



UvA-DARE (Digital Academic Repository)

On the Optimality of the LR Test for Mediation

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

DOI

[10.3390/sym14010178](https://doi.org/10.3390/sym14010178)

Publication date

2022

Document Version

Final published version

Published in

Symmetry

License

CC BY

[Link to publication](#)

Citation for published version (APA):

van Garderen, K. J., & van Giersbergen, N. (2022). On the Optimality of the LR Test for Mediation. *Symmetry*, 14(1), Article 178. <https://doi.org/10.3390/sym14010178>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Article

On the Optimality of the LR Test for Mediation

Kees Jan Van Garderen ^{*,†}  and Noud Van Giersbergen [†] 

UvA Econometrics, Amsterdam School of Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands; N.P.A.vanGiersbergen@uva.nl

* Correspondence: K.J.vanGarderen@uva.nl

† These authors contributed equally to this work.

Abstract: Testing for mediation, or indirect effects, is empirically very important in many disciplines. It has two obvious symmetries that the testing procedure should be invariant to. The ordered absolute t-statistics from two ordinary regressions are maximal invariant under the associated groups of transformations. Sobel's (1982) Wald-type and the LR test statistic are both functions of this maximal invariant and satisfy two logical coherence requirements: (1) size coherence: rejection at level α implies rejection at all higher significance levels; and (2) information coherence: more (less) evidence against the null implies continued (non) rejection of the null. The LR test statistic is simply the smallest of the two absolute t-statistics, and we show that the LR test is the Uniformly Most Powerful (information and size) Coherent Invariant (UMPCI) test. In short: the LR test for mediation is simple and best.

Keywords: coherence; indirect effect; maximal invariant; Sobel; statistics



Citation: Van Garderen, K.J.; Van Giersbergen, N. On the Optimality of the LR Test for Mediation. *Symmetry* **2022**, *14*, 178. <https://doi.org/10.3390/sym14010178>

Academic Editor: Juan Alberto Rodríguez Velázquez

Received: 20 December 2021

Accepted: 10 January 2022

Published: 17 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Testing for mediation is empirically extremely important in many scientific disciplines. For example, in psychology: Baron and Kenny [1], which has more than 100,000 citations; in accounting: Coletti et al. [2]; in marketing: MacKenzie et al. [3]; in sociology: Alwin and Hauser [4]; in economics: Huber [5], amongst many others. It has also generated much methodological interest because it is a non-standard problem. Despite its great popularity, Sobel's original test has serious shortcomings, including extremely low rejection probabilities when the indirect effect is small or estimated with large variance. See MacKinnon et al. [6] for an overview of various testing procedures. The aim of mediation testing is to establish if the causal mechanism of the effect that an independent variable (X) has on a dependent variable (Y) is via a mediating variable (M). The most basic version of the model with all variables in deviation from their means is:

$$Y = \beta X + \theta_2 M + u, \quad (1)$$

$$M = \theta_1 X + v, \quad (2)$$

where disturbances u and v are assumed to be independent, since Y does not influence M , e.g., because of the experimental set up. The causal variable X affects Y via two different pathways. The first is the *direct* effect of X on Y and is quantified by β . The second is the *indirect* effect of X on Y via the mediating variable M . This mediation effect is quantified by $\theta_1\theta_2$ and only exists if both θ_1 and θ_2 are non-zero. The null hypothesis of no mediation is commonly expressed as:

$$H_0 : \theta_1\theta_2 = 0,$$

but it is not a standard testing problem, since the null has a singular point in the origin where the two axes ($\theta_1 = 0$ and $\theta_2 = 0$) cross (see Drton [7]) and the distribution of the test statistic is non-standard as a result.

The famous test by Sobel [8] is a Wald-type test that standardizes the estimator $\hat{\theta}_1 \hat{\theta}_2$. The test can be expressed in terms of the absolute values of ordinary t -statistics ($|T_1|, |T_2|$) in (1) and (2). This renders a test that is invariant to the parameter value β and to the variances of u and v . This is no coincidence: all invariant testing procedures that exploit all the information in the data and the model must be based on $(|T_1|, |T_2|)$ since it is a maximal invariant as we will see below.

An obvious alternative to the Wald test is the Likelihood Ratio (LR) test, which provides an optimal test when the null and alternative are both simple. Testing for mediation, however, involves composite null and alternative hypotheses, as well as nuisance parameters. The Neyman-Pearson lemma is not applicable, and no uniformly most powerful test exists.

We show, nevertheless, that a uniformly most powerful test exists within a class of procedures that are information coherent and invariant. Information coherence is the logical requirement introduced in Section 3, that, when information against the null increases, a test should continue to reject. It differs from the common size coherence that requires the same when the size (maximum probability of a Type I error) is increased. Both these coherency requirements are very mild. The LR test is the best coherent invariant test. Establishing this result is the main contribution of this paper.

A likelihood analysis of the estimators and LR test for mediation requires a distributional assumption. We will assume a normal distribution for each observation $1 \leq i \leq n$: $(u_i, v_i)' \sim IIN(0, \text{diag}(\sigma_{11}, \sigma_{22}))$. This is convenient, but the analysis is also valid asymptotically without assuming normality of the errors, since the t -statistics used are still asymptotically normal.

