	.064	.101	.148	.206	.275	.357
$i =$	10	11	12	13	14	15	16	17	18
z_i	.455	.571	.708	.873	1.074	1.323	1.6424	2.0722	2.7055

The augmenting critical region w_r *is shown in blue in Figure 1 for the case* $\alpha = .05$ $(r = 18)$; CR_{LR} *is the green region. The red line will be explained shortly.*

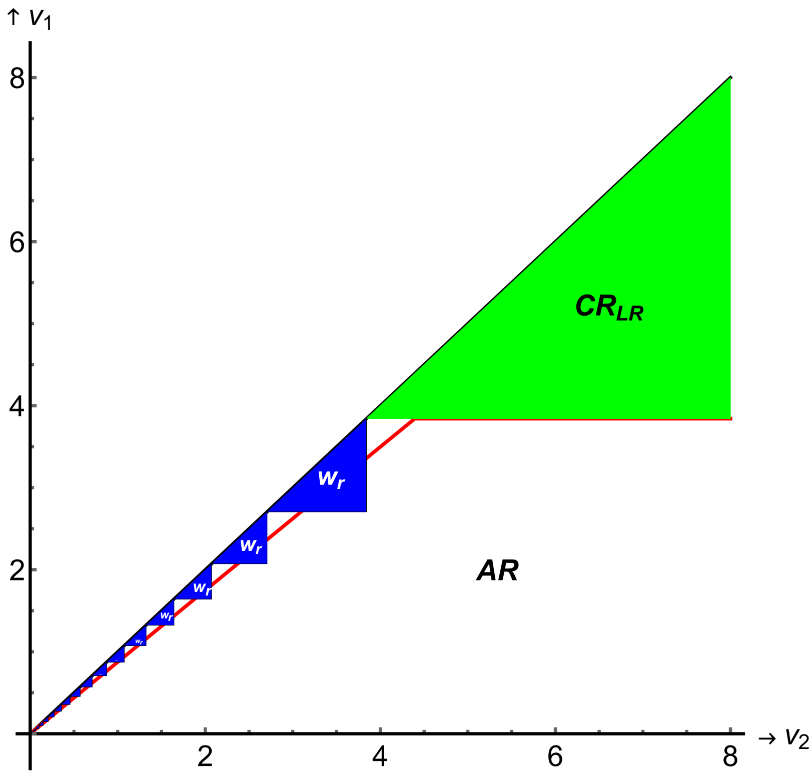

Figure 1: CR_{LR} (green) and augmenting region w_r (blue) for the exact test; $\alpha = .05$ (r=18).

Remark 4 *This construction of exact regions (tests) obviously only works for* $\alpha \leq \frac{1}{2}$. *Using a different constructive proof, Van Garderen and Van Giersbergen (2021) show that, within a class of test with weakly increasing cadlag boundary, a similar test exists if and only* $1/\alpha \in \mathbb{N}$ *including* $\alpha = 1$ *and trivially* $\alpha = 0$. *Their test with* $\alpha = 1/(r+2)$ *is essentially the same as the exact test here.*

4.1 Power gain

The power function of the exact test described above is obviously above that of the LR test for all (λ_1, λ_2), whatever the value of r. It is easy to check that the content of the region A_3 under the alternative is $G(z_\alpha; \lambda_1)G(z_\alpha; \lambda_2)$. The power added by the augmenting region is therefore given by

$$\begin{aligned}
P_{w_r}(\lambda_1,\lambda_2) &= G(z_\alpha;\lambda_1)G(z_\alpha;\lambda_2) \\
&\quad -\sum_{i=1}^{r}\int_{0<v_1<z_i}\int_{z_i<v_2<z_{i+1}}[g(v_1;\lambda_1)g(v_2;\lambda_2)+g(v_2;\lambda_1)g(v_1;\lambda_2)]\,dv_1dv_2 \\
&= G(z_\alpha;\lambda_1)G(z_\alpha;\lambda_2)-\sum_{i=1}^{r}G(z_i;\lambda_1)[G(z_{i+1};\lambda_2)-G(z_i;\lambda_2)] \\
&\quad -\sum_{i=1}^{r}G(z_i;\lambda_2)[G(z_{i+1};\lambda_1)-G(z_i;\lambda_1)]. \qquad (38)
\end{aligned}$$

This is naturally symmetric in (λ_1, λ_2), and vanishes in the limit as either noncentrality parameter goes to infinity. That is, there is no power gain over the LR test in the limit, but there certainly is for (λ_1, λ_2) near the origin. The NRP gain at the origin is obviously $\alpha(1-\alpha) = (r+1)\alpha^2$. The power function behaves similarly for points (λ_1, λ_2) close to the origin with a power gain around .0475 over the LR test when $\alpha = .05$.

4.2 Pros and Cons

For a restricted, but relevant, range of nominal sizes (α of the form $(r+2)^{-1}$) the construction described above provides, for the first time, a non-randomized exact test of the no-mediation hypothesis. And, whilst the augmenting critical region does contain points close to the origin, which might be considered counter-intuitive, over 90% of the area of the augmenting critical region in the case of $\alpha = .05$ is accounted for by the four largest triangular regions, and these regions are well away from the origin. Nevertheless, the critical region of the exact test does have several undesirable properties. First, the region $CR_{LR} \cup w_r = CR_r$, say, is not monotone in (v_1, v_2). That is, $(v_1, v_2) \in CR_r$ does not imply that $(v_1', v_2') \in CR_r$ when $v_1' \geq v_1$ and $v_2' \geq v_2$. Similarly, the acceptance region for the test is not convex, which is somewhat counter-intuitive. Also, unlike the LR test itself, the exact test does not possess an important coherence property, namely, that rejection at level α does not imply rejection at every level smaller than α. That is, as the reader may easily confirm, the critical region for $\alpha = (r+2)^{-1}$ is not a subset of that for $\alpha = (r+1)^{-1}$. The simply-augmented LR test introduced in the next section will remedy some of these deficiencies.

To motivate the class of tests we consider next, observe that one could approximate the exact augmenting region (i.e., the blue triangles in Figure 1) with the region above a line $v_1 = bv_2$, for some suitable choice of b. For instance, the (geometric) area of the augmenting region of the exact test in the case $r = 18$ ($\alpha = .05$) is 1.08403, which is equal to the area above the line $v_1 = (.87187)v_2$. This is the red line in Figure 1. We denote this value of b by b_r. Unlike the exact CR, the NRP of the region defined in this way does vary slightly with λ (see Figure 4 below). However, this suggests that an augmenting region consisting of the the region above a line $v_1 = bv_2$ but outside CR_{LR} might produce a test that is almost exact, but without the negative properties of the exact CR. [9] Obviously the acceptance region corresponding to such a CR will be convex by construction. We consider the general class of such tests next.

[9] The paper by Van Garderen and Van Giersbergen (2021) discusses (in present notation) more general augmenting regions close to the $v_1 = v_2$ line, but bounded below by a piecewise linear spline. The exact test just described is of this form, but a very special case: the alternate knots are constrained to lie on the line $v_1 = v_2$, and the linear components are constrained to be alternately horizontal and vertical.

