UvA ECONOMETRICS

Discussion Paper: 2014/10

Inference about the Indirect Effect: a Likelihood Approach

Noud P.A. van Giersbergen

www.ase.uva.nl/uva-econometrics

Amsterdam School of Economics
Department of Economics & Econometrics
Valckenierstraat 65-67
1018 XE AMSTERDAM
The Netherlands

UvA UNIVERSITEIT VAN AMSTERDAM

Inference about the Indirect Effect: a Likelihood Approach

Noud P.A. van Giersbergen*
Email: N.P.A.vanGiersbergen@UvA.nl
UvA-Econometrics, Amsterdam School of Economics
Department of Econometrics, University of Amsterdam,
Valckenierstraat 65-67, 1018 XE, Amsterdam, The Netherlands

December 30, 2014

Abstract

Prior research for constructing confidence intervals for an indirect effect has focused on a Wald statistic. In this paper, however, the inference problem is analyzed from a likelihood ratio (LR) perspective. When testing the null hypothesis $H_0 : \alpha\beta = 0$, the LR test statistic leads to the minimum of two *t-ratios*, whose size can be controlled. A confidence interval is obtained by inverting the LR statistic. Another confidence interval is obtained by inverting the sum of two pivotal *t-statistics*. In the Monte Carlo simulations, this latter confidence interval is the best performer: it outperforms the commonly used existing methods.

1 Introduction

An indirect effect implies a causal relation between an independent variable, a mediating variable and a dependent variable; see Sobel (1990) for a detailed analysis. Indirect effects are important in sociology, psychology and economics. Recently, the statistical properties of estimators of indirect effects have received much attention, especially by MacKinnon and coauthors. Key papers in this field are MacKinnon et al. (2002) and MacKinnon et al. (2004). The latter paper concludes that confidence intervals for the indirect effect based on the distribution of products are the most accurate and they perform even better than bootstrap intervals. These confidence intervals are based on critical values reported by Meeker et al. (1981).

*Helpful comments from Kees Jan van Garderen and Hans van Ophem are gratefully acknowledged.

The indirect effect can be written as a product of two parameters: $\alpha\beta$ (to be defined later). Research has concentrated on the following Wald statistic:

$$\frac{\hat{\alpha}\hat{\beta} - \alpha\beta}{\sqrt{V(\hat{\alpha}\hat{\beta})}}.$$

All kinds of different variance estimators have been proposed, although the simple formula based on the multivariate delta method derived by Sobel (1982) seems to work best:

$$V(\hat{\alpha}\hat{\beta}) = \hat{\alpha}^2 V(\hat{\beta}) + \hat{\beta}^2 V(\hat{\alpha}). \tag{1}$$

In contrast to the existing literature, a likelihood based approach will be investigated in this paper. However, since the parameter of interest is a product, the analysis will not be standard. We shall focus on the likelihood ratio (LR) statistic. For testing $\alpha\beta = 0$, the LR test statistic leads to an exact test, i.e. there exists a critical value such that the size is never greater than the nominal significance level in finite samples. However, the LR test is conservative, just like the Wald statistic. Besides size-considerations, the LR test statistic is the most powerful test due to the Neyman-Pearson lemma in the simple hypothesis setup. Although our analysis concerns the composite case, it is hoped that the LR statistic is still more powerful than the Wald statistic.

The paper is organized as follows. Section 2 introduces the model and its likelihood. In Section 3, the LR Statistic for testing $H_0 : \alpha\beta = 0$ is analyzed. The construction of confidence intervals is discussed in Section 4. The fifth section presents some simulation results. Finally, the main conclusions are summarized in the last section.

2 The Model and Likelihood

We consider the following data generating process (DGP):

$$\begin{aligned} y &= \delta x + \beta_0 + \beta x_m + \varepsilon = X \begin{bmatrix} \delta \\ \beta_0 \end{bmatrix} + \beta x_m + \varepsilon, & (2) \\ x_m &= \alpha x + \alpha_0 + \xi = X \begin{bmatrix} \alpha \\ \alpha_0 \end{bmatrix} + \xi, & (3) \end{aligned}$$

where $X = [x : \imath]$ is a $n \times 2$ matrix, y, x, x_m, ε and ξ are $n \times 1$ vectors. For the likelihood function, normality is assumed: $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ is independent of $\xi_i \sim N(0, \sigma_\xi^2)$. It turns out that the LR approach delivers a test statistic whose distribution does not dependent too much on the normality assumption. Although the structural equation in (2) does not include additional exogenous variables, they can easily be included in the equation like

$$y = \delta x + \beta_0 + \beta x_m + Z\lambda + \varepsilon,$$

where Z denotes a $n \times m$ matrix of exogenous variables. However, equation (2) results if the additional exogenous regressors Z have been partialled out and the degrees of freedom are modified as needed. The same reasoning can be applied to equation (3).

The regression equations shown in (2) and (3) can be derived as the marginal model of x_m given x and the conditional model of y given x and x_m, when the tripled (y, x_m, x) is assumed to be normally distributed. In econometrics, regression (2) is also referred to as the structural model, while regression (3) is known as the reduced model, although usually ε is then assumed to be correlated with ξ. Note that when ε is correlated with ξ, OLS estimators are not consistent anymore and some kind of instrumental variables (IV) estimator has to be used for the structural model. The coefficient δ denotes the direct effect of x on y (i.e. the partial effect of x on y holding the mediating variable x_m constant). Substituting formula (3) into (2) gives

$$\begin{aligned} y &= X \begin{bmatrix} \delta + \alpha\beta \\ \beta_0 + \alpha_0\beta \end{bmatrix} + \beta\xi + \varepsilon, \\ &= X \begin{bmatrix} \tau \\ \tau_0 \end{bmatrix} + \varepsilon^+, \qquad \varepsilon^+ = \beta\xi + \varepsilon. \end{aligned} \tag{4}$$

The total effect is given by τ, while $\alpha\beta$ is the so-called indirect effect. Since model (4) wrongly omits the regressor x_m when $\beta \neq 0$, the disturbance term ε^+ is correlated with ξ in this case.

Defining $\theta = (\delta, \beta_0, \beta)'$ and $\psi = (\alpha, \alpha_0)'$, the likelihood is given by

$$\begin{aligned} L(\theta, \psi, \sigma_\varepsilon^2, \sigma_\xi^2) &= f(y|x, x_m, \delta, \beta_0, \beta)\, f(x_m|x, \alpha, \alpha_0) \\ &= \left(2\pi\sigma_\varepsilon^2\right)^{-\frac{n}{2}} \exp(-\frac{1}{2\sigma_\varepsilon^2} \sum \varepsilon_i^2) \left(2\pi\sigma_\xi^2\right)^{-\frac{n}{2}} \exp(-\frac{1}{2\sigma_\xi^2} \sum \xi_i^2). \end{aligned}$$

Since there are no restrictions between the parameters of the structural and reduced form equation and ε_i is independent of ξ_i, the likelihood is just the product of the likelihood of the structural and reduced form equation. The ML estimators of θ and ψ are the OLS estimators.