The joint density of (Y, M) given X can be written as:

$$f_{Y,M|X}(y, m|x; \lambda) = f_{Y|M,X}(y|m, x; \lambda_1) \times f_{M|X}(m|x; \lambda_2),$$

with $\lambda_1 = (\beta, \theta_2, \sigma_{11})'$, $\lambda_2 = (\theta_1, \sigma_{22})'$, and $\lambda = (\lambda_1', \lambda_2')'$. The parameters λ_1 and λ_2 vary freely as a result of the triangular (recursive) structure of the model. The mediation variable is the endogenous variable in (2), but is exogenous for θ_2 in (1). There is no feedback or causal relation from Y to M . This can easily be extended to include more regressors/covariates. This will affect the degrees of freedom in finite samples, but not the asymptotic distribution of the t -statistics. One could also use instrumental variables, but X and M appear in both equations and in the standard setup u and v are independent because of the experimental interpretation of M and, without it, parameters are not identified. The log-likelihood given n independent observations is the sum of two Gaussian log-likelihoods corresponding to (1) and (2):

$$\ell(\lambda) \propto -\frac{1}{2\sigma_{11}} \sum_{i=1}^n (y_i - \beta x_i - \theta_2 m_i)^2 - \frac{n}{2} \log(\sigma_{11}) - \frac{1}{2\sigma_{22}} \sum_{i=1}^n (m_i - \theta_1 x_i)^2 - \frac{n}{2} \log(\sigma_{22}). \quad (3)$$

As a consequence, the Maximum Likelihood Estimators (MLEs) for θ_1 and θ_2 are the usual OLS estimators for the two equations separately. The MLE for the full λ is minimal sufficient, and its dimension is equal to the number of parameters. The model is a full exponential model (see Van Garderen [9]), and the MLE is a complete sufficient statistic. Inference on the parameters can, therefore, be based on the MLE without loss of information. Randomization in the test procedure will lead to a loss of information and power. Hence, we will not consider randomized tests or bootstrap procedures. Finally, the information matrices will be block diagonal in λ_1 and λ_2 as well as in $(\beta, \theta_2)'$, σ_{11} , θ_1 , and σ_{22} . As a result the standard t -statistics, T_1 and T_2 for θ_1 and θ_2 are asymptotically independent and normally distributed.

2. Symmetry and Invariance

It is clear that the problem has a number of symmetries and invariances. The null hypothesis of no mediation $\theta_1 \theta_2 = 0$ remains true or false if we:

- (i) Interchange θ_1 and θ_2 ;
- (ii) Change the signs of θ_1 or θ_2 ;
- (iii) Change the values of the nuisance parameters σ_{11} , σ_{22} , or β .

We want a test procedure that respects these symmetry and invariance properties. This may not be straightforward because the distributions of test statistics will generally depend on nuisance parameters, even under the null. Moreover, the t -statistics have different degrees of freedom. We will first establish the maximal invariance of the two t -statistics under a group of location-scale transformations in finite samples. This covers the invariance with respect to nuisance parameters in (iii). Invariant procedures based on the minimal sufficient statistics should depend on the t -statistics only, in finite samples as well as asymptotically. By moving to the asymptotic normal distribution, we can abstract from the difference in the degrees of freedom and consider the permutation invariance in (i) and the reflection invariance in (ii).

The finite-sample distribution involves t -distributions with different degrees of freedom and is further complicated by the fact that the standard deviation of $\hat{\theta}_2$, $\sigma_{\hat{\theta}_2}$, depends on M . Even in finite samples, however, we can derive exact results that provide a strong justification for restricting attention to the t -statistics. Proposition 1, derived in Hillier et al. [10], is a particular case in point and establishes that T is a maximal invariant regardless of sample size.

Proposition 1. *The testing problem is invariant under the group \mathbf{K} of transformations acting on $(\hat{\beta}, \hat{\theta}_2, s_{11}, \hat{\theta}_1, s_{22})$ defined by:*

$$(\hat{\beta}, \hat{\theta}_2, s_{11}, \hat{\theta}_1, s_{22}) \rightarrow (\sqrt{a_1}(\hat{\beta} + a_0), \sqrt{a_1/a_2}\hat{\theta}_2, a_1s_{11}, \sqrt{a_2}\hat{\theta}_1, a_2s_{22}),$$

$$a_0 \in \mathbb{R}, a_1, a_2 \in \mathbb{R}^+.$$

The induced group of transformations $\bar{\mathbf{K}}$ acting on the parameter space is defined by:

$$(\beta, \theta_2, \sigma_{11}, \theta_1, \sigma_{22}) \rightarrow (\sqrt{a_1}(\beta + a_0), \sqrt{a_1/a_2}\theta_2, a_1\sigma_{11}, \sqrt{a_2}\theta_1, a_2\sigma_{22})$$

The vector of t -statistics is a maximal invariant statistic under the group of transformations \mathbf{K} :

$$T = (T_1, T_2)' = \left(\hat{\theta}_1 / \sqrt{\frac{1}{n-2} s_{22} (x'x)^{-1}}, \hat{\theta}_2 / \sqrt{\frac{1}{n-3} s_{11} / s_{22}} \right)'.$$

A parameter-space maximal invariant under the induced group $\bar{\mathbf{K}}$ is:

$$\mu = (\mu_1, \mu_2)' = \left(\theta_1 / \sqrt{\sigma_{22} (x'x)^{-1}}, \theta_2 / \sqrt{\frac{\sigma_{11}}{\sigma_{22} / n}} \right)'.$$

The distribution of $(T_1, T_2)'$ depends only on $(\mu_1, \mu_2)'$.