5 Simpler augmented LR tests

5.1 Simply-augmented LR tests

In view of these comments, we now consider tests defined by the following simple augmentation of CR_{LR} : the region w_b bounded below by the line $v_1 = bv_2$, with $0 \leq b \leq 1$, and bounded above by $\min\{v_2, z_\alpha\}$. Since the line $v_1 = bv_2$ meets the line $v_1 = z_\alpha$ at $v_2 = z_\alpha/b$, w_b has the following form:

$$w_b = \{(v_1, v_2) : (bv_2 < v_1 < v_2, 0 < v_2 < z_\alpha) \cup (bv_2 < v_1 < z_\alpha, z_\alpha < v_2 < z_\alpha/b)\}, \quad (39)$$

for $0 < b \leq 1$, and $w_b = A_2 \cup A_3$ for $b = 0$. For a fixed value of $b \in (0, 1]$ the nominal level-α test therefore has the simple form: reject H_0 if either $v_1 \geq z_\alpha$, or $v_1 < z_\alpha$ and $v_1/v_2 > b$. Note that w_b is empty when $b = 1$, so the critical region is CR_{LR} in that case. For $0 < b < 1$ w_b is part of the region $A_2 \cup A_3$, and does not intersect CR_{LR} (see Figure 2). We denote the critical region $CR_{LR} \cup w_b$ by CR_b, call tests of this form the class of *simply-augmented LR tests,* and call a test with given value of b a $LR(b)$ test. The LR test itself is thus the $LR(1)$ test. Some properties of this class of tests are given next; the proofs are in Appendix A.

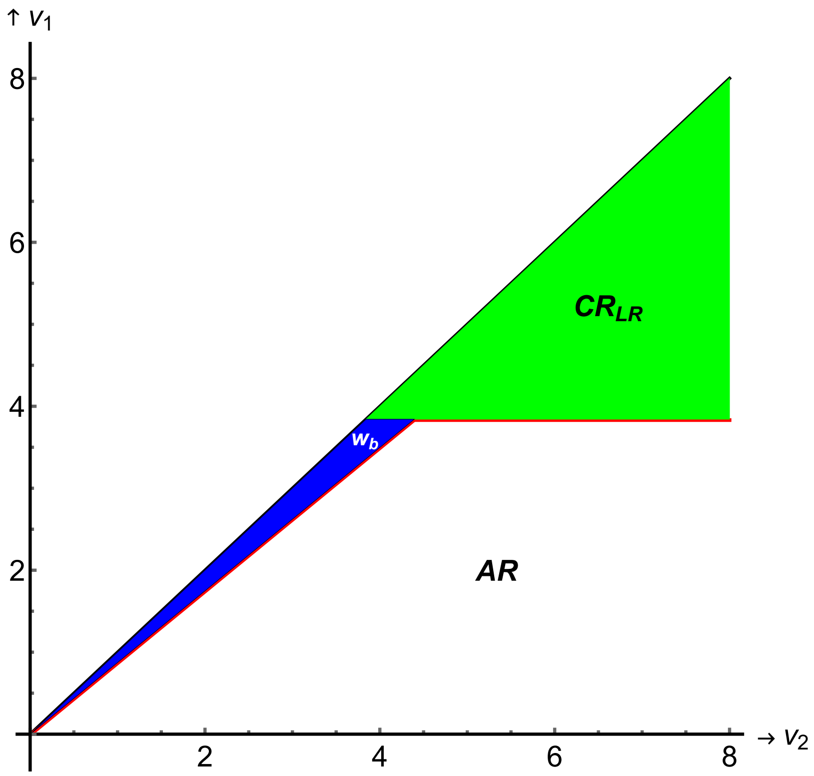

Figure 2: $LR(b)$ A simply augmented LR test for $\alpha = 0.05$.

Proposition 3 *The NRP of the simply-augmented* $LR(b)$ *test with parameter* $b < 1$ *is given by* [10]

$$P_{CR_b}(\lambda) = \alpha + \int_{0<v<z_\alpha} [g(v)G(v/b; \lambda) + g(v; \lambda)G(v/b)]dv - G(z_\alpha; \lambda). \quad (40)$$

[10]The following properties of the NRP are easily seen from this Proposition: (i) $P_{CR_1}(\lambda) = \alpha - \alpha G(z; \lambda)$ (the NRP for the LR test); (ii) $P_{CR_b}(\lambda) \to 1$ as $b \to 0$ for all λ; (iii) $P_{CR_1}(0) = \alpha^2$, and (iv) $P_{CR_1}(\lambda) \to \alpha$ as $\lambda \to \infty$.

When $\lambda = 0$,

$$P_{CR_b}(0) = 2\alpha + 2\int_{0<v<z_\alpha} g(v)G(v/b)dv - 1. \tag{41}$$

Let us next define the discrepancy function

$$D_\alpha(b,\lambda) = A_\alpha(b;\lambda) - G(z_a;\lambda), \tag{42}$$

where

$$A_\alpha(b;\lambda) = \int_{0<v<z_\alpha} [g(v;\lambda)G(v/b) + g(v)G(v/b;\lambda)]dv. \tag{43}$$

When $D_\alpha(b,\lambda) < 0$ the test has NRP $< \alpha$, and vice versa. The following result says that there is no member of this class of tests that has NRP equal to α (i.e., $D_\alpha(b,\lambda) = 0$) for all λ :

Proposition 4 *There is no value of* b *for which* $D_\alpha(b,\lambda) = 0$ *for all* $\lambda \geq 0$.

Now $D_\alpha(b,\lambda)$ is obviously continuous in b on the interval $(0,1]$, and is (strictly) monotonic decreasing in b, with $D_\alpha(0,\lambda) = 1-\alpha > 0$ and $D_\alpha(1,\lambda) = -\alpha G(z_\alpha;\lambda) < 0$. The following result is therefore clear:

Proposition 5 *For each finite* λ *there is a unique* $b(\lambda) \in (0,1]$ *for which* $D_\alpha(b,\lambda) = 0$. *At this point* $D_\alpha(b,\lambda)$ *changes sign, from positive to negative.*

This result says that the equation $D_\alpha(b,\lambda) = 0$ implicitly defines a function $b(\lambda) : \mathbb{R}^+ \longmapsto [0,1]$. Table 1 illustrates the behaviour of $b(\lambda)$ for a few values of λ when $\alpha = .05$.

$\lambda \rightarrow$	0	.1	.5	1	2	5	20
$b(\lambda)$	.8588	.8599	.8634	.8666	.8707	.8743	.8685

Table 1: Values of $b(\lambda)$ for various values of $\lambda; \alpha = .05$.

Next, the LR test of nominal size α has the property that its NRP$\rightarrow \alpha$ as $\lambda \rightarrow \infty$ (since $G(z_\alpha;\lambda) \rightarrow 0$ as $\lambda \rightarrow \infty$). This property is shared by all members of the class of $LR(b)$ tests:

Proposition 6 *For any fixed* $b \in (0,1]$, *and any* α,

$$\lim_{\lambda\rightarrow\infty} D_\alpha(b,\lambda) = 0. \tag{44}$$

That is, for any $b \in (0,1]$,

$$\lim_{\lambda\rightarrow\infty} P_{CR_b}(\lambda) = \alpha. \tag{45}$$

Note that, since $P_{CR_{LR}}(\lambda) \rightarrow \alpha$ as $\lambda \rightarrow \infty$, this result implies that, for any $b \in (0,1]$, $P_{w_b}(\lambda) \rightarrow 0$, as $\lambda \rightarrow \infty$. The result in Proposition 6 can be seen in Figure 4 below.

It remains to decide which member of the class of simply-augmented LR tests to recommend, i.e., which choice of b is best. In the next two subsections we consider two possible choices, each of which has merit.

