This is a "preproof" accepted article for *Psychometrika*.
This version may be subject to change during the production process.
DOI: 10.1017/psy.2024.28

Robust and Efficient Mediation Analysis via Huber Loss

WenWu Wang[a], Xiujin Peng[a], Tiejun Tong[b]

[a]*School of Statistics and Data Science, Qufu Normal University, Qufu, China*
[b]*Department of Mathematics, Hong Kong Baptist University, Hong Kong*

Abstract

Mediation analysis is one of the most popularly used methods in social sciences and related areas. To estimate the indirect effect, the least-squares regression is routinely applied, which is also the most efficient when the errors are normally distributed. In practice, however, real data sets are often non-normally distributed, either heavy-tailed or skewed, so that the least-squares estimators may behave very badly. To overcome this problem, we propose a robust M-estimation for the indirect effect via a general loss function, with a main focus on the Huber loss which is more slowly varying at large values than the squared loss. We further propose a data-driven procedure to select the optimal tuning constant by minimizing the asymptotic variance of the Huber estimator, which is more robust than the least-squares estimator facing outliers and non-normal data, and more efficient than the least-absolute-deviation estimator. Simulation studies compare the finite sample performance of the Huber loss with the existing competitors in terms of the mean square error, the type I error rate, and the statistical power. Finally, the usefulness of the proposed method is also illustrated using two real data examples.

Keywords: Data-driven tuning constant, Huber loss, Indirect effect, Iteratively reweighted least-squares, M-regression

https://doi.org/10.1017/psy.2024.28 Published online by Cambridge University Press

1. Introduction

In social sciences and related areas, the effect of an exposure on the outcome variable is often mediated by an intermediate variable. Mediation analysis aims to identify the direct effect of the predictor on the outcome and the indirect effect between the same predictor and the outcome via the change in a mediator (MacKinnon, 2008). Since the seminal paper of Baron and Kenny (1986), mediation analysis has become one of the most popular statistical methods in social sciences. Empirical applications of mediation analysis have dramatically expanded in sociology, psychology, epidemiology, and medicine (Ogden et al., 2010; Lockhart et al., 2011; Rucker et al., 2011; Newland et al., 2013; Richiardi et al., 2013). In practice, however, researchers have found that the assumptions of traditional mediation analysis methods, e.g. normality and no outliers, do not match the data they collected, which may lead to misleading results (Yuan and MacKinnon, 2014; Preacher, 2015). To overcome the problem, it is often required to adopt some sophisticated models for mediation analysis (VanderWeele and Tchetgen, 2017; Frölich and Huber, 2017; Lachowicz et al., 2018). For more details on mediation analysis, one may refer to the recent books including, for example, MacKinnon (2008), VanderWeele (2015), and Hayes (2023).

One important issue in mediation analysis is to conduct the inference on the indirect effect, with a main focus on testing its statistical significance. In this direction, the first approach is the causal steps approach (Baron and Kenny, 1986), which specifies a series of tests of links in a causal chain.

https://doi.org/10.1017/psy.2024.28 Published online by Cambridge University Press

Moreover, some variants of this method that test three different hypotheses have also been proposed (Allison, 1995; Kenny et al., 1998). The second approach is the difference in coefficients approach (Freedman and Schatzkin, 1992), which takes the difference between a regression coefficient before and after being adjusted by the intervening variable. The third approach is the product of coefficients approach which involves paths in a path model (Sobel, 1982; MacKinnon et al., 1998, 2004). MacKinnon et al. (2002) compared 14 methods of testing the statistical significance of the indirect effect and found that the difference in coefficients approach and the product of coefficients approach have a better control on the type I error rate as well as a higher power in most cases. And between them, the product of coefficients method is more widely used mainly thanks to its clear causal path explanation (MacKinnon et al., 2004; Preacher and Hayes, 2008; Preacher and Selig, 2012; Yuan and MacKinnon, 2014).