This proposition justifies restricting the attention to the two t -statistics, even in finite samples. The proof is given in Hillier et al. [10] and further establishes that T_1 and T_2 are independent. The t -statistic T_2 has one degree of freedom less than T_1 , since (1) has one more variable than (2). This difference compromises the permutation symmetry but is practically unimportant, unless the number of observations is small, for instance, less than 30. Throughout the rest of the paper, we will, therefore, use the limiting normal distribution for the t -statistics. Hence, let $\mu = (\mu_1, \mu_2)' = (\theta_1^0 / \sigma_{\theta_1}, \theta_2^0 / \sigma_{\theta_2})'$ where θ_1^0, θ_2^0 denote the true parameter values and $\sigma_{\theta_1}, \sigma_{\theta_2}$ the standard deviations of the OLS estimators, then asymptotically: $(T - \mu) \xrightarrow{d} N(0, I_2)$, but we continue as if this is the exact distribution and state this explicitly in:

Assumption 1. *Gaussian Shift Experiment*

$$T \sim N(\mu, I_2)$$

Turning to symmetries (i) and (ii), let G_1 be the group of permutations and G_2 the group of reflections or sign changes. The two groups only have the identity element in common, and the full group $G = G_1 \times G_2$, generated by G_1 and G_2 , has $2!2^2 = 8$ elements. The density after a sign change in T_k is obtained by a corresponding sign change in μ_k , and density for a permutation of T is also obtained by permuting μ accordingly. Hence, for any element $g \in G$, we have $g \cdot T \sim N(g \cdot \mu, I_2)$ or $P_{g\mu}[gT \in A] = P_\mu[T \in A]$, so the distribution is invariant; see Lehmann and Romano [11].

It is clear that the absolute order statistic, denoted $(|T|_{(1)}, |T|_{(2)})$ with $0 \leq |T|_{(1)} \leq |T|_{(2)}$, is invariant because its value does not change by permuting or changing signs in T . It is also maximal invariant because, for two T and \tilde{T} such that $(|T|_{(1)}, |T|_{(2)}) = (|\tilde{T}|_{(1)}, |\tilde{T}|_{(2)})$, this can only occur if the two elements in \tilde{T} are a permutation of the two elements in T and possible changes in signs of the elements. There must exist a transformation $g = g_1 \cdot g_2 \in G_1 \times G_2$ s.t. $\tilde{T} = g \cdot T$. The same argument holds for the ordered absolute parameter $(|\mu|_{(1)}, |\mu|_{(2)})$ since the group of transformations is the same on parameter and sample space. We have, therefore, established the following:

Proposition 2. *The testing problem $H_0 : \mu_1\mu_2 = 0$ is invariant under the group of transformations $G = G_1 \times G_2$ acting on T and μ , given Assumption 1. The absolute order statistic $(|T|_{(1)}, |T|_{(2)})$ with $0 \leq |T|_{(1)} \leq |T|_{(2)}$ is a maximal invariant statistic, and the absolute order parameter $(|\mu|_{(1)}, |\mu|_{(2)})$ with $0 \leq |\mu|_{(1)} \leq |\mu|_{(2)}$ is a maximal invariant parameter under the group of transformations $G = G_1 \times G_2$. The distribution of $(|T|_{(1)}, |T|_{(2)})$ depends only on $(|\mu|_{(1)}, |\mu|_{(2)})$.*

Note that $(|T|_{(1)}, |T|_{(2)})$ is a function of the complete sufficient statistics and, hence, a maximal invariant that exploits all the information in the data. A randomized test or another procedure that is not a function of $(|T|_{(1)}, |T|_{(2)})$ cannot do better in terms of power.

For deriving the joint distribution of this maximal invariant $(|T|_{(1)}, |T|_{(2)})$, first note that if scalar $T \sim N(\mu, 1)$, then $|T|$ has the folded normal distribution (noncentral Chi-distribution with one degree of freedom) with density:

$$f_{|T|}(t, \mu) = \phi(t + \mu) + \phi(t - \mu) = \sqrt{\frac{2}{\pi}} \exp(-(t^2 + \mu^2)/2) \cosh(\mu t), \quad \text{for } t \geq 0, \quad (4)$$

with $\phi(\cdot)$ the standard normal density function. Second, $|T_1|$ and $|T_2|$ are independent and, hence, by Equation (6) of Vaughan and Venables [12], we obtain after simplification:

Lemma 1. *The density of the maximal invariant stated in Proposition 2 equals:*

$$f_{|T|_{(1)}, |T|_{(2)}}(t_1, t_2, \mu_1, \mu_2) = \frac{2}{\pi} \{ \cosh(t_1\mu_1) \cosh(t_2\mu_2) + \cosh(t_1\mu_2) \cosh(t_2\mu_1) \} \times e^{-(t_1^2+t_2^2+\mu_1^2+\mu_2^2)/2}, \quad \text{for } 0 \leq t_1 \leq t_2. \quad (5)$$

3. Test Statistics, Critical Regions, and Coherence

The previous section established that invariant test procedures should be based on the ordered absolute t -statistic. The classic test by Sobel [8] and the LR test are both functions of $(|T|_{(1)}, |T|_{(2)})$. The Sobel test statistic is the square root of the Wald statistic W based on $\hat{\theta}_1\hat{\theta}_2$ and its standard error derived by the delta method:

$$\sqrt{W} = \left(\frac{T_1^2 T_2^2}{T_1^2 + T_2^2} \right)^{1/2} = \frac{|T|_{(1)} |T|_{(2)}}{\sqrt{|T|_{(1)}^2 + |T|_{(2)}^2}}. \tag{6}$$

The LR test statistic is easily obtained by maximizing the log-likelihood with and without the restriction that either θ_1 or θ_2 is zero and:

$$LR = \min\{|T_1|, |T_2|\} = |T|_{(1)}. \tag{7}$$

An equivalent LR test can also be based on the minimum of two F-statistics $f_1 = T_1^2$ and $f_2 = T_2^2$.

In both cases, we reject when the test statistic is larger than the two-sided critical value from a standard normal distribution. The rejection probability for both tests is monotonically increasing in $|\mu|_{(2)}$ under the null with $|\mu|_{(1)} = 0$. The size, i.e., the maximum rejection probability under H_0 , is attained when $|\mu|_{(2)} \rightarrow \infty$. The largest absolute t -value will diverge in that case, and the Sobel and LR test simply reject when the smallest absolute t -value is larger than the usual critical value, 1.96 for a 5% level test.