5.2 Best simply-augmented LR test of size α

Proposition 6 says that all members of the class of $LR(b)$ tests have the correct limiting NRP as $\lambda \to \infty$, but *not* that the NRP is less than α for all λ. We have already noted that the LR test $v_1 > z_\alpha$ has the correct size α, i.e., that $P_{CR_{LR}}(\lambda) \leq \alpha$ for all $\lambda \geq 0$. This raises the question: do there exist values $b < 1$ with the property that $D_\alpha(b, \lambda) < 0$ for all λ? In fact there are such values, and there is a smallest value of b, b_u say, satisfying $D_\alpha(b, \lambda) \leq 0$ for all $\lambda \geq 0$. This value cannot be computed directly, but is easily located numerically. Since b_u depends on α we write it as $b_u(\alpha)$. We may then state:

Theorem 4 *For given α, all members of the class of simply-augmented LR tests with $b_u(\alpha) \leq b \leq 1$ have size α. Of these, the test with $b = b_u(\alpha)$ has maximum power.*

Figure 6 in Section 6.3 is a graph of the values $b_u(\alpha)$ as a function of α, and in Table 4 in Appendix B we provide the values $b_u(\alpha)$ for a fine grid of values of $\alpha \in [0, 1]$. The table also gives the chi-squared critical values, z_α, needed to implement the test. Thus, the table allows the implementation of the optimal test in this class. For the commonly used test sizes $\alpha = .01, .05, .1, .2$, the values required are given in the third and fourth columns of Table 2 below.

α	$b_0(\alpha)$	$b_u(\alpha)$	z_α
.01	0.9693	0.9697	6.6349
.05	0.8588	0.8744	3.8415
.10	0.7408	0.8202	2.7055
.20	0.5503	0.7419	1.6424

Table 2: The values $b_0(\alpha)$, $b_u(\alpha)$, and z_α for tests of size $\alpha = .01, .05, .10, .20$.

For λ near zero the $LR(b_u)$ test has NRP well above that of the LR test itself. For instance, when $\lambda = 0$ we have, for a test of nominal level $\alpha = .05$, the value $b_u(\alpha) = .8744$, and $P_{CR_{b_u(\alpha)}}(0) = .0444$, whereas the LR test itself has NRP .0025 at $\lambda = 0$. Figure 3 shows the improved NRP behaviour (as a function of λ) of the $LR(b_u)$ test for the cases $\alpha = .05$ and $\alpha = .10$.

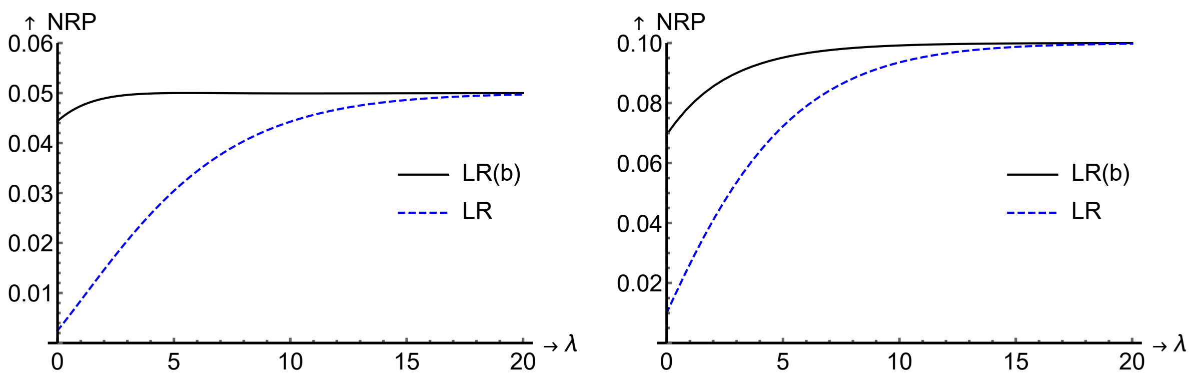

Figure 3: NRP of LR test and the optimal simply-augmented test; $\alpha = .05$ and .10

5.3 Slightly over-sized tests

The optimal test of size α just discussed has the correct size for all λ. However, since our initial objective was to improve the behaviour of the NRP near the origin, it is worth considering the test whose NRP is correct at the origin, i.e, the test based on the choice $b = b(0)$. Such a choice for b, which we denote below by $b_0(\alpha)$, produces a test whose NRP is correct at $\lambda = 0$, since $D_\alpha(b_0(\alpha), 0) = 0$ by definition, and also as $\lambda \to \infty$, but which can have NRP slightly above the nominal level for intermediate values of λ. Values of $b_0(\alpha)$ are given in the second column of Table 2. As for b_u, the values of $b_0(\alpha)$ need to be located numerically. The actual sizes of the $LR(b_0)$ test for nominal sizes .01,.05,.1, and .2 are .0101,.0536,.1095, and .2185, respectively, so the excess size is minimal.

Since our aim was to find a test with improved behaviour near the origin, and this can be achieved at very little cost in terms of size, the $LR(b_0)$ can be entertained as an alternative to the optimal $LR(b_u)$ test. And, since typically $b_0(\alpha) < b_u(\alpha)$, the power of the $LR(b_0)$ test is guaranteed to exceed that of the $LR(b_u)$ test for all (λ_1, λ_2). Figure 4 below shows the discrepancies $D_\alpha(b_u(\alpha); \lambda)$ (blue), $D_\alpha(b_r(\alpha); \lambda)$ (red), and $D_\alpha(b_0(\alpha); \lambda)$ (green) as functions of λ when $\alpha = .05$. In this case $D_\alpha(b_0(\alpha); \lambda)$ is always non-negative, and $D_\alpha(b_u(\alpha); \lambda)$ is always non-positive. Similar behaviour is exhibited for other values of α, except the case $\alpha = .01$, where $D_\alpha(b_0(\alpha); \lambda)$ can be very slightly positive or negative, depending on λ. However, the NRP of the test $LR(b_0)$ is never greater than .0101, and is correct at $\lambda = 0$. In this case, though, b_0 and b_u are very similar (.969338 and .9697, respectively), so the tests are essentially the same.

Notwithstanding that the $LR(b(0))$ test is an option, the properties discussed in the next section are for the $LR(b_u)$ test alone. The properties of the $LR(b(0))$ test are similar.