This leads to the following concentrated log-likelihood function (concentrated with respect to θ and ψ)

$$l_c \propto -\frac{n}{2}\log(\hat{\sigma}_\varepsilon^2) - \frac{1}{2\hat{\sigma}_\varepsilon^2}\hat{\varepsilon}'\hat{\varepsilon} - \frac{n}{2}\log(\hat{\sigma}_\xi^2) - \frac{1}{2\hat{\sigma}_\xi^2}\hat{\xi}'\hat{\xi}.$$

The ML estimators of σ_ε^2 and σ_ξ^2 are respectively

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n}\hat{\varepsilon}'\hat{\varepsilon} \qquad \text{and} \qquad \hat{\sigma}_\xi^2 = \frac{1}{n}\hat{\xi}'\hat{\xi}.$$

Hence, the unrestricted (u) maximized likelihood is

$$L_u = \max L(\hat{\theta}, \hat{\psi}, \hat{\sigma}_\varepsilon^2, \hat{\sigma}_\xi^2) = \left(\frac{2\pi}{n}\right)^{-n} \exp(-n) SSR_\varepsilon^{-n/2} SSR_\xi^{-n/2}.$$

where $SSR_\varepsilon = \hat{\varepsilon}'\hat{\varepsilon}$ and $SSR_\xi = \hat{\xi}'\hat{\xi}$.

3 The LR statistic for $\alpha\beta = 0$

The likelihood ratio test compares the unrestricted maximized likelihood with the maximized likelihood subject to the restriction $H_0 : \alpha\beta = 0$. Maximizing the likelihood under the null hypothesis can be done in two distinct steps: (i) first, maximize subject to $\alpha = 0$ and (ii) secondly, maximize subject to $\beta = 0$. The maximized concentrated log-likelihood function under $\alpha = 0$ is given by

$$l_c(\alpha = 0) \propto -\frac{n}{2}\log(\hat{\sigma}_\varepsilon^2) - \frac{1}{2\hat{\sigma}_\varepsilon^2}\hat{\varepsilon}'\hat{\varepsilon} - \frac{n}{2}\log(\tilde{\sigma}_\xi^2) - \frac{1}{2\tilde{\sigma}_\xi^2}\tilde{\xi}'\tilde{\xi},$$

where $\tilde{\xi} = x_m - \tilde{\alpha}_0$ are the restricted residuals. This leads to the restricted estimator of $\tilde{\sigma}_\xi^2 = n^{-1} SSR_{\alpha=0}$ with $SSR_{\alpha=0} = \tilde{\xi}'\tilde{\xi}$, so that the maximized likelihood under $\alpha = 0$ is given by

$$L_{\alpha=0} = \max_{\text{s.t. } \alpha=0} L(\hat{\theta}, \tilde{\psi}, \hat{\sigma}_\varepsilon^2, \tilde{\sigma}_\xi^2) = \left(\frac{2\pi}{n}\right)^{-n} \exp(-n) SSR_\varepsilon^{-n/2} SSR_{\alpha=0}^{-n/2}.$$

Analogously, we obtain under $\beta = 0$

$$L_{\beta=0} = \max_{\text{s.t. } \beta=0} L(\tilde{\theta}, \hat{\psi}, \tilde{\sigma}_\varepsilon^2, \hat{\sigma}_\xi^2) = \left(\frac{2\pi}{n}\right)^{-n} \exp(-n) SSR_{\beta=0}^{-n/2} SSR_\xi^{-n/2},$$

where $SSR_{\beta=0} = \tilde{\varepsilon}'\tilde{\varepsilon}$ with $\tilde{\varepsilon} = y - \tilde{\delta}x - \tilde{\beta}_0$. The likelihood ratio becomes

$$\begin{aligned} \lambda &= \frac{L_u}{\max\{L_{\alpha=0}, L_{\beta=0}\}} = \min\left\{\frac{L_u}{L_{\alpha=0}}, \frac{L_u}{L_{\beta=0}}\right\} \\ &= \min\left\{\left(\frac{SSR_\xi}{SSR_{\alpha=0}}\right)^{-n/2}, \left(\frac{SSR_\varepsilon}{SSR_{\beta=0}}\right)^{-n/2}\right\} = \min(\lambda_\alpha, \lambda_\beta). \end{aligned}$$

As usual, the F-ratio is a monotone transformation of the likelihood ratio:

$$f_{\min} = \min\left\{\frac{n-2}{1}(\lambda_\alpha^{n/2}-1), \frac{n-3}{1}(\lambda_\beta^{n/2}-1)\right\}.$$

Since we are only testing one parameter at a time, this is also equivalent to the minimum of the t-statistics for $\alpha = 0$ and $\beta = 0$:

$$t^2_{\min} = \min\left\{t^2_\alpha = \frac{\hat{\alpha}^2}{SE(\hat{\alpha})^2}, t^2_\beta = \frac{\hat{\beta}^2}{SE(\hat{\beta})^2}\right\}. \tag{5}$$

For a critical region, we look for a critical value such that

$$\max_{\alpha\beta=0} P[t^2_{\min} > c] \leq \omega,$$

where ω denotes the nominal significance level. When $\alpha = 0$, $\beta \in \mathbb{R}$, we have

$$t_\alpha \sim T_{n-2} \qquad \text{and} \qquad t_\beta \sim T_{n-3}(\beta/\sigma_{\hat{\beta}}),$$

so t_β follows a non-central T-distribution with $n-3$ degrees of freedom and non-centrality parameter $\beta/\sigma_{\hat{\beta}}$. When $|\beta|$ becomes large, $|t_\beta|$ also gets large, so that $\min\{t^2_\alpha, t^2_\beta\} \approx t^2_\alpha$ under $\alpha = 0, \beta \in \mathbb{R}$. For $|\beta| \to \infty$, the α-upper quantile of the $t^2_{n-2}(= F_{1,n-1})$-distribution can be taken as critical value. Analogously, we get for $\beta = 0, \alpha \in \mathbb{R}$:

$$t_\alpha \sim T_{n-2}(\alpha/\sigma_{\hat{\alpha}}) \qquad \text{and} \qquad t_\beta \sim T_{n-3}.$$

Now, for $|\alpha| \to \infty$, we can take the α-upper quantile of the $t^2_{n-3}(= F_{1,n-3})$-distribution as critical value. Since the critical value for $n-3$ is larger than for $n-2$, the quantile should be taken from the $F_{1,n-3}$-distribution or equivalently $|t_{n-3}|$-distribution.

Figure 1 shows the 95%-quantile of $\min\{t^2_\alpha, t^2_\beta\}$ for $n = 50$ based on Monte Carlo simulation (100,000 replications). In the upper figure, $\beta = 0$ and α varies. The critical value is lowest when $\alpha = 0$ and it converges to $F_{5\%}(1, 48) = 4.0427$ when $|\beta|$ becomes large. In the lower figure, $\alpha = 0$ and β varies. For t^2_β, the same behavior is observed: the critical value becomes smaller as $|\beta|$ becomes smaller.

Insert Figure 1 around here.