Rather than the one-dimensional test statistics, we can define tests in terms of their critical region (CR) in the sample space of dimension two. The sample space for $(|T_1|, |T_2|)$ is the quadrant \mathbb{R}_+^2 , and the sample space for the maximal invariant $(|T|_{(1)}, |T|_{(2)})$ is an octant in \mathbb{R}^2 : with $\mathbb{T} = \{(|T|_{(1)}, |T|_{(2)}) \in \mathbb{R}_+^2 \mid 0 \leq |T|_{(1)} \leq |T|_{(2)} \leq \infty\}$. The LR critical region is:

$$CR_{LR}(\alpha) = \{ \mathbb{T} \mid z_{\alpha/2} \leq |T|_{(1)} \leq |T|_{(2)} \leq \infty \},$$

with $z_{\alpha/2}$ the (two-sided) critical value of standard normal variate, i.e., $P[|Z| > z_{\alpha/2}] = \alpha$.

The boundary of the critical regions for the LR and Sobel test in the sample space for $(|T_1|, |T_2|)$ are given in Figure 1.

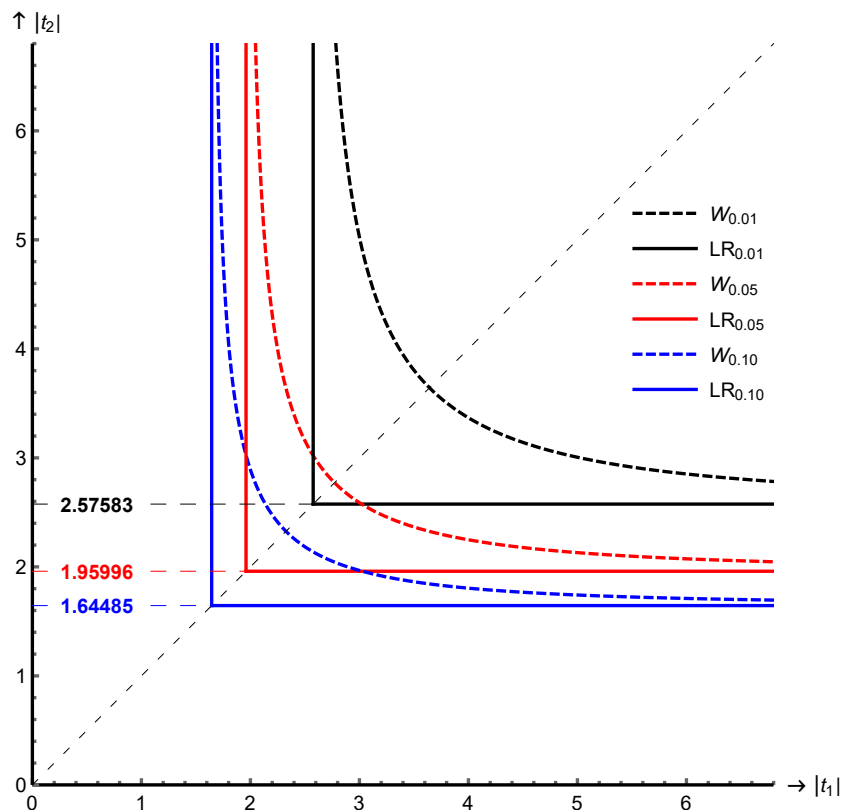


Figure 1. Boundaries ∂CR for Sobel-(Wald) and LR tests in \mathbb{R}_+^2 , the sample space for $|t|$.

Both the LR and Wald CR have two desirable properties. First, if they reject at level α , then they also reject at any other level that is larger than α . Second, if evidence against the null is accumulating by increasing one or both of the t -statistics, the tests will continue to reject, and, if evidence is decreasing, they will continue to accept (not reject) the null hypothesis. The first property we call *size* or α coherence, with α added to the usual term to distinguish it from the second property we call *information* coherence. We formalize these concepts next. This allows us to prove the main theorem that the LR is the most powerful CR that respects both coherence properties.

Information and Size Coherence

In general, it is desirable for testing procedures to continue to reject if the evidence against the null hypothesis is increased and to continue to accept (i.e., not reject) when evidence against the null is decreasing. Thus, if a t -statistic for a single parameter in a one-sided alternative is increasing and we reject for $t = 2$, then we should also want to reject for $t > 2$ because this represents a value that is even less likely, more extreme, under the null and commonly interpreted as more evidence against the null. In multivariate settings, this is less trivial because no uniformly most powerful test exists, and the separate test statistics might be correlated. Nevertheless, in the mediation case with $H_0 : \theta_1 = 0 \vee \theta_2 = 0$, the two t -tests are independent, and it seems reasonable to require that one continues to reject when either t_1 and/or t_2 is increasing in absolute value, since the information against the null is strengthening, and continue to accept when either t_1 and/or t_2 is decreasing towards 0, since the information against the null is weakened. We define a class of size α critical regions that formalizes this requirement and show that the LR test is optimal in this class of tests that respects information coherence. Consider CR and acceptance region (AR) for $(|T_1|, |T_2|) \in \mathbb{R}_+^2$.

Definition 1. Information Coherence. \mathbb{C}_α is the class of all information coherent critical regions of size α defined by the property that, with $\delta_1, \delta_2 \in \mathbb{R}_+$

- (i) for any $CR \in \mathbb{C}_\alpha$ if $(|T_1|, |T_2|) \in CR \Rightarrow (|T_1| + \delta_1, |T_2| + \delta_2) \in CR$,
or, equivalently,
- (ii) for any $AR \in \mathbb{C}_\alpha$ if $(|T_1|, |T_2|) \in AR \Rightarrow (|T_1| - \delta_1, |T_2| - \delta_2) \in AR$.

Traditional (α) coherence is a property of a family of CRs when the size of the test varies, but information coherence considers a fixed α and varying values of the test statistics.