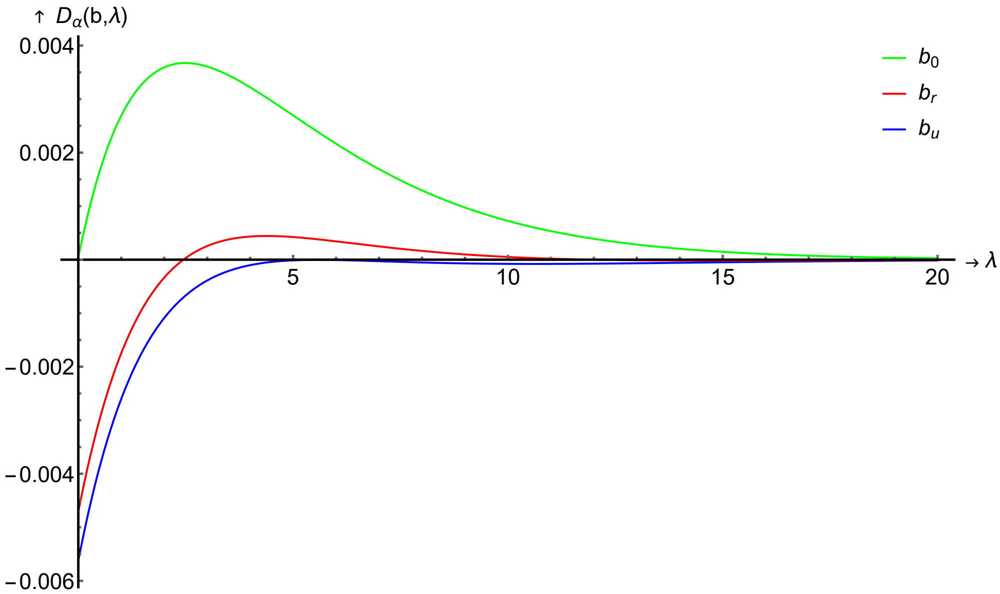

Figure 4: Discrepancy as a function of λ for the case $\alpha = .05$: $D_\alpha(b_0; \lambda)$ in green, $D_\alpha(b_u; \lambda)$ in blue, and $D_\alpha(b_r; \lambda)$ in red.

6 Power, coherence, p-values, and an example

6.1 Power

It is trivially true that the power of the $LR(b_u)$ test cannot be less than that of the LR test. The power function of the augmented LR test can be calculated in exactly the same way as we have done for the size. The result is, after simplifying:

$$\begin{aligned} P(CR_{b_u}|\lambda_1,\lambda_2) &= [1-G(z_\alpha;\lambda_1)][1-G(z_\alpha;\lambda_2)] - G(z_\alpha;\lambda_1)G(z_\alpha;\lambda_2) \\ &+ \int_{0<v<z_\alpha} [g(v;\lambda_1)G(v/b_u;\lambda_2) + g(v;\lambda_2)G(v/b_u;\lambda_1)]\, dv. \end{aligned} \tag{46}$$

the first term being the power of the LR test. The power function is obviously symmetric in (λ_1, λ_2). And again, as either noncentrality parameter goes to infinity the power of the augmented test approaches that of the LR test. The power function, together with that of the LR test itself (in brackets), is given in Table 3 for a selection of values of (λ_1, λ_2). The table is, of course, symmetric.

$\lambda_1\backslash\lambda_2$	.1	.5	1	2	5	20
.1	.0454 (.0038)					
.5	.0475 (.0067)	.0528 (.0119)				
1	.0498 (.0105)	.0588 (.0185)	.0694 (.0289)			
2	.0530 (.0180)	.0690 (.0319)	.0882 (.0498)	.1240 (.0858)		
5	.0578 (.0375)	.0893 (.0663)	.1287 (.1035)	.2052 (.1784)	.3917 (.3707)	
20	.0615 (.0612)	.1087 (.1083)	.1696 (.1690)	.2918 (.2913	.6057 (.6052)	.988 (.988)

Table 3: Power of $LR(b)$ and LR tests; $\alpha = .05, b = b_u(.05) = .8744$.

It is clear that, for (λ_1, λ_2) near the origin, the LR test has poor power, and that the simply-augmented LR test of size .05 improves considerably upon it. In Appendix B we display the power difference $P_{CR_{b_u}}(\lambda_1, \lambda_2) - P_{CR_{LR}}(\lambda_1, \lambda_2)$ for values of the $\lambda_i \in [0, 10]$. It is evident that the power difference is quite small for large λ_i, but substantial for (λ_1, λ_2) near the origin. It can be seen in both Table 3, and the power surfaces, that as either noncentrality parameter increases the two converge, and for large departures from the null they are identical.

6.2 Coherence

The likelihood ratio critical region has the desirable property that, if an observed point (v_1, v_2) falls in the rejection region at level α, it also falls in the rejection region at every level smaller than α. That is, the CR at a given level properly contains that at any smaller level. The LR test is thus coherent for inference on H_0. An important property of the augmented LR test as we have constructed it is that *it retains this coherency property*. This is perhaps best illustrated graphically. Figure 5 shows the respective critical regions for the levels $\alpha = .01, .05$, and $.1$. It is clear that the proposed approach provides coherent inference on H_0 in this sense.

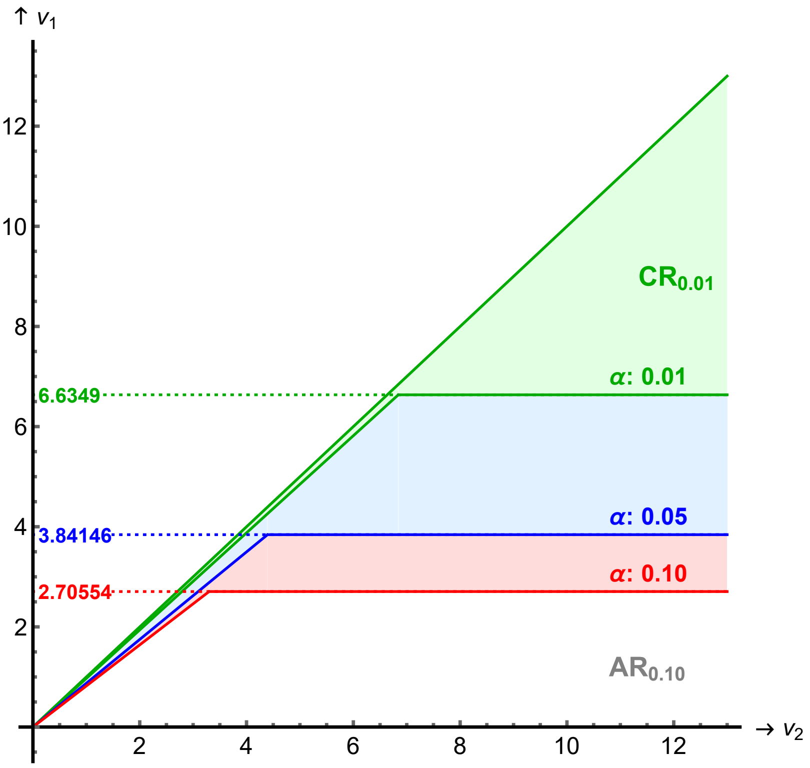

Figure 5: Augmented LR critical regions for various $\alpha = 0.01,\ 0.05,\ 0.10$

6.3 p-values

The coherence property just mentioned suggests that we can define a p-value for any observed point (v_1, v_2) by reference to the critical regions $CR_{b_u(\alpha)}$. To do so, we simply locate the value of α, say α_0, for which the observed point lies on the boundary of the critical region $CR_{b_u(\alpha_0)}$. All points in the region $CR_{b_u(\alpha_0)}$ lie in critical regions at levels smaller than α_0, and in this sense are "more extreme" than the observed point under the null hypothesis. The value α_0 then has a natural interpretation as the p-value for the observed point (v_1, v_2).

To define α_0 explicitly we make three observations: First, since $b_u(\alpha)$ is monotonic it has an inverse, so for each $b \in [0, 1]$ the value α_b satisfying $b = b_u(\alpha_b)$ is well-defined. The value α_b can be (approximately) located by using Table 4 in the Appendix, or the graph of $b_u(\alpha)$ in Figure 6. Next, observe that all points in the region $V = \{(v_1, v_2); 0 \leq v_1 \leq v_2 < \infty\}$ can be associated with two coordinates, $v_1 \geq 0$, and $b = v_1/v_2 \in [0, 1]$.[11] And, each $v_1 \geq 0$

[11] The coordinate b identifies the "direction" of the point (v_1, v_2), and v_1 its distance from the origin.