4 Confidence Interval for $\alpha\beta$

Most of the literature has focused on setting up confidence intervals. One way to determine a confidence interval is by inverting a test statistic, so a confidence interval contains all the values γ such that $H_0 : \alpha\beta = \gamma$ cannot be rejected. Contrary to the previous section, we cannot assume that one of the parameters is zero. Hence, we have to maximize the log-likelihood under the restriction $\alpha\beta = \gamma$. Recall that $\xi = x_m - \alpha x - \alpha_0$, so the log-likelihood becomes under the restriction $\alpha = \gamma/\beta$:

$$l(\theta, \alpha_0, \sigma_\varepsilon^2, \sigma_\xi^2) \propto -\frac{n}{2}\log(\sigma_\varepsilon^2) - \frac{1}{2\sigma_\varepsilon^2}\sum \varepsilon_i^2 - \frac{n}{2}\log\left(\sigma_\xi^2\right) - \frac{1}{2\sigma_\xi^2}\sum (x_{m,i} - \frac{\gamma}{\beta}x_i - \alpha_0)^2.$$

Now, the log-likelihood of the reduced model is connected to the structural model. The first-order conditions (FOC) with respect to β and σ_ξ^2 change to:

$$\begin{aligned} \frac{\partial l}{\partial \beta} &= \frac{1}{\tilde{\sigma}_\varepsilon^2}\sum (y_i - \tilde{\delta}x_i - \tilde{\beta}_0 - \tilde{\beta}x_{m,i})x_{m,i} - \frac{1}{\tilde{\sigma}_\xi^2}\sum (x_{m,i} - \frac{\gamma}{\tilde{\beta}}x_i - \tilde{\alpha}_0)\frac{\gamma}{\tilde{\beta}^2}x_i = 0, \quad (6) \\ \frac{\partial l}{\partial \sigma_\xi^2} &= -\frac{n}{2\tilde{\sigma}_\xi^2} + \frac{1}{2\tilde{\sigma}_\xi^4}\sum (x_{m,i} - \frac{\gamma}{\tilde{\beta}}x_i - \tilde{\alpha}_0)^2 = 0. \end{aligned}$$

From the other FOCs, we can express all other estimators as a function of $(\tilde{\beta}, \gamma)$ and the data. For the regression coefficients, we get

$$\begin{aligned} \tilde{\delta} &= \frac{S_x S_y - nS_{x,y} + (nS_{x,xm} - S_x S_{xm})\tilde{\beta}}{S_x^2 - nS_{x,x}}, \\ \tilde{\beta}_0 &= \frac{S_x S_{x,y} - S_{x,x} S_y + (S_{xm} S_{x,x} - S_x S_{x,xm})\tilde{\beta}}{S_x^2 - nS_{x,x}}, \\ \tilde{\alpha}_0 &= \frac{S_{xm}\tilde{\beta} - \gamma S_x}{n\tilde{\beta}}, \end{aligned}$$

where $S_v = \sum_{i=1}^n v_j$ for $v \in \{y, xm, x\}$ denotes the sum of a variable and $S_{v,w} = \sum_{i=1}^n v_i w_i$ for $v, w \in \{y, xm, x\}$ denotes the cross-product of two variables. Using these expressions and the FOCs for $\tilde{\sigma}_\varepsilon^2$ and $\tilde{\sigma}_\xi^2$, we can get a FOC for β in $(\tilde{\beta}, \gamma)$ and the data only:

$$\frac{\gamma^2(nS_{x,x} - S_x^2) + N11\tilde{\beta}}{\gamma^2(nS_{x,x} - S_x^2)\tilde{\beta} + D11\tilde{\beta}^2 + D12\tilde{\beta}^3} + \frac{N20 + N21\tilde{\beta}}{D20 + D21\tilde{\beta} + D22\tilde{\beta}^2} = 0,$$

where

$$
\begin{aligned}
N11 &= \gamma\,(S_x S_{xm} - nS_{x,xm}),\\
D11 &= 2\gamma\,(S_x S_{xm} - nS_{x,xm}), \qquad D12 = (nS_{xm,xm} - S_{xm}^2),\\
N20 &= S_x^2 S_{xm,y} - nS_{xm,y}S_{xx} - S_x S_{xm} S_{x,y} + nS_{x,xm}S_{x,y} + S_{xm}S_{x,x}S_y - S_x S_{x,xm}S_y,\\
N21 &= nS_{xm,xm}S_{x,x} - S_x^2 S_{xm,xm} - S_{xm}^2 S_{xx} + 2S_x S_{xm}S_{x,xm} - nS_{x,xm}^2,\\
D20 &= nS_{x,y}^2 - 2S_x S_{x,y}S_y + S_{x,x}S_y^2 + S_x^2 S_{y,y} - nS_{x,x}S_{y,y},\\
D21 &= 2(nS_{xm,y}S_{x,x} - S_x^2 S_{xm,y} + S_x S_{xm}S_{x,y} - nS_{x,xm}S_{x,y} - S_{xm}S_{x,x}S_y + S_x S_{x,xm}S_y),\\
D22 &= Sx^2 S_{xm,xm} + S_{xm}^2 S_{x,x} - nS_{xm,xm}S_{x,x} - 2S_x S_{xm}S_{x,xm} + nS_{x,xm}^2.
\end{aligned}
$$

Taking the two fractions together, we see that the roots of the FOC condition yields a quartic polynomial in $\tilde{\beta}$, with a zero coefficient for $\tilde{\beta}^2$. The advantage of having a closed form polynomial is that there exist fast and accurate procedures for finding *all* the roots. Once the roots are determined, the $\tilde{\beta}$ with the highest log-likelihood value is the ML estimator. Note that commonly used optimization procedures have difficulties with determining a global optimum. A $100(1-\omega)\%$ confidence region is obtained by determining all γ values for which $H_0 : \alpha\beta = \gamma$ is not rejected at significance level ω, i.e.

$$
CI(\gamma)_{LR}^{Asymp} = \{\gamma\,:\, LR(\gamma) \le \chi_1^2(1-\omega)\}. \tag{7}
$$

This set can be accurately approximated by some search technique, for instance grid search or bisection. Note that there is no guarantee that the region is indeed an interval, i.e. the region could consist of a union of disjoint intervals.

Next, we shall derive an interval that can be determined by just finding roots of a polynomial. This interval is based on the following idea: a $100(1-\omega)\%$ confidence region for the mean of a p-dimensional $N_p(\mu, \Sigma)$ is the ellipsoid determined by all μ such that

$$
n(\bar{u}-\mu)'S^{-1}(\bar{u}-\mu) \le \frac{p(n-1)}{n-p}F_{p,n-p}(\omega),
$$

where $\bar{u} = n^{-1}\sum_{i=1}^n u_i$ and $S = (n-1)^{-1}\sum_{i=1}^n (u_i-\bar{u})(u_i-\bar{u})'$. When S is diagonal, this boils down to the sum of squared t-statistics. In our case, we have

$$
\tau_\alpha(\alpha)^2 + \tau_\beta(\beta)^2,
$$

where $\tau_\alpha(\alpha) = (\hat{\alpha}-\alpha)/SE(\hat{\alpha})$ and $\tau_\beta(\beta) = (\hat{\beta}-\beta)/SE(\hat{\beta})$ are two pivotal statistics; here τ is used since the statistic also depends upon an unknown parameter. However, a confidence interval for γ is required, which is only one parameter. So, instead of taking the critical values from a χ^2-distribution with two degrees of freedom in case of (α, β), we now take the quantiles from a χ^2-distribution with only one degree of freedom. This leads to the following $100(1-\omega)\%$ asymptotic confidence interval:

$$CI(\gamma)_{t^2}^{Asymp} = \{\gamma : \tau_\alpha(\alpha)^2 + \tau_\beta(\beta)^2 \leq \chi_1^2(1-\omega)\}. \tag{8}$$

It should be noted that the intervals defined by (7) and (8) are not numerically identical, i.e. in finite samples the two intervals are numerically slightly different.