Definition 2. Size Coherence. A family of $CR(\alpha)$ indexed by its size α is size or α coherent iff $CR(\alpha) \subseteq CR(\tilde{\alpha}) \Leftrightarrow \alpha \leq \tilde{\alpha}$.

A smaller significance level requires more extreme observations and, hence, a smaller CR. Note that the definition of α coherence does not require the definition of the statistics involved, but information coherence uses explicit statistics.

We show in Appendix A that a family of information coherent CRs has the following properties, with ∂CR the boundary of CR in \mathbb{R}_+^2 that separates the CR from the AR, and $cv_1(t)$ and $cv_2(t)$ the critical values for $|T_1|$ and $|T_2|$, respectively, either simultaneously or conditional on the other:

Proposition 3. Any $CR \in \mathbb{C}_\alpha$ and its boundary ∂CR have the following topological and statistical properties:

- (i) CR is simply connected.
- (ii) ∂CR is a continuous plane curve.
- (iii) ∂CR is monotonically weakly decreasing.
- (iv) ∂CR can be parametrically represented globally as $\partial CR(\tau) = (cv_1(\tau), cv_2(\tau))$ using a one-dimensional $\tau \in \mathbb{R}$, and locally as $\partial CR(t_1) = (t_1, cv_2(t_1))$ if not vertical, and/or $\partial CR(t_2) = (t_2, cv_1(t_2))$ if not horizontal.

- (v) The class \mathbb{C}_α contains critical regions that may not be convex, but no critical regions that are strictly concave.
- (vi) A test with $CR \in \mathbb{C}_\alpha$ can only be size correct and admissible if $\lim_{t_1 \rightarrow \infty} cv_2(t_1) = z_{\alpha/2}$.

Proposition 3 is instrumental in proving the optimality of the LR test and showing that the Sobel and LR tests are information and size coherent.

Proposition 4. The Sobel (Wald) test and LR test of size α , both respect

- information coherence: $CR_{\sqrt{W}}(\alpha), CR_{LR}(\alpha) \in \mathbb{C}_\alpha$ as well as
- size coherence: $CR_{LR}(\alpha) \subset CR_{LR}(\tilde{\alpha})$ and $CR_{\sqrt{W}}(\alpha) \subset CR_{\sqrt{W}}(\tilde{\alpha})$ iff $\tilde{\alpha} > \alpha$.

This result only states that both tests share two desirable properties, but does not imply that both tests are equally good. The power of the LR is much better than Sobel's test, which suffers from extremely low power when the mediation effect is small or is inaccurately estimated. In fact, the LR test is better than any other coherent test in \mathbb{C}_α , which we are now able to state and prove in the main theorem of the paper.

Theorem 1. The LR test of size α is the uniformly most powerful test in \mathbb{C}_α .

Proof. The $CR_{LR}(\alpha) = \{\min\{|T_1|, |T_2|\} \geq z_{\alpha/2}\}$ sets the critical value for the $cv_1(t_2) = z_{\alpha/2}$ for all $t_2 \geq t_1 \geq 0$, which is the smallest value that cv_1 can take for all values of t_2 while still being size-correct according to (iii) and (iv) of Proposition 3. Analogously, $cv_2(t_1) = z_{\alpha/2}$ for all $t_1 \geq t_2$. So $CR_{LR}(\alpha)$ is the closure of \mathbb{C}_α and any member $CR(\alpha) \subseteq CR_{LR}(\alpha)$. Hence, $P[CR_{LR}(\alpha) | H_a]$ is larger than for any other $CR(\alpha) \in \mathbb{C}_\alpha$ that is not equal to $CR_{LR}(\alpha)$ a.s. This holds uniformly for all values of μ under H_0 . \square

This optimality property of the LR test is derived under coherence requirements that are very weak: it seems more than reasonable to require that any test continues to reject if more extreme outcomes are observed or if the level of the test is increased.

4. Discussion and Conclusions

This paper has exploited the symmetry present in the mediation testing hypothesis $H_0 : \theta_1\theta_2 = 0$ and used invariance arguments to reduce the sample space to an eighth of \mathbb{R}^2 . We have developed a coherence framework to formulate and analyze the requirement that increasing or decreasing information against the null leads to coherent decisions. We call tests or CRs with this property information coherent, which is distinct from the more standard (size) coherent property that tests may possess. The Sobel test is both information and size coherent, but has very poor null rejection and power properties.

The LR test is much better than the Sobel test, and this paper shows the LR test to be the best possible of all tests that satisfy the basic coherence requirements. The optimality lends support to Perlman and Wu [13] on their preference for the LR test.

Nevertheless, the LR test has some serious shortcomings, like the Sobel test, when detecting deviations from H_0 . In particular, when both θ_1 and θ_2 are close to 0, or are estimated inaccurately, it is extremely conservative and the power deteriorates and goes to α^2 when $\mu \rightarrow (0, 0)'$.

The bootstrap does not provide an answer because of the (strong) dependency on the nuisance parameter. Of course, it may provide small-sample corrections and avoid asymptotic approximations and strong distributional assumptions. However, bootstrap versions of the LR and Sobel test still lack power and are neither size nor information coherent.

Van Garderen and Van Giersbergen [14] show there is an opportunity to gain power without violating the size condition by adding a small area near the diagonal where the opportunity to detect deviations from H_0 are best. It is uniformly more powerful than the LR test, but is not in \mathbb{C}_α since the boundary ∂CR is monotonically increasing and, hence, violating property (iii) of Proposition 3. Therefore, the LR test remains the optimal coherent choice for mediation testing. It is simple and best.