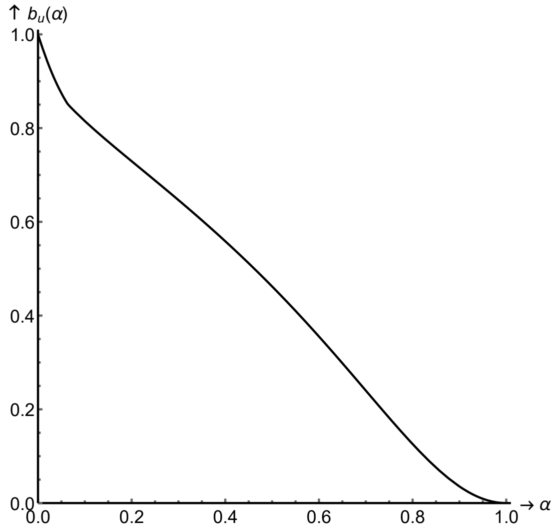

Figure 6: $b_u(\alpha)$ as function of α

yields a value $G(v_1)$, hence a value $\alpha_1 = 1 - G(v_1) \in [0, 1]$. Finally, every point $(v_1, v_2) \in V$ lies on either the horizontal part of the boundary of some $CR_{b_u(\alpha)}$, or on the sloping part.

If $b < b_u(\alpha_1)$, then (v_1, v_2) lies on the horizonal part of the boundary of $CR_{b_u(\alpha_1)}$, and $\alpha_0 = \alpha_1$. That is because $b < b_u(\alpha_1)$ means that $v_2 > v_1/b_u(\alpha_1)$, which defines the horizontal part of the boundary of $CR_{b_u(\alpha_1)}$. On the other hand, if $b > b_u(\alpha_1)$, (v_1, v_2) must lie on the sloping part of the boundary of some region $CR_{b_u(\alpha)}$, and the value of α for which this holds is defined by $b = b_u(\alpha)$, i.e., $\alpha = \alpha_b$. Hence, we just need to calculate α_b from either the table or the graph of $b_u(\alpha)$, and the p-value is $\alpha_0 = \alpha_b$.

6.4 An application

To illustrate our best simply-augmented LR test, data from Hayes (2017, Section 4.2) is considered; this data set is called ESTRESS and can be downloaded from www.afhayes.com. The study involves entrepreneurs who were members of a networking group for small business owners; see Pollack et al. (2012). They answered an online survey about the recent performance of their business and also about their emotional and cognitive reactions to the economic climate. Hence, Y, X and M denote disengagement from entrepreneurial activities ($withdraw$), economic stress ($estress$) and depressed affect ($affect$) respectively. There are also three confounding variables C_1, C_2 and C_3 that are related to entrepreneurial self-efficacy (ese), gender (sex) and length of time in the business ($tenure$); see Figure 4 of Hayes (2017) for a causal diagram. We focus on females with short $tenure$ (less than 0.6 years). OLS gives the following results (showing t-values in parentheses):

$$\begin{aligned}
\hat{Y} &= \underset{(0.652)}{1.3229} - \underset{(-0.661)}{0.1776}X + \underset{(1.130)}{0.5249}M + \underset{(0.286)}{0.0772}C_1 + \underset{(1.318)}{1.6398}C_3, \\
\hat{M} &= \underset{(2.037)}{2.0328} + \underset{(1.120)}{0.1606}X - \underset{(-1.802)}{0.2451}C_1 + \underset{(0.705)}{0.4795}C_3.
\end{aligned}$$

From these estimation results, we get the following test statistics: $(f_1, f_2) = (1.277, 1.254)$ and $(v_1, v_2) = (1.254, 1.277)$, so that $LR = 1.254$. Since $LR < 3.84$, the null of no mediation is not rejected at 5% using the likelihood ratio test. However, we see that both F-statistics are very similar, leading to a ratio $v_1/v_2 = .982$ that is larger than $b_u(.05) = .8744$. Hence, based on the best simply-augmented critical region $CR_{b_u(.05)}$, there is sufficient evidence to reject the null of no mediation. In fact, based on Table 4, the p-value is approximately .59% using linear interpolation. In summary, performing the best simply-augmented LR test requires just one additional step: if the LR test statistic is smaller than the critical value, simply compare the ratio v_1/v_2 to the appropriate value $b_u(\alpha)$ from Table 2 (or Table 4 for other values of α).

7 A higher-dimensional problem

A model analogous to (1)-(2) but with x and m of higher dimension would be:

$$y = X\tau + Y\beta + u, \tag{47}$$

$$Y = X\Pi + U, \tag{48}$$

where X is now $n \times k$ and Y is $n \times p$, say $(k + p < n)$. If we assume that the rows of U are independent with covariance matrix Σ, and that the elements of u are independent with variance σ^2, the second equation here is a multivariate linear model (Muirhead (1982)). A natural analogue of the null hypothesis for the case $k = p = 1$ would be[12]

$$H_0 : \Pi = 0, \text{ or } \beta = 0, \text{ or both.} \tag{50}$$

Tests may, as before, be based on the sufficient statistics (eliminating $\hat{\tau}$)

$$\hat{\beta} = (Y'M_XY)^{-1}Y'M_Xy, ; s_{11} = y'M_{X,Y}y \tag{51}$$

for equation (1), and

$$\hat{\Pi} = (X'X)^{-1}X'Y; S = Y'M_XY. \tag{52}$$

for (2). Then, given Y, the conditional density of $\hat{\beta}$ depends only on S :

$$\hat{\beta}|S \sim N(\beta, \sigma^2 S^{-1}); \quad s_{11}/\sigma^2 \sim \chi^2_{n-k-p}, \tag{53}$$

and

$$\hat{\Pi} \sim N(\Pi, (X'X)^{-1} \otimes \Sigma); \quad S \sim W_p(n - k, \Sigma). \tag{54}$$

Invariance arguments analogous to those used for the simpler case reduce the problem to consideration of the statistics

$$f_1 = \hat{\beta}' S \hat{\beta}/s_{11}; \quad f_2 = \hat{\pi}'((X'X) \otimes S^{-1})\hat{\pi} = tr\left[S^{-1}\hat{\Pi}'X'X\hat{\Pi}\right], \tag{55}$$

where $\hat{\pi} = vec[\hat{\Pi}]$ is $kp \times 1$. Simple arguments give the asymptotic distributions

$$f_1 \sim \chi^2_p(\lambda_1); \quad f_2 \sim \chi^2_{pk}(\lambda_2), \tag{56}$$

where

$$\lambda_1 = \beta'\Sigma\beta/\sigma^2, \lambda_2 = \pi'((X'X) \otimes \Sigma^{-1})\pi = tr\left[\Sigma^{-1}\Pi'X'X\Pi\right], \tag{57}$$

[12] An altenative formulation of the null would be

$$H_0 : \Pi\beta = 0, \tag{49}$$

which would lead to a different analysis altogether. The problem described in the text seems a more relevant formulation.

with $\pi = vec[\Pi]$. The hypothesis becomes, as before,

$$H_0 : \lambda_1 \lambda_2 = 0. \tag{58}$$

However, in this case the problem is invariant under permutations of the variates only if $k = 1$. For the case $k = 1$ all of the results given above will go through unchanged, except that all densities and cdfs have p degrees of freedom rather than 1. For $k > 1$ this permutation invariance does not hold, because the joint density changes under permutations. Thus, when $k > 1$ invariant tests tests are based on (f_1, f_2), and the LR test rejects for $(f_1, f_2) \in \{f_1 > z_\alpha(p), f_2 > z_\alpha(pk)\}$. The size of the test will be either $[1 - G_p(z_\alpha(p); \lambda_1)][1 - G_{pk}(z_\alpha(pk)]$ when $\lambda_2 = 0$, or $[1 - G_p(z_\alpha(p))][1 - G_{pk}(z_\alpha(pk); \lambda_2]$ when $\lambda_1 = 0$. Evidently, because of the loss of symmetry, the null density does depend on which noncentrality parameter is zero, if both are not. At the origin the size is $[1 - G_p(z_\alpha(p))][1 - G_{pk}(z_\alpha(pk))] = \alpha^2$ again, and it will be near this in either case for points near the origin. Thus, to remedy this there is again an incentive to augment the LR test critical region. We leave the remaining details for this case to further work.