Note that $\tau_\alpha(\alpha)^2 + \tau_\beta(\beta)^2 \leq q$ defines an elliptical confidence region for (α, β). The extreme values for β are obtained for $\alpha = \tilde{\alpha}$, while the extreme values for α are obtained for $\beta = \tilde{\beta}$. To compare the confidence region approach with the test statistic approach, we look at the event that the confidence region contains the value zero, i.e. $\alpha\beta = 0$. When $\alpha = \tilde{\alpha}$, the t-statistic for α equals zero, i.e. $\tau_\alpha(\tilde{\alpha}) = 0$, and the set defined by $\{\beta : \tau_\beta(\beta)^2 \leq q_\beta\}$ contains zero if $\tilde{\beta}/SE(\tilde{\beta}) \leq \sqrt{q_\beta}$. The probability of this inequality is highest when $\beta = 0$, so that $q_\beta = F_{1,n-3}(\omega)$. Analogously for α, the set defined by $\{\alpha : \tau_\alpha(\alpha)^2 \leq q_\alpha\}$ contains zero if $\tilde{\alpha}/SE(\tilde{\alpha}) \leq \sqrt{q_\alpha}$ when $\beta = \tilde{\beta}$. This leads to $q_\alpha = F_{1,n-2}(\omega)$. Combining these two results, we see that zero is included in the interval if

$$\max\{t_\alpha^2 = \tau_\alpha(0)^2, t_\beta^2 = \tau_\beta(\beta)^2\} \leq F_{1,n-3}(\omega),$$

since $F_{1,n-3}(\omega) > F_{1,n-2}(\omega)$. Note that this is the dual of the test statistic given in (5). This result suggests a better finite-sample performance by replacing the critical value based on the χ^2-distribution with the appropriate quantile of the $F_{1,n-3}$-distribution:

$$CI(\gamma)_{t^2}^{F} = \{\gamma : \tau_\alpha(\alpha)^2 + \tau_\beta(\beta)^2 \leq F_{1,n-3}(\omega)\}. \tag{9}$$

A confidence interval based on $\tau_\alpha(\alpha)^2 + \tau_\beta(\beta)^2 \leq q$ has the advantage that no search technique has to be used for finding the lower and upper confidence limit. The equation

$$\frac{(\tilde{\alpha}-\alpha)^2}{V(\tilde{\alpha})} + \frac{(\tilde{\beta}-\beta)^2}{V(\tilde{\beta})} = q, \tag{10}$$

can be solved with respect to β:

$$\beta_\pm^* = \frac{\tilde{\beta}V(\tilde{\alpha}) \pm \sqrt{V(\tilde{\alpha})V(\tilde{\beta})(qV(\tilde{\alpha}) - (\tilde{\alpha}-\alpha)^2}}{V(\tilde{\alpha})}. \tag{11}$$

The equation defined by the $+$-sign describes the upper contour of the ellipse, while the '$-$'-equation describes the lower contour. The upper confidence limit for γ is defined as

$$\max \alpha\beta \quad \text{s.t. equation (10).}$$

We can substitute equation (11) into $\alpha\beta$ to get a formula in α and estimates only. The optimum is found by solving the FOC for α, i.e. set the derivative of $\alpha\beta^*_\pm$ to zero. For the β^*_+-solution, we obtain after reducing fractions to the same denominator:

$$\frac{V(\tilde{\beta})\{qV(\tilde{\alpha}) - (\tilde{\alpha} - 2\alpha)(\tilde{\alpha} - \alpha)\} + \tilde{\beta}\sqrt{V(\tilde{\alpha})V(\tilde{\beta})(qV(\tilde{\alpha}) - (\tilde{\alpha} - \alpha)^2}}{\sqrt{V(\tilde{\alpha})V(\tilde{\beta})(qV(\tilde{\alpha}) - (\tilde{\alpha} - \alpha)^2}} = 0. \tag{12}$$

This fraction is zero if the numerator is zero and the denominator different from zero. The numerator consists of the sum $h_1(\alpha)+h_2(\alpha)$. Solving $h_1(\alpha)+h_2(\alpha) = 0$, implies calculating the roots of the following polynomial:

$$C_4\alpha^4 + C_3\alpha^3 + C_2\alpha^2 + C_1\alpha + C_0 = 0, \tag{13}$$

where

$$\begin{aligned}
C_0 &= \tilde{\alpha}^2\tilde{\beta}^2V(\tilde{\alpha})V(\tilde{\beta}) - q\tilde{\beta}^2V(\tilde{\alpha})^2V(\tilde{\beta}) + \tilde{\alpha}^4V(\tilde{\beta})^2 - 2q\tilde{\alpha}^2V(\tilde{\alpha})V(\tilde{\beta})^2 + q^2V(\tilde{\alpha})^2V(\tilde{\beta})^2,\\
C_1 &= -2\tilde{\alpha}\tilde{\beta}^2V(\tilde{\alpha})V(\tilde{\beta}) - 6\tilde{\alpha}^3V(\tilde{\beta})^2 + 6q\tilde{\alpha}V(\tilde{\alpha})V(\tilde{\beta})^2,\\
C_2 &= \tilde{\beta}^2V(\tilde{\alpha})V(\tilde{\beta}) + 13\tilde{\alpha}^2V(\tilde{\beta})^2 - 4qV(\tilde{\alpha})V(\tilde{\beta})^2,\\
C_3 &= -12\tilde{\alpha}V(\tilde{\beta})^2,\\
C_4 &= 4V(\tilde{\beta})^2.
\end{aligned}$$

We get the same equation for the β^*_--part (only multiplied by -1). In general there are 4 roots, although some may be complex. If the roots are calculated, the corresponding β-values are easily found using equation (11). The upper and lower confidence limit is obtained by finding the largest and smallest value for the set $\{\alpha_i\beta_i\}$, where α_i enumerates all the real roots of equation (13); see Figure 2 for an illustration

In Figure 2, the results of inverting the LR statistic is shown for two different samples. The blue line shows the trajectory of $(\tilde{\alpha}, \gamma/\tilde{\alpha})$ when varying γ such that the LR statistic does not reject $H_0 : \alpha\beta = \gamma$. The ellipses are the boundaries of two sets (α, β) such that $\tau_\alpha(\alpha)^2 + \tau_\beta(\beta)^2 \leq q$. For the upper figure, we have $\tilde{\alpha}$ and $\tilde{\beta}$ significantly different from zero and both confidence limits are positive. Furthermore, there are only two real roots for

α, so there are four corresponding β-values indicate by circles. In the lower figure, $\tilde{\alpha}$ and $\tilde{\beta}$ are both not significantly different from zero. There are four real roots for α, which leads to eight possible (α, β)-values. The upper limit is positive, while the lower limit is negative.