Author Contributions: Both authors contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proofs

Throughout this Appendix we use:

$$t = (t_1, t_2)', v = (v_1, v_2)' \in \mathbb{R}^2, |t| = (|t_1|, |t_2|)', \delta = (\delta_1, \delta_2)' \in \mathbb{R}_+^2.$$

In order to prove the properties of $CR \in \mathbb{C}_\alpha$ and its boundary, we first note from the definition of \mathbb{C}_α that if $|t| \in CR \subseteq \mathbb{R}_+^2$, then CR also contains the top right quarter defined by the intersection of two half-planes:

$$Q(|t|) = \{v \in \mathbb{R}^2 | v_1 \geq |t_1|, v_2 \geq |t_2|\},$$

i.e., a quarter with $|t|$ as its bottom left corner. The interior of $Q(|t|)$ will be denoted:

$$Q^\circ(|t|) = \{v \in \mathbb{R}^2 | v_1 > |t_1|, v_2 > |t_2|\}.$$

Proposition A1. *In Definition 1, the conditions on CR and AR , i.e., (i) and (ii), are equivalent.*

Proof. If there is a $|t| \in CR$ and a $\delta = (\delta_1, \delta_2) \in \mathbb{R}_+^2$ such that $|\tilde{t}| = (|t_1| + \delta_1, |t_2| + \delta_2) \in AR$, then there is an element in AR such that $|\tilde{t}| - \delta \in CR$. So for any $|t|$ for which (i) does not hold, there is an $|\tilde{t}|$ for which (ii) does not hold. Similarly, (i) cannot hold if (ii) does not hold. Hence, (i) and (ii) are equivalent. \square

Proof of Proposition 3. Topological and statistical properties (i)–(vi).

\square

Property (i). Any $CR \in \mathbb{C}_\alpha$ is simply connected.

Proof. If $|t|, |\tilde{t}| \in CR$ then a path connecting the two points can be constructed inside $Q(|t|) \cup Q(|\tilde{t}|)$. This path can be a single straight line if $|t| \in Q(|\tilde{t}|)$ or $|\tilde{t}| \in Q(|t|)$, with $|t_1| \geq |\tilde{t}_1| \wedge |t_2| \geq |\tilde{t}_2|$ or $|\tilde{t}_1| \geq |t_1| \wedge |\tilde{t}_2| \geq |t_2|$, respectively. If $|t_1| > |\tilde{t}_1|$ and $|t_2| < |\tilde{t}_2|$, then a path by two linear parts exists: first, $(|t_1|, |t_2|) \rightarrow (|t_1|, |\tilde{t}_2|)$ linearly, by a linear contraction $C : [0, 1] \times \mathbb{R} \rightarrow \mathbb{R}^2 : (\lambda, |t_2|) \mapsto (|t_1|, |t_2| + \lambda(|\tilde{t}_2| - |t_2|))$. This line piece lies in $Q(|t|)$ by definition since $|t_2| + \lambda(|\tilde{t}_2| - |t_2|) > |t_2|$ and $|t_1|$ is fixed. Any closed path $\gamma : \mathbb{S} \mapsto Q(|t|)$, with $\mathbb{S} \subseteq \mathbb{R}$, can be shrunk to $(|t_1|, |\tilde{t}_2|)$. Second, $(|t_1|, |\tilde{t}_2|) \rightarrow (|\tilde{t}_1|, |\tilde{t}_2|)$ linearly, which lies in $Q(|\tilde{t}|)$ because $|t_1| > |\tilde{t}_1|$ and $|\tilde{t}_2|$ is fixed. Any closed path $\gamma : \mathbb{S} \mapsto Q(|\tilde{t}|)$ can be shrunk to $(|\tilde{t}_1|, |\tilde{t}_2|)$. For $|t_1| < |\tilde{t}_1|$ and $|t_2| > |\tilde{t}_2|$, the result can be shown analogously but moving first from $(|\tilde{t}_1|, |\tilde{t}_2|) \rightarrow (|\tilde{t}_1|, |t_2|)$ and then from $(|\tilde{t}_1|, |t_2|) \rightarrow (|t_1|, |t_2|)$. Combining these two results, we have that any closed path $\gamma : \mathbb{S} \mapsto Q(|t|) \cup Q(|\tilde{t}|)$ can be shrunk to $(\max(|t_1|, |\tilde{t}_1|), \max(|t_2|, |\tilde{t}_2|))$. There are no holes and the genus of $Q(|t|) \cup Q(|\tilde{t}|)$ is 0. Hence, $CR \in \mathbb{C}_\alpha$ is simply connected. \square

Properties (ii) and (iii). The boundary ∂CR of any $CR \in \mathbb{C}_\alpha$ is a continuous plane curve that is monotonically (weakly) decreasing in the sense that, if $t, \tilde{t} \in \partial CR$, then $\tilde{t}_1 \geq t_1 \Leftrightarrow \tilde{t}_2 \leq t_2$, but may be non-differentiable at infinitely many points.

Proof. If $|t| \in \partial CR$ of any $CR \in \mathbb{C}_\alpha$, then any $|\tilde{t}|$ with $|\tilde{t}_1| \geq |t_1|$ and $|\tilde{t}_2| = |t_2|$ is either on the boundary ∂CR or in the interior of CR . This implies that the boundary for this value of

$|\tilde{t}_1|$ is equal or below $|t_2|$ and, therefore, weakly decreasing. That the boundary is a plane curve and continuous follows from Lemma A1 below, and the construction of the paths in the proof of the simple connectedness. The curve can be a step function with an infinite number of infinitesimal steps. \square

Property (iv). ∂CR can be parametrically represented globally as $\partial CR(\tau) = (cv_1(\tau), cv_2(\tau))$ and locally as $\partial CR(t_1) = (t_1, cv_2(t_1))$ if not vertical and/or $\partial CR(t_2) = (t_2, cv_1(t_2))$ if not horizontal.