To estimate the indirect effect, the least-squares (LS) regression is routinely applied, which is also the most efficient when the errors are normally distributed. In practice, however, real data sets are often non-normally distributed, either heavy-tailed or skewed (Field and Wilcox, 2017). As an example, Micceri (1989) examined 440 data sets from the psychological and educational literature and found that none of them were normally distributed at the $\alpha = 0.01$ significance level. When applied to non-normal data sets, the LS estimators may behave very badly (Huber and Ronchetti, 2009). To circumvent such drawbacks, some robust approaches have recently emerged in the mediation literature. Zu and Yuan (2010) adopted the local influence function to identify the strongly-affected outliers. Yuan and MacKinnon (2014) proposed the least-absolute-deviation (LAD) regression when the errors are heavy-tailed, and moreover, Wang and Yu (2023) established the statistical theory for the LAD estimation of the indirect effect. Lastly, as claimed by Preacher (2015), mediation analysis for non-normal variables has become an active research field.

To move forward, it is noteworthy that the LS and LAD estimators are special cases of the M-estimators, which minimize a specified loss function (Serfling, 2001; Hansen, 2022). Another popular loss function in the M-regression is known as the Huber loss function, which utilizes a tuning parameter to adjust the tail of the standard normal distribution (Huber, 1964). This tuning parameter controls the trade-off between the efficiency and robustness. Wang et al. (2007) found that the Huber loss function with the optimal tuning parameter can greatly improve the efficiency when maintaining the robustness. To the best of our knowledge, little work has been done on estimating the indirect effect from the perspective of the optimal loss.

This paper proposes to further advance the literature by developing robust estimation of the indirect effect. To be specific, our approach mainly alleviates effects in the response variable and implicitly assumes that there is no large leverage points in the independent variables. In Section 2, we introduce the M-regression in the simple mediation model with a general loss function. An iteratively reweighted least-squares algorithm is also proposed

https://doi.org/10.1017/psy.2024.28 Published online by Cambridge University Press

to numerically solve the M-regression, as well as to construct two robust confidence intervals. In Section 3, we propose a data-driven approach to select the optimal tuning constant, and moreover study the statistical properties specifically for the Huber loss. In Section 4, we conduct simulation studies to assess the finite sample performance of the Huber loss and compared it with the existing competitors used in mediation analysis. We further illustrate the advantages of our method by an empirical example in Section 5, and conclude the paper in Section 6 with some discussion.

2. Simple Mediation Model

The simplest mediation model is given in Figure 1, where X is the independent variable, Y is the dependent variable, and M is the mediating variable that mediates the effects of X on Y. Given the observations (X_i, M_i, Y_i) for $i = 1, \dots, n$, this simple mediation model consists of three linear regression equations as

$$Y_i = \beta_1 + cX_i + \epsilon_{1,i}, \tag{1}$$

$$M_i = \beta_2 + aX_i + \epsilon_{2,i}, \tag{2}$$

$$Y_i = \beta_3 + c'X_i + bM_i + \epsilon_{3,i}, \tag{3}$$

where c represents the total effect of X on Y, a represents the relation between X and M, c' represents the direct effect of X on Y after adjusting the effect of M, b represents the relation between M and Y after adjusting

https://doi.org/10.1017/psy.2024.28 Published online by Cambridge University Press

the effect of X, and the random errors $\epsilon_{j,i}, j = 1, 2, 3$, are independent of the corresponding regressors.