Insert Figure 2 around here.

5 Monte Carlo Results

Since the LR approach has only an asymptotic justification, the finite-sample behavior is investigated through Monte Carlo simulation. All simulations were carries out by MATLAB, although a connection with R was established using statconnDCOM to obtain the Meeker critical values using the RMediation package; see Tofighi and MacKinnon (2011).

Confidence limits based on a Wald statistic are given by

$$\left(\hat{\alpha}\hat{\beta} - q(1-\omega/2)SE(\hat{\alpha}\hat{\beta}), \hat{\alpha}\hat{\beta} - q(\omega/2)SE(\hat{\alpha}\hat{\beta})\right), \tag{14}$$

where $q(\omega)$ denote the ω-quantile of some distribution. Wald-based intervals in this paper use the Sobel standard error as shown in formula (1). The normal approximation uses the quantiles of the standard normal distribution, i.e. $q(\omega) = \Phi^{-1}(\omega)$ where $\Phi(\cdot)$ denotes the cdf of the standard normal distribution. The Meeker interval uses the same standard errors, but other quantiles:

$$q(\omega) = \frac{q_{\hat{\alpha}\hat{\beta}}(\omega) - t_\alpha t_\beta}{\sqrt{t_\alpha^2 + t_\beta^2 + 1}},$$

where $q_{\hat{\alpha}\hat{\beta}}(\omega)$ is the ω-quantile of the product of two normally distributed random variables. Note that the Meeker interval can be asymmetric. For the intervals based on inversion, we consider the interval based on inverting the LR statistic and the asymptotic χ^2-distribution given in formula (7). Furthermore, we consider the interval based on $\{\gamma : \tau_\alpha(\alpha)^2 + \tau_\beta(\beta)^2 \leq F_{1,n-3}(\omega)\}$ and the roots of formula (13); this is called the sum T^2-inversion method.

Since the inferential problem is almost symmetric in α and β, we only have to consider halve the number of parameter combinations. Moreover, the two pivotal statistics $\tau_\alpha(\alpha)^2$ and $\tau_\beta(\beta)^2$ are invariant with respect to the variances and the regression coefficients δ, β_0 and α_0, so we can set without loss of generality $\sigma_\varepsilon^2 = \sigma_\xi^2 = 1$ and $\delta = \beta_0 = \alpha_0 = 0$ in the DGP. The simulation study is based on 10,000 replications with sample sizes of 50, 200 and 500 for $\gamma \in \{0, 0.04, 0.16, 0.36, 0.64, 1\}$ for various values of α and β; see Figure 3.

We focus on the type-I error rates denoted by $\hat{\omega}$. These error rates are significantly different from the nominal significance level ω if

$$|\hat{\omega} - \omega| > 2.5758\sqrt{\omega(1-\omega)/10,000},$$

where $2.5758 = -z_{0.005}$. For $\omega = 2.5\%$, we have $|\hat{\omega} - \omega| > 0.4\%$. So, one out of hundred simulations ($1\% = 2 \cdot 0.005$) is expected to lead to error rates that are significantly different from ω. Note that a conservative confidence interval leads to error rates that are too low, although this might still lead to exact inference. This seems to be unavoidable if no knowledge about either α or β is known. However, error rates that are significantly too high are incompatible with exact inference procedures.

Insert Figure 3 around here.

Table 1 shows the results for $n = 50$. As expected, inference based on symmetric normal-based intervals lead to error rates that are significantly too low (i.e. γ is to the left; 26 in total) or too high (i.e. γ is to the right; 15 in total). The Meeker-based interval performs much better, but also suffers from error rates that are too high over all values of γ considered. In total, 13 error rates are too low and 25 are too high. The performance of interval based on inverting the LR statistic and the χ^2-distribution is very similar to the Meeker-based intervals: 6 error rates are too low and 25 are too high. Finally, intervals based on the sum T^2-inversion method and the F-distribution is the best performer: 15 error rates are too low, but only 2 are too high. All conclusions also qualitatively hold for $n = 200$, although the differences are now smaller. The number of error rates that are too high reduces to 9 for the Meeker intervals, to 4 for the χ^2-LR inverted statistic and only 3 for the sum T^2-inversion method. The results for $n = 500$ show that the performance of the two confidence intervals based on inverting the LR statistic are very similar, although the use of the F-quantile still leads to a small advantage. Note that for this quite large sample size, the Meeker interval still significantly overrejects in 7 out of 54 parameter combinations.

Insert Tables 1-3 around here.

6 Conclusion

In this paper, we have looked at various confidence intervals for the indirect effect. Prior research has focused on a Wald-type statistic. However, in the likelihood approach, there are three classical tests: Wald, Lagrange Multiplier (or Score) and LR tests. The LR test plays a central role in the Neyman-Pearson lemma and it is usually the most powerful test. When testing the null hypothesis $H_0 : \alpha\beta = 0$, the LR test statistic leads to the minimum of two *t-ratios* (t-statistics for testing zero values under the null). It can be shown that the size of this test, i.e. probability to reject a true null over the whole restricted parameter space, can be bounded.

A confidence interval is obtained by inverting the LR statistic. Looking from a different perspective, a confidence interval can also be based on inversion of the sum of two pivotal *t-statistics*. The lower and upper limits are easily obtained by calculating the roots of a quartic polynomial. Analyzing the rejection rule for $H_0 : \alpha\beta = 0$ based on the confidence interval, a finite-sample correction is suggested, i.e. usage of critical values from a particular F-distribution. The Monte Carlo simulations show that the latter confidence interval performs best: it has the least type-I errors different from the nominal significance level.

References

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), 83–104.

MacKinnon, D. P., C. M. Lockwood, and J. Williams (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate behavioral research 39*(1), 99–128.

Meeker, W. Q., L. W. Cornwell, and L. A. Aroian (1981). The product of two normally distributed random variables. In W. J. Kennedy and R. E. Odeh (Eds.), *Selected tables in mathematical statistics*, Volume VII.

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

Sobel, M. E. (1990). Effect analysis and causation in linear structural equation models. *Psychometrika 55*(3), 495–515.

Tofighi, D. and D. P. MacKinnon (2011). Rmediation: An R package for mediation analysis confidence intervals. *Behavior research methods 43*(3), 692–700.

Figure 1: Critical values for $\alpha\beta = 0$ based on 100,000 replications and $n = 50$.