Proof. By properties (i)–(iii) and their proofs, it follows immediately that we can parameterize the boundary as a continuous function of a one-dimensional parameter τ . If the boundary is not vertical, we can locally use $|t_1|$ as parameter and, if the boundary is not horizontal, then we can locally use $|t_2|$ as parameter. Where it is vertical, the boundary associates a given $|t_1|$ with a set of points for $|t_2|$, and, when it is horizontal, it associates a set of points $|t_1|$ with a given $|t_2|$. \square

Property (v). The class \mathbb{C}_α contains critical regions that may not be convex, but no critical regions that are strictly concave.

Proof. \mathbb{C}_α contains CRs with step function boundaries. They are not convex. Nonconvexity is obvious since ∂CR is monotonically (weakly) decreasing. \square

Property (vi). A test can only be size correct and admissible if $\lim_{t_1 \rightarrow \infty} cv_2(t_1) = z_{\alpha/2}$.

Proof. The boundary ∂CR is weakly decreasing, and the critical value for $|t_2|$ will be smallest when $|t_1| \rightarrow \infty$. So we can define the minimum and limiting value:

$$cv_2^* = \min_{|t_1| \geq 0} cv_2(|t_1|) = \lim_{|t_1| \rightarrow \infty} cv_2(|t_1|).$$

and, similarly, $cv_1^* = \min cv_1(|t_2|) = \lim_{|t_2| \rightarrow \infty} cv_1(|t_2|)$. The normal distribution $N(\mu, 1)$ has survival function $S_\mu(c) = \frac{1}{2} - \frac{1}{2} \operatorname{erf}\left(\frac{c - \mu}{\sqrt{2}}\right)$, which has the property that for any fixed finite $0 \leq c < \infty$:

$$\lim_{\mu \rightarrow \infty} \int_c^\infty f(t, \mu) dt = \lim_{\mu \rightarrow \infty} S_\mu(c) = 1 \quad (\text{A1})$$

Without loss of generality, we can set $\mu_2 = 0$ under H_0 and write the Null Rejection Probability (NRP) as a function of μ_1 :

$$\begin{aligned} NRP(\mu_1) &= P(\text{Reject} \mid \mu_1) = \int_{cv_1^*}^\infty \int_{cv_2(|t_1|)}^\infty f(|t_1|; \mu_1) f(|t_2|) dt_2 dt_1 \\ &= \int_{cv_1^*}^\infty f(|t_1|; \mu_1) \int_{cv_2(|t_1|)}^\infty f(|t_2|) dt_2 dt_1 \\ &= \int_{cv_1^*}^\infty f(|t_1|; \mu_1) S_0(cv_2(|t_1|)) dt_1. \end{aligned}$$

Consider the two cases where cv_2^* is below or above $z_{\alpha/2}$

(a) In case $cv_2^* < z_{\alpha/2}$ there will exist a $\varepsilon > 0$ such that $cv_2^* = z_{\alpha/2} - \varepsilon$. The survival function is monotonically decreasing so $S_0(cv_2^*) = S_0(z_{\alpha/2} - \varepsilon) > \alpha$. Hence:

$$NRP(\mu) = \int_{cv_1^*}^\infty f(|t_1|; \mu) S_0(cv_2(|t_1|)) dt_1 > \alpha \int_{cv_1^*}^\infty f(|t_1|; \mu) dt_1,$$

The last term converges to 1 when $\mu \rightarrow \infty$ by Equation (A1), and we have $\lim_{\mu \rightarrow \infty} NRP(\mu) > \alpha$. The size condition is, therefore, violated.

- (b) In case $cv_2^* > z_{\alpha/2}$ there will exist a $\varepsilon > 0$ such that $cv_2^* = z_{\alpha/2} + \varepsilon$ and $\lim_{\mu \rightarrow \infty} NRP(\mu) < \alpha$. Choosing $cv_2(|t_1|) = z_{\alpha/2}$ (the LR test with size α) has $\lim_{\lambda \rightarrow \infty} NRP(\mu) = \alpha$ and dominates any test with $cv_2^* > z_{\alpha/2}$ in terms of power. The same arguments hold for cv_1^* , so, for a size α test to be admissible, it must satisfy $cv_1^* = cv_2^* = z_{\alpha/2}$. \square

Lemma A1. Given four corner points of any regular square in \mathbb{R}_+^2 ,

$$t^{(1)} = (t_1, t_2), t^{(2)} = (t_1, t_2 + \varepsilon), t^{(3)} = (t_1 + \varepsilon, t_2 + \varepsilon), t^{(4)} = (t_1 + \varepsilon, t_2),$$

for any real $\varepsilon > 0$, at least one corner is not on the boundary.

Proof. Since $|t|^{(1)}$ is the bottom left corner and $|t|^{(3)}$ the top right, $|t|^{(3)} \in Q^o(|t|^{(1)})$. Hence, if $|t|^{(1)}$ is on the boundary, then $|t|^{(3)} \in CR \setminus \partial CR$ and not all four points can be on the boundary.

The number of points could be zero if the square is entirely inside AR or CR . There could be two points on the boundary: $|t|^{(1)}$ and $|t|^{(2)}$ (vertical), $|t|^{(1)}$ and $|t|^{(4)}$ (horizontal), or $|t|^{(2)}$ and $|t|^{(4)}$ (strictly decreasing). There could be three points on the boundary if it contains the path from $|t|^{(2)}$ vertically down via $|t|^{(1)}$ horizontally to $|t|^{(4)}$. \square

Proof of Proposition 3. In order to show that Sobel and LR tests are information coherent, we prove that the tests are (weakly) increasing in $|t_1|$ and $|t_2|$. The CR for the Sobel test is:

$$CR_{\sqrt{W}}(\alpha) = \left\{ |t| \in \mathbb{R}_+^2 \mid \sqrt{W} \equiv \frac{|t_1||t_2|}{\sqrt{|t_1|^2 + |t_2|^2}} > z_{\alpha/2} \right\}$$