8 Conclusion and closing comments

We have demonstrated constructively that exact tests of the no-mediation hypothesis exist for tests of the nominal levels that are typically used in practice. Because the exact test described has some undesirable features, we have also proposed a very simple modification of the LR test for the absence of mediation effects that to a large extent remedies its two main deficiencies: very small size for small values of the nuisance parameter, and very poor power near the origin of the parameter space. The test is extremely easy to understand, and to implement, and it has the important property of coherence. The earlier paper by Van Garderen and Van Giersbergen (2021), in which the idea of augmenting the LR critical region was first proposed for this problem, provides a much more sophisticated method of augmentation. Although discussed and implemented in terms of the variates (t_1, t_2), rather than the order statistics (v_1, v_2) used here, it essentially uses a carefully constructed piecewise-linear spline in place of our simple straight line as the lower boundary of the augmenting region. Correspondingly, it performs somewhat better than the simple method proposed here in both size and power. Our method has the important advantage of extreme simplicity.

There is no doubt that both the approach discussed in this paper, and that of Van Garderen and Van Giersbergen, fall into the class of tests - departures from the likelihood ratio method - that is frowned upon by Perlman and Wu (1999). Both lead to critical regions that imply rejection of the null hypothesis when the observed sample point is close to the origin. And, the arguments claiming an improvement over the LR test are certainly based firmly on the Neyman-Pearson criteria of size and power. If one does not approve, the LR test is still available of course.

9 Appendix A: Proofs

9.1 Proof of Theorem 1

As mentioned in Section 2, the joint distribution of the sufficient statistics is a product of the form

$$f(\hat{\tau},\hat{\beta},s_{11},\hat{\theta},s_{22}) = N\left(\tau-\hat{\theta}\left(\hat{\beta}-\beta\right),\sigma_{11}\left(x'x\right)^{-1}\right)\times N\left(\beta,\sigma_{11}/s_{22}\right)\times\left(\sigma_{11}\chi^2_{n-2}\right)$$
$$\times N(\theta_1,\sigma_{22}(x'x)^{-1})\times\left(\sigma_{22}\chi^2_{n-1}\right)$$

The transformations $s_{11}\to a_1 s_{11}$ and $s_{22}\to a_2 s_{22}$ with $a_1,a_2>0$ leave the joint density of (s_{11},s_{22}) in the same family with $(\sigma_{11},\sigma_{22})$ replaced by $(a_1\sigma_{11},a_2\sigma_{22})$ and have no bearing on the hypothesis under test. The same parameters σ_{11} and σ_{22} are present in the other components, so we need to transform the remaining variables accordingly, namely by: $\hat{\theta}\to\sqrt{a_2}\hat{\theta}$ and $\hat{\beta}\to\sqrt{a_1/a_2}\hat{\beta}$. And, since τ is not involved in the inference problem, we may transform $\hat{\tau}$ by the affine transformation $\hat{\tau}\to\sqrt{a_1}\left(\hat{\tau}+c\right)$

These transformations preserve the family of distributions for the sufficient statistics (and MLEs), and the induced transformation on the mediation effect is that $\theta\beta\to\sqrt{a_1}\theta\beta$. Thus, the transformations do not change the truth or falsity of the hypothesis under test (i.e. H_0 is true before iff it is true after the transformation). The transformations on $\hat{\tau}$ are transitive, so no invariant test can depend on $\hat{\tau}$. We can therefore restrict attention to the four remaining statistics $\left(\hat{\beta},s_{11},\hat{\theta},s_{22}\right)$, and the group $\mathbf{K}$, say, of (scale) transformations of them. The invariance of (T_1,T_2) under the transformations is obvious. To show that (T_1,T_2) are maximal we need to show that $T_1(\hat{\beta},\hat{\theta},s_{11},s_{22})=T_1(\tilde{\beta},\tilde{\theta},\tilde{s}_{11},\tilde{s}_{22})$ and $T_2(\hat{\beta},\hat{\theta},s_{11},s_{22})=T_2(\tilde{\beta},\tilde{\theta},\tilde{s}_{11},\tilde{s}_{22})$ implies that there exists a group element $K\in\mathbf{K}$ such that $(\tilde{\beta},\tilde{\theta},\tilde{s}_{11},\tilde{s}_{22})=K(\hat{\beta},\hat{\theta},s_{11},s_{22})$.

Thus, assume that

$$\hat{\beta}/\sqrt{s_{11}/s_{22}}=\tilde{\beta}/\sqrt{\tilde{s}_{11}/\tilde{s}_{22}} \tag{59}$$

and

$$\hat{\theta}/\sqrt{s_{22}/s_{xx}}=\tilde{\theta}/\sqrt{\tilde{s}_{22}/s_{xx}} \tag{60}$$

Then $\tilde{\theta}=\sqrt{a_2}\hat{\theta}$, with $a_2=\tilde{s}_{22}/s_{22}$, and $\tilde{\beta}=\sqrt{a_1/a_2}\hat{\beta}$ with $a_1=\tilde{s}_{11}/s_{11}$ and a_2 as above. Since also $\tilde{s}_{11}=a_1 s_{11}$, and $\tilde{s}_{22}=a_2 s_{22}$, this shows that the invariance of (T_1,T_2) implies that the two sets of statistics are related by a group element, so (T_1,T_2) are indeed maximal. The same argument applies for the induced group acting on the parameter space, and the last statement is a well-known property of maximal invariants.

9.2 Proof of Theorem 3

Excluding the region $A_r(z)$ from A_3 leaves $r+1$ disjoint triangles lying along the 45° line, the region $w_r(z)\subset A_3$, and it is easy to see that the null probability content of this region is

$$P_{w_r(z)}(\lambda)=(1-\alpha-G(z_r))G(z_\alpha;\lambda)-\sum_{i=1}^{r}\left[G(z_{i+1})-2G(z_i)+G(z_{i-1})\right]G(z_i;\lambda).$$

This differs from the target value $\alpha G(z_\alpha;\lambda)$ by

$$(1-2\alpha-G(z_r))G(z_\alpha;\lambda)-\sum_{i=1}^{r}\left[G(z_{i+1})-2G(z_i)+G(z_{i-1})\right]G(z_i;\lambda). \tag{61}$$

This is a linear combination of $r+1$ non-central chi-square CDFs, the $G(z_i;\lambda)$, and $G(z_{r+1};\lambda) = G(z_\alpha;\lambda)$, and vanishes for all λ if and only if the coefficients of all $r+1$ terms that involve λ vanish. With $\alpha = 1-G(z_{r+1})$, these conditions give rise to a system of $r+1$ linear equations in $r+1$ unknowns $G(z_i), i = 1,...,r+1$ (we take $z_0 = 0$, so that $G(z_0) = 0$), and these determine $z_1,...,z_r$ and $z_{r+1} = z_\alpha$, hence α. It is easy to see that the matrix of the system is non-singular, so the solution is unique, and it is straightforward to check that the solution is:

$$G(z_i) = \frac{i}{r+2}, i = 1,...,r+1, \tag{62}$$

so that $\alpha = 1 - G(z_{r+1}) = (r+2)^{-1}$.