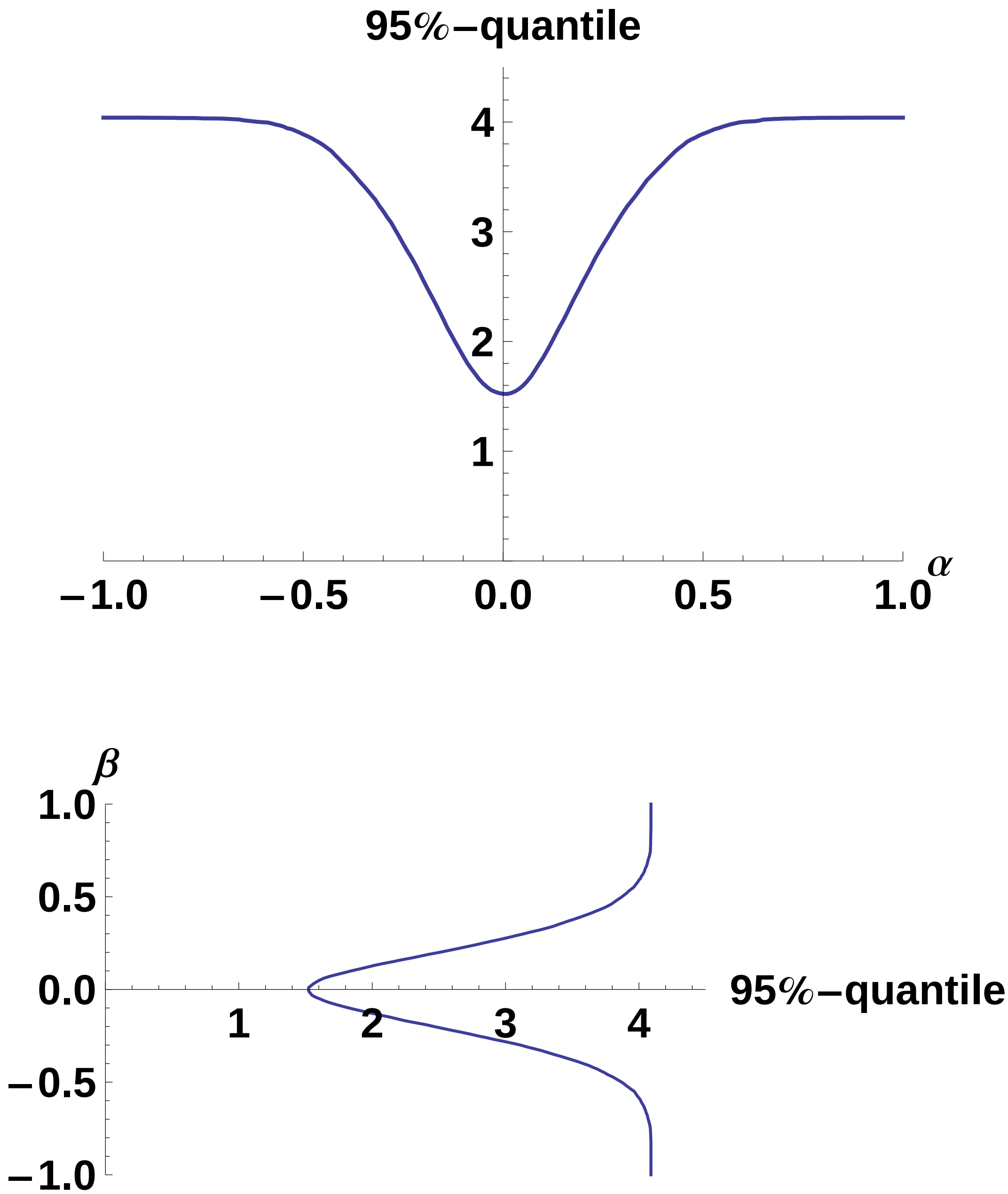

Figure 2: Confidence interval for γ based on inverting LR test statistic. Green points denote the lower and upper confidence limits.

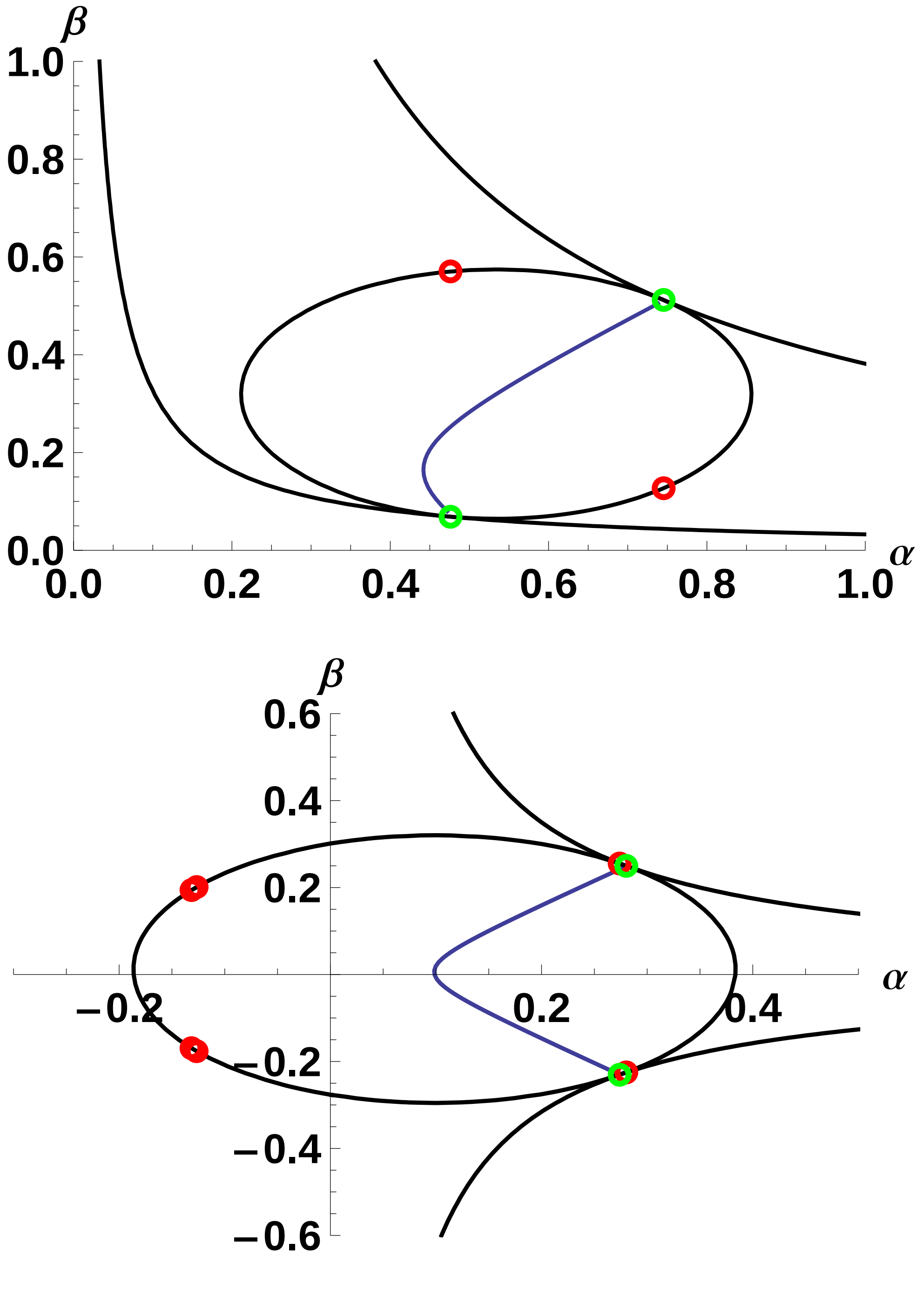

Figure 3: Parameter combinations (red circles) used in the Monte Carlo simulation. Combinations of (α, β) that lead to the same indirect effect $\gamma = \alpha\beta$ are joined by lines.