The partial derivative $\frac{\partial \sqrt{W}}{\partial |t_1|} = |t_2|^3 \cdot (|t_1|^2 + |t_2|^2)^{-3/2} > 0$ since $|t_2| > 0$ a.s. So \sqrt{W} is increasing in $|t_1|$ and by symmetry \sqrt{W} is increasing in $|t_2|$ for any $|t| \in \mathbb{R}_+^2$. Hence, if $\sqrt{W} > z_{\alpha/2}$ for $|t|$, then the test will continue to reject for $|\tilde{t}|$ with one of its elements larger than $|t|$ since $\sqrt{\tilde{W}} > \sqrt{W} > z_{\alpha/2}$. This shows information coherence for fixed α , and α coherence when α is varying since $CR_{\sqrt{W}}(\alpha) \subset CR_{\sqrt{W}}(\tilde{\alpha})$ if $\alpha < \tilde{\alpha} \Leftrightarrow z_{\alpha/2} > z_{\tilde{\alpha}/2}$.

The CR for the LR test is:

$$CR_{LR}(\alpha) = \left\{ |t| \in \mathbb{R}_+^2 \mid LR \equiv \min\{|t_1|, |t_2|\} > z_{\alpha/2} \right\}$$

LR is never decreasing in $|t_1|$ or $|t_2|$. That $z_{\alpha/2}$ is the correct critical value in both cases follows from the fact that $W \rightarrow |t_2|$ as $|t_1| \rightarrow \infty$ (and $|t_1|$ as $|t_2| \rightarrow \infty$) and $LR = |t_2|$ as $|t_1| \rightarrow \infty$ (and $|t_1|$ as $|t_2| \rightarrow \infty$). The marginal distribution of $|t_1|$ (or $|t_2|$) is the folded standard normal distribution with critical value $z_{\alpha/2}$. \square

Alternative Proof of Main Theorem 1. The boundary $\partial CR_{LR}(\alpha)$ is defined by the quarter $Q((z_{\alpha/2}, z_{\alpha/2})')$. The corner $(z_{\alpha/2}, z_{\alpha/2})'$ contains the smallest values for $|t_1|$ and $|t_2|$ such that the NRP does not exceed α , even as $(\mu_1, \mu_2) \rightarrow (0, \infty)$ or $(\mu_1, \mu_2) \rightarrow (\infty, 0)$ and $cv_1(|t_2|) = z_{\alpha/2} = cv_2(|t_1|)$ for all $|t_1|, |t_2|$. Any other $CR \in \mathbb{C}_\alpha$, $CR \subset CR_{LR}(\alpha)$ since it could not satisfy the size condition if $cv_1^* < z_{\alpha/2}$ by Proposition 3 (vi). Hence, $P[CR] < P[CR_{LR}(\alpha)]$ unless $CR = CR_{LR}$ a.s. This holds under the null as well as under the alternative, so $CR_{LR}(\alpha)$ is at least as powerful as any other test in \mathbb{C}_α . The LR test of size α is the most powerful test in \mathbb{C}_α . \square

References

- Baron, R.M.; Kenny, D.A. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *J. Personal. Soc. Psychol.* **1986**, *51*, 1173. [CrossRef]
- Coletti, A.L.; Sedatole, K.L.; Towry, K.L. The effect of control systems on trust and cooperation in collaborative environments. *Account. Rev.* **2005**, *80*, 477–500. [CrossRef]

3. MacKenzie, S.B.; Lutz, R.J.; Belch, G.E. The role of attitude toward the ad as a mediator of advertising effectiveness: A test of competing explanations. *J. Mark. Res.* **1986**, *23*, 130–143. [[CrossRef](#)]
4. Alwin, D.F.; Hauser, R.M. The decomposition of effects in path analysis. *Am. Sociol. Rev.* **1975**, *40*, 37–47. [[CrossRef](#)]
5. Huber, M. Mediation analysis. In *Handbook of Labor, Human Resources and Population Economics*; Springer Nature: Cham, Switzerland, 2020; pp. 1–38.
6. MacKinnon, D.P.; Lockwood, C.M.; Hoffman, J.M.; West, S.G.; Sheets, V. A comparison of methods to test mediation and other intervening variable effects. *Psychol. Methods* **2002**, *7*, 83. [[CrossRef](#)] [[PubMed](#)]
7. Drton, M. Likelihood ratio tests and singularities. *Ann. Stat.* **2009**, *37*, 979–1012. [[CrossRef](#)]
8. Sobel, M.E. Asymptotic confidence intervals for indirect effects in structural equation models. *Soc. Meth.* **1982**, *13*, 290–312. [[CrossRef](#)]
9. Van Garderen, K.J. Curved exponential models in econometrics. *Econom. Theory* **1997**, *13*, 771–790. [[CrossRef](#)]
10. Hillier, G.H.; Van Garderen, K.J.; Van Giersbergen, N.P.A. Improved Tests for Mediation. Working Paper CWP01/22 Centre for Microdata Methods and Practice. 2022. Available online: <http://dx.medra.org/10.47004/wp.cem.2022.0122> (accessed on 1 December 2021).
11. Lehmann, E.L.; Romano, J.P. *Testing Statistical Hypotheses*, 3rd ed.; Springer Science & Business Media: New York, NY, USA, 2005.
12. Vaughan, R.J.; Venables, W.N. Permanent expressions for order statistic densities. *J. R. Stat. Soc. Ser. B (Methodol.)* **1972**, *34*, 308–310. [[CrossRef](#)]
13. Perlman, M.D.; Wu, L. The emperor's new tests. *Stat. Sci.* **1999**, *14*, 355–369. [[CrossRef](#)]
14. Van Garderen, K.J.; Van Giersbergen, N.P.A. A Nearly Similar Powerful Test for Mediation. Mimeo, University of Amsterdam. 2021. Available online: <https://arxiv.org/abs/2012.11342> (accessed on 1 December 2021).