9.3 Proof of Proposition 1

The region $W \leq z$ has the form

$$\{0 < v_1 < v_2, 0 < v_2 < 2z\} \cup \{v_1 \leq \frac{zv_2}{v_2 - z}, v_2 \geq 2z\}. \tag{63}$$

Thus,

$$\begin{aligned}
\Pr\{W \leq z;\lambda_1,\lambda_2\} &= \int_{0<v_1<v_2}\int_{0<v_2<2z} [g(v_1;\lambda_1)g(v_2;\lambda_2) + g(v_2;\lambda_1)g(v_1;\lambda_2)]\, dv_1 dv_2 \\
&\quad + \int_{0<v_1<\frac{zv_2}{v_2-z}}\int_{v_2>2z} [g(v_1;\lambda_1)g(v_2;\lambda_2) + g(v_2;\lambda_1)g(v_1;\lambda_2)]\, dv_1 dv_2 \\
&= G(2z;\lambda_1)G(2z;\lambda_2) \\
&\quad + \int_{v_2>2z}\left[G\left(\frac{zv_2}{v_2-z};\lambda_1\right)g(v_2;\lambda_2) + G\left(\frac{zv_2}{v_2-z};\lambda_2\right)g(v_2;\lambda_1)\right] dv_2,
\end{aligned}$$

as stated. The expression for $P_W(\lambda)$ follows on putting $\lambda_1 = \lambda$, $\lambda_2 = 0$, and $z = z_\alpha$.

9.4 Proof of Proposition 3

Under H_0 the probability content of the augmenting region is given by

$$\begin{aligned}
\Pr\{(v_1,v_2) \in w_b;\lambda) &= \int_{0<v_2<z}\int_{bv_2<v_1<v_2} pdf(v_1,v_2;\lambda)dv_1dv_2 \\
&\quad + \int_{z<v_2<z/b}\int_{bv_2<v_1<z} pdf(v_1,v_2;\lambda)dv_1dv_2.
\end{aligned}\tag{64}$$

Substituting for the density and evaluating the integral over v_1 produces, after simplification,

$$\Pr\{(v_1,v_2) \in w_b;\lambda\} = G(z)[G(z/b;\lambda) - G(z;\lambda)] + G(z;\lambda)G(z/b) \tag{65}$$

$$- \int_{0<v_2<z/b} [g(v_2;\lambda)G(bv_2) + g(v_2)G(bv_2;\lambda)]\, dv_2. \tag{66}$$

Integrating each term in the second line by parts gives

$$\Pr\{(v_1,v_2) \in w_b;\lambda\} = b\int_{0<v_2<z/b} [G(v_2;\lambda)g(bv_2) + G(v_2)g(bv_2;\lambda)]\, dv_2 - G(z)G(z;\lambda). \tag{67}$$

Then, transforming to $v = bv_2$ in the integral, we obtain

$$\Pr\{(v_1,v_2) \in w_b;\lambda\} = \int_{0<v<z} [G(v/b;\lambda)g(v) + G(v/b)g(v;\lambda)]\, dv - (1-\alpha)G(z;\lambda). \tag{68}$$

The result follows.

9.5 Proof of Proposition 4

Expanding the two non-central components in the integrand in $A_\alpha(b;\lambda)$ as Poisson mixtures we have

$$A_\alpha(b;\lambda) = e^{-\frac{1}{2}\lambda}\sum_{j=0}^{\infty}\frac{(\lambda/2)^j}{j!}\int_{0<v<z_\alpha}[g_{2j+1}(v)G(v/b)+g(v)G_{2j+1}(v/b)]dv, \tag{69}$$

and also

$$G(z_a;\lambda) = e^{-\frac{1}{2}\lambda}\sum_{j=0}^{\infty}\frac{(\lambda/2)^j}{j!}G_{2j+1}(z_\alpha). \tag{70}$$

The two power series coincide for all λ if and only if all coefficients agree, that is

$$\int_{0<v<z_\alpha}[g_{2j+1}(v)G(v/b)+g(v)G_{2j+1}(v/b)]dv = G_{2j+1}(z_\alpha) \tag{71}$$

for all j. There is no $b \in (0,1]$ satisfying this equation for all j.

9.6 Proof of Proposition 6

It is well-known that $\lim_{\lambda\to\infty} G(z;\lambda) = 0$ for any finite $z > 0$. The term $A_\alpha(b;\lambda)$ in $D_\alpha(b,\lambda)$ is evidently positive, and is less than $G(z_\alpha/b)G(z_\alpha;\lambda)+G(z_\alpha)G(z_\alpha/b;\lambda)$ for all λ. Since this $\to 0$ as $\lambda \to \infty$ for any $b > 0$, both terms in $D_\alpha(b,\lambda)$ go to zero as $\lambda \to \infty$.

10 Appendix B: Additional Graph and Table

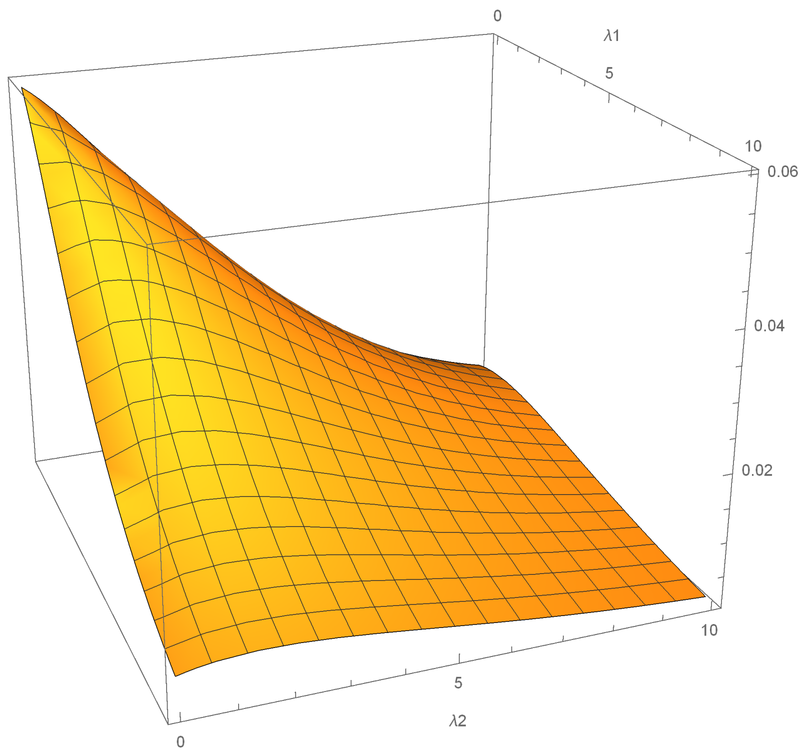

Figure 7: Power difference between the $LR(b_u)$ test and the LR test; $\alpha = .05$.