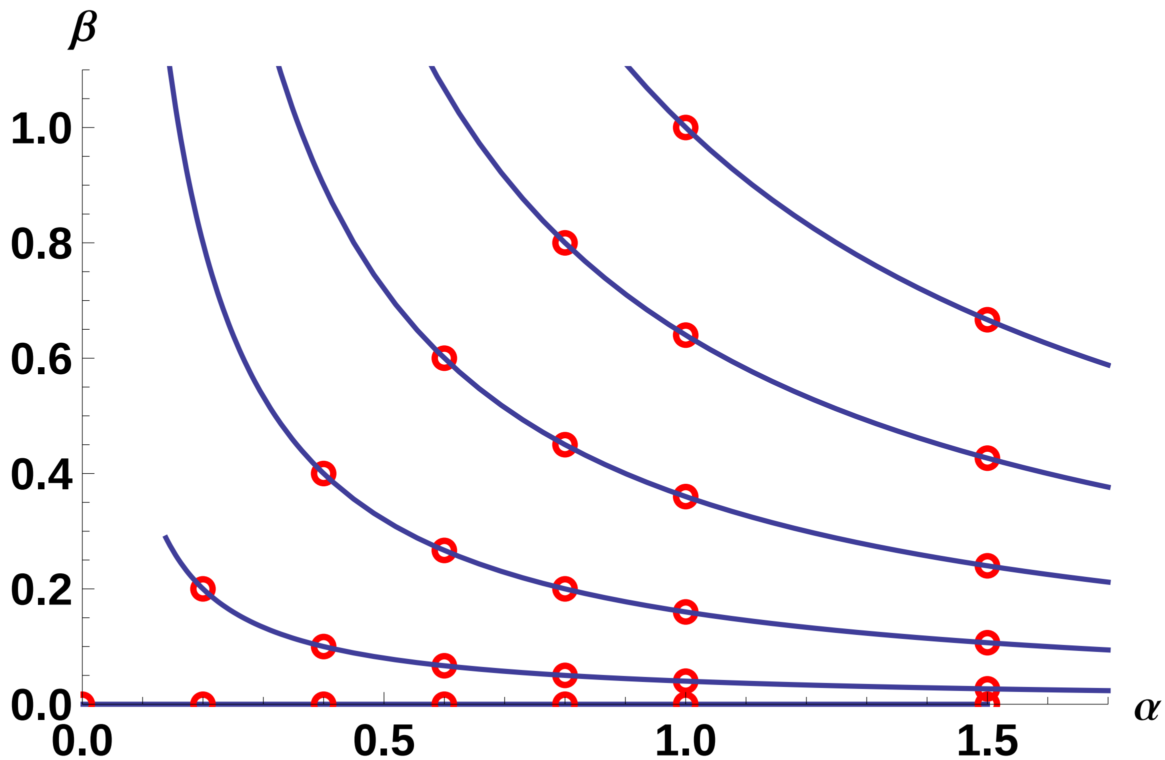

Table 1: Type-I Error Rates in percentage points based on 10,000 replications, nominal significance level of 5% (two-sided) and $n = 50$.

			Normal		Meeker		LR(χ2)		sumT2(*F*)	
γ	α	β	L	R	L	R	L	R	L	R
0	0.0	0	0.00 -	0.01 -	0.16 -	0.20 -	0.18 -	0.25 -	0.14 -	0.17 -
0	0.2	0	0.10 -	0.06 -	0.79 -	0.73 -	0.86 -	0.83 -	0.61 -	0.58 -
0	0.4	0	0.62 -	0.77 -	2.32	2.68	2.23	2.61	1.76 -	2.17
0	0.6	0	1.35 -	1.61 -	2.96 +	3.29 +	2.83	3.10 +	2.39	2.72
0	0.8	0	2.12	1.88 -	3.28 +	2.85	3.16 +	2.82	2.66	2.42
0	1.0	0	2.26	2.50	2.89	3.06 +	2.89	3.06 +	2.48	2.67
0	1.5	0	2.31	2.75	2.60	2.99 +	2.66	3.05 +	2.25	2.62
0.04	0.2	0.2	0.26 -	8.89 +	0.89 -	9.12 +	1.37 -	2.19	1.13 -	1.60 -
0.04	0.4	0.1	0.60 -	3.23 +	1.68 -	5.25 +	2.10	2.95 +	1.77 -	2.50
0.04	0.6	0.06667	1.22 -	2.09 -	2.66	3.54 +	2.86	3.03 +	2.38	2.50
0.04	0.8	0.05	1.63 -	2.68	2.59	3.60 +	2.66	3.36 +	2.24	2.93 +
0.04	1.0	0.04	2.45	2.30	3.46 +	2.71	3.58 +	2.66	2.89	2.42
0.04	1.5	0.02667	2.36	2.67	2.67	2.98 +	2.75	3.04 +	2.35	2.53
0.16	0.4	0.4	0.63 -	7.35 +	1.47 -	4.95 +	1.87 -	3.44 +	1.57 -	2.89
0.16	0.6	0.26667	0.91 -	5.07 +	1.89 -	4.20 +	2.30	3.35 +	1.97 -	2.91 +
0.16	0.8	0.2	1.25 -	3.64 +	2.13	3.53 +	2.49	3.17 +	2.03 -	2.74
0.16	1.0	0.16	1.87 -	3.02 +	2.56	3.20 +	2.78	2.97 +	2.40	2.54
0.16	1.5	0.10667	2.47	2.70	2.95 +	2.84	3.05 +	2.92 +	2.54	2.43
0.36	0.6	0.6	0.88 -	5.62 +	1.60 -	3.78 +	2.12	3.16 +	1.67 -	2.60
0.36	0.8	0.45	1.23 -	4.90 +	2.04 -	3.68 +	2.55	3.16 +	2.05 -	2.69
0.36	1.0	0.36	1.67 -	3.95 +	2.36	3.38 +	2.70	3.16 +	2.29	2.66
0.36	1.5	0.24	2.46	3.07 +	2.89	3.03 +	3.16 +	3.01 +	2.66	2.61
0.64	0.8	0.8	1.36 -	4.90 +	1.91 -	3.62 +	2.29	3.25 +	1.90 -	2.74
0.64	1.0	0.64	1.33 -	4.33 +	1.88 -	3.40 +	2.25	3.15 +	1.89 -	2.58
0.64	1.5	0.42667	1.91 -	3.48 +	2.41	3.09 +	2.62	3.05 +	2.31	2.55
1	1.0	1	1.44 -	4.19 +	2.03 -	3.21 +	2.35	3.04 +	1.99 -	2.48
1	1.5	0.66667	1.93 -	3.55 +	2.28	3.09 +	2.45	2.99 +	2.18	2.60

Remarks: sumT2(F) is based on $\tau_\alpha(\alpha)^2+\tau_\beta(\beta)^2$ and the F-distribution. Columns indicated with an L show the estimated error rates when γ is to the left, i.e. γ <lower confidence limit. R-columns show the estimated error rates when γ is to the right, i.e. γ >upper confidence limit. A '$-$'-sign indicates that the error rate is significantly too low, while a '$+$'-sign indicates that the error rate is significantly too high.

Table 2: Type-I Error Rates in percentage points based on 10,000 replications, nominal significance level of 5% (two-sided) and $n = 200$.

			Normal		Meeker		LR(χ2)		sumT2(F)	
γ	α	β	L	R	L	R	L	R	L	R
0	0.0	0	0.00 -	0.00 -	0.11 -	0.13 -	0.10 -	0.13 -	0.10 -	0.13 -
0	0.2	0	0.49 -	0.43 -	2.25	2.11	2.05 -	1.90 -	1.94 -	1.86 -
0	0.4	0	1.71 -	1.82 -	2.68	2.70	2.54	2.55	2.46	2.44
0	0.6	0	2.13	2.05 -	2.66	2.45	2.60	2.43	2.44	2.35
0	0.8	0	2.14	2.44	2.28	2.63	2.28	2.63	2.22	2.54
0	1.0	0	2.54	2.70	2.62	2.89	2.62	2.89	2.54	2.77
0	1.5	0	2.21	2.44	2.25	2.46	2.28	2.46	2.18	2.43
0.04	0.2	0.2	0.60 -	7.34 +	1.53 -	4.96 +	1.96 -	3.34 +	1.88 -	3.21 +
0.04	0.4	0.1	1.07 -	3.18 +	1.94 -	3.08 +	2.25	2.52	2.09 -	2.45
0.04	0.6	0.06667	1.96 -	2.52	2.64	2.70	2.65	2.54	2.57	2.45
0.04	0.8	0.05	2.10	2.55	2.38	2.70	2.39	2.61	2.31	2.54
0.04	1.0	0.04	2.46	2.50	2.72	2.59	2.73	2.58	2.62	2.46
0.04	1.5	0.02667	2.51	2.94 +	2.59	2.97 +	2.63	2.98 +	2.50	2.88
0.16	0.4	0.4	1.24 -	4.65 +	1.87 -	3.33 +	2.13	2.78	2.04 -	2.71
0.16	0.6	0.26667	1.61 -	3.51 +	2.24	2.93 +	2.47	2.58	2.36	2.42
0.16	0.8	0.2	1.54 -	3.33 +	1.92 -	3.14 +	2.04 -	2.99 +	1.97 -	2.94 +
0.16	1.0	0.16	2.02 -	2.79	2.38	2.70	2.48	2.59	2.36	2.48
0.16	1.5	0.10667	2.34	2.50	2.47	2.50	2.51	2.51	2.39	2.40
0.36	0.6	0.6	1.46 -	4.02 +	2.00 -	3.12 +	2.32	2.87	2.16	2.74
0.36	0.8	0.45	1.76 -	3.38 +	2.11	2.91 +	2.34	2.71	2.19	2.60
0.36	1.0	0.36	2.08 -	3.24 +	2.55	2.90	2.72	2.74	2.57	2.62
0.36	1.5	0.24	2.49	2.84	2.66	2.71	2.70	2.65	2.62	2.51
0.64	0.8	0.8	1.74 -	3.88 +	2.11	3.33 +	2.36	3.14 +	2.23	2.96 +
0.64	1.0	0.64	1.81 -	3.38 +	2.22	2.83	2.45	2.70	2.30	2.58
0.64	1.5	0.42667	2.22	3.01 +	2.51	2.88	2.58	2.87	2.49	2.79
1	1.0	1	1.84 -	3.33 +	2.16	2.86	2.35	2.78	2.23	2.65
1	1.5	0.66667	2.08 -	2.91 +	2.39	2.62	2.46	2.55	2.42	2.46

Remarks: see Table 1.

Table 3: Type-I Error Rates in percentage points based on 10,000 replications, nominal significance level of 5% (two-sided) and $n = 500$.

			Normal		Meeker		LR(χ2)		sumT2(*F*)	
γ	α	β	L	R	L	R	L	R	L	R
0	0.0	0	0.00 -	0.00 -	0.08 -	0.14 -	0.09 -	0.16 -	0.09 -	0.16 -
0	0.2	0	1.28 -	1.46 -	2.67	3.18 +	2.44	2.88	2.40	2.85
0	0.4	0	2.61	2.23	3.08 +	2.54	3.04 +	2.50	2.95 +	2.46
0	0.6	0	2.48	2.58	2.61	2.77	2.58	2.76	2.55	2.71
0	0.8	0	2.22	2.79	2.29	2.92 +	2.28	2.91 +	2.23	2.83
0	1.0	0	2.30	2.44	2.32	2.46	2.32	2.46	2.30	2.44
0	1.5	0	2.78	2.29	2.82	2.31	2.83	2.32	2.78	2.29
0.04	0.2	0.2	1.08 -	5.57 +	1.87 -	3.75 +	2.24	2.84	2.20	2.81
0.04	0.4	0.1	1.70 -	3.08 +	2.24	2.93 +	2.32	2.68	2.27	2.64
0.04	0.6	0.06667	2.11	2.80	2.34	2.81	2.39	2.75	2.30	2.67
0.04	0.8	0.05	2.26	2.29	2.42	2.34	2.43	2.29	2.39	2.26
0.04	1.0	0.04	2.43	2.47	2.48	2.48	2.49	2.48	2.47	2.45
0.04	1.5	0.02667	2.38	2.77	2.40	2.77	2.41	2.79	2.38	2.73
0.16	0.4	0.4	1.82 -	3.96 +	2.35	3.05 +	2.49	2.89	2.46	2.86
0.16	0.6	0.26667	1.66 -	3.29 +	2.15	2.86	2.32	2.63	2.27	2.59
0.16	0.8	0.2	2.08 -	3.02 +	2.39	2.84	2.46	2.81	2.44	2.75
0.16	1.0	0.16	2.16	2.88	2.31	2.83	2.37	2.77	2.34	2.74
0.16	1.5	0.10667	2.39	2.30	2.50	2.30	2.53	2.30	2.46	2.30
0.36	0.6	0.6	1.87 -	3.20 +	2.33	2.73	2.50	2.64	2.46	2.61
0.36	0.8	0.45	1.96 -	3.13 +	2.27	2.74	2.42	2.60	2.36	2.57
0.36	1.0	0.36	2.22	3.05 +	2.43	2.78	2.53	2.66	2.47	2.62
0.36	1.5	0.24	2.39	2.88	2.53	2.82	2.59	2.80	2.51	2.72
0.64	0.8	0.8	1.93 -	2.86	2.23	2.50	2.32	2.45	2.31	2.39
0.64	1.0	0.64	2.17	3.20 +	2.39	2.89	2.45	2.79	2.43	2.77
0.64	1.5	0.42667	2.52	2.79	2.62	2.68	2.67	2.64	2.62	2.56
1	1.0	1	2.18	3.06 +	2.39	2.86	2.44	2.81	2.40	2.75
1	1.5	0.66667	2.13	3.15 +	2.23	2.99 +	2.31	2.93 +	2.25	2.91 +

Remarks: see Table 1.