α	$b_u(\alpha)$	z_α	α	$b_u(\alpha)$	z_α	α	$b_u(\alpha)$	z_α
0.00	1.0000000	∞	0.33	0.6416808	0.9488978	0.66	0.3177177	0.1935236
0.01	0.9696672	6.6348966	0.34	0.6336214	0.9104313	0.67	0.3058554	0.1816021
0.02	0.9418996	5.4118944	0.35	0.6254586	0.8734571	0.68	0.2937820	0.1701258
0.03	0.9168437	4.7092922	0.36	0.6172028	0.8378932	0.69	0.2816199	0.1590854
0.04	0.8943913	4.2178846	0.37	0.6088914	0.8036645	0.70	0.2694031	0.1484719
0.05	0.8744076	3.8414588	0.38	0.6004699	0.7707019	0.71	0.2570733	0.1382770
0.06	0.8568178	3.5373846	0.39	0.5919717	0.7389420	0.72	0.2447520	0.1284927
0.07	0.8468088	3.2830203	0.40	0.5834148	0.7083263	0.73	0.2323988	0.1191116
0.08	0.8376316	3.0649017	0.41	0.5747498	0.6788007	0.74	0.2199257	0.1101266
0.09	0.8287311	2.8743734	0.42	0.5659807	0.6503152	0.75	0.2075763	0.1015310
0.10	0.8201655	2.7055435	0.43	0.5570348	0.6228235	0.76	0.1951608	0.0933185
0.11	0.8118556	2.5542213	0.44	0.5479180	0.5962824	0.77	0.1828105	0.0854831
0.12	0.8036448	2.4173209	0.45	0.5387653	0.5706519	0.78	0.1705896	0.0780191
0.13	0.7957459	2.2925045	0.46	0.5295689	0.5458947	0.79	0.1583169	0.0709213
0.14	0.7878599	2.1779592	0.47	0.5202967	0.5219760	0.80	0.1464034	0.0641848
0.15	0.7800219	2.0722509	0.48	0.5108375	0.4988633	0.81	0.1344952	0.0578047
0.16	0.7723088	1.9742261	0.49	0.5012439	0.4765263	0.82	0.1228917	0.0517767
0.17	0.7647238	1.8829433	0.50	0.4915770	0.4549364	0.83	0.1114934	0.0460968
0.18	0.7570699	1.7976241	0.51	0.4817126	0.4340671	0.84	0.1003974	0.0407610
0.19	0.7494630	1.7176176	0.52	0.4715323	0.4138933	0.85	0.0896371	0.0357658
0.20	0.7419282	1.6423744	0.53	0.4615902	0.3943916	0.86	0.0792623	0.0311078
0.21	0.7344272	1.5714263	0.54	0.4513865	0.3755398	0.87	0.0693260	0.0267841
0.22	0.7267313	1.5043712	0.55	0.4409336	0.3573172	0.88	0.0598636	0.0227917
0.23	0.7192705	1.4408614	0.56	0.4303885	0.3397042	0.89	0.0509320	0.0191281
0.24	0.7116751	1.3805940	0.57	0.4197496	0.3226825	0.90	0.0425622	0.0157908
0.25	0.7040628	1.3233037	0.58	0.4089752	0.3062346	0.91	0.0348502	0.0127777
0.26	0.6964302	1.2687570	0.59	0.3980636	0.2903443	0.92	0.0277903	0.0100869
0.27	0.6886447	1.2167470	0.60	0.3869441	0.2749959	0.93	0.0214448	0.0077167
0.28	0.6810198	1.1670899	0.61	0.3756828	0.2601749	0.94	0.0158541	0.0056656
0.29	0.6733092	1.1196214	0.62	0.3644757	0.2458676	0.95	0.0110635	0.0039321
0.30	0.6654990	1.0741942	0.63	0.3528319	0.2320608	0.96	0.0071027	0.0025154
0.31	0.6575840	1.0306758	0.64	0.3413743	0.2187422	0.97	0.0039968	0.0014144
0.32	0.6495982	0.9889465	0.65	0.3296234	0.2059001	0.98	0.0017679	0.0006285
0.33	0.6416808	0.9488978	0.66	0.3177177	0.1935236	0.99	0.0004349	0.0001571
0.34	0.6336214	0.9104313	0.67	0.3058554	0.1816021	1.00	0.0000000	0.0000000

Table 4: The values $b_u(\alpha)$ and z_α for values of $\alpha \in [0, 1]$.

References

Alwin, D. F. and R. M. Hauser (1975). The decomposition of effects in path analysis. *American sociological review 40*, 37–47.

Baron, R. M. and D. A. Kenny (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology 51*(6), 1173.

Berger, R. L. (1989). Uniformly more powerful tests for hypotheses concerning linear inequalities and normal means. *Journal of the American Statistical Association 84*(405), 192–199.

Coletti, A. L., K. L. Sedatole, and K. L. Towry (2005). The effect of control systems on trust and cooperation in collaborative environments. *The Accounting Review 80*(2), 477–500.

Freedman, L. S. and A. Schatzkin (1992). Sample size for studying intermediate endpoints within intervention trials or observational studies. *American Journal of Epidemiology 136*(9), 1148–1159.

Ghosh, B. (1973). Some monotonicity theorems for $\chi 2$, f and t distributions with applications. *Journal of the Royal Statistical Society: Series B (Methodological) 35*(3), 480–492.

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford publications.

Heckman, J. and R. Pinto (2015a). Causal analysis after haavelmo. *Econometric Theory 31*(1), 115–151.

Heckman, J. and R. Pinto (2015b). Econometric mediation analyses: Identifying the sources of treatment effects from experimentally estimated production technologies with unmeasured and mismeasured inputs. *Econometric reviews 34*(1-2), 6–31.

Lehmann, E. (1952). Testing multiparameter hypotheses. *The Annals of Mathematical Statistics*, 541–552.

MacKenzie, S. B., R. J. Lutz, and G. E. Belch (1986). The role of attitude toward the ad as a mediator of advertising effectiveness: A test of competing explanations. *Journal of marketing research 23*(2), 130–143.

MacKinnon, D. P., C. M. Lockwood, J. M. Hoffman, S. G. West, and V. Sheets (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological methods 7*(1), 83.

Muirhead, R. (1982). *Aspects of Multivariate Statistical Theory*. Wiley Series in Probability and Statistics.

Nomakuchi, K. and T. Sakata (1987). A note on testing two-dimensional normal mean. *Annals of the Institute of Statistical Mathematics 39*(3), 489–495.

Perlman, M. D. and L. Wu (1999). The emperor's new tests. *Statistical Science 14*(4), 355–369.

Pollack, J. M., E. M. Vanepps, and A. F. Hayes (2012). The moderating role of social ties on entrepreneurs' depressed affect and withdrawal intentions in response to economic stress. *Journal of Organizational Behavior 33*(6), 789–810.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological methodology 13*, 290–312.

Van Garderen, K. J. and N. P. A. Van Giersbergen (2021). A nearly similar powerful test for mediation. Mimeo,UvA.

Vaughan, R. J. and W. N. Venables (1972). Permanent expressions for order statistic densities. *Journal of the Royal Statistical Society: Series B (Methodological) 34*(2), 308–310.