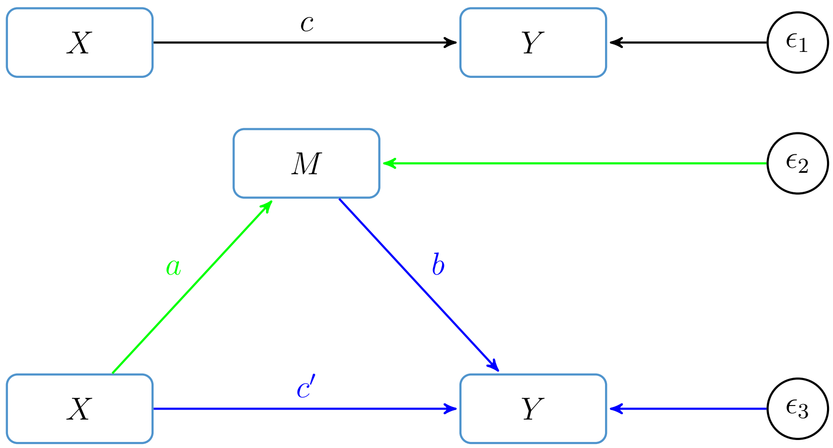

Figure 1: Causal diagram of the simple mediation model.

2.1. M-regression

To alleviate the effects of influential observations in the least-squares fitting, we adopt the M-regression to estimate the regression parameters, which can be regarded as a generalization of the maximum likelihood estimation as follows:

$$(\hat{\beta}_1, \hat{c})^T = \arg\min_{\beta_1, c} \sum_{i=1}^{n} \rho(Y_i - \beta_1 - cX_i), \tag{4}$$

$$(\hat{\beta}_2, \hat{a})^T = \arg\min_{\beta_2, a} \sum_{i=1}^{n} \rho(M_i - \beta_2 - aX_i), \tag{5}$$

$$(\hat{\beta}_3, \hat{c'}, \hat{b})^T = \arg\min_{\beta_3, c', b} \sum_{i=1}^{n} \rho(Y_i - \beta_3 - c'X_i - bM_i), \tag{6}$$

https://doi.org/10.1017/psy.2024.28 Published online by Cambridge University Press

where $\rho(\cdot)$ is the loss function with three properties: (i) nonnegativity such that $\rho(\epsilon) \geq 0$ with $\rho(0) = 0$, (ii) symmetricity such that $\rho(\epsilon) = \rho(-\epsilon)$, and (iii) monotonicity such that $\rho(\epsilon) \geqslant \rho(\epsilon')$ for any $|\epsilon| \geqslant |\epsilon'|$.

Let $\psi(\epsilon) = (d/d\epsilon)\rho(\epsilon)$ be the first derivative of the loss function, referred to as the influence curve. Let also $X = (X_1, \ldots, X_n)^T$, $M = (M_1, \ldots, M_n)^T$, $Y = (Y_1, \ldots, Y_n)^T$, $I = (1, \ldots, 1)^T$, $\tilde{\boldsymbol{X}} = (I, X)$, and $\check{\boldsymbol{X}} = (I, X, M)$. For large samples, we further assume that $\boldsymbol{U}$ is the limiting matrix of $(n^{-1}\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{X}})^{-1}$, and $\boldsymbol{V}$ is the limiting matrix of $(n^{-1}\check{\boldsymbol{X}}^T\check{\boldsymbol{X}})^{-1}$. Then by Huber and Ronchetti (2009), we have the following asymptotic normality for the M-estimators of the regression parameters.

Lemma 1. *For the mediation model linked with (1)-(3), under the regularity conditions given on pages 163-164 of Huber and Ronchetti (2009), the M-estimators in (4)-(6) are all normally distributed:*

$$\sqrt{n}(\hat{c} - c) \sim N\left(0, \frac{\mathrm{E}_{\epsilon_1}[\psi^2]}{(\mathrm{E}_{\epsilon_1}[\psi'])^2}\boldsymbol{U}_{[2,2]}\right), \quad \sqrt{n}(\hat{a} - a) \sim N\left(0, \frac{\mathrm{E}_{\epsilon_2}[\psi^2]}{(\mathrm{E}_{\epsilon_2}[\psi'])^2}\boldsymbol{U}_{[2,2]}\right),$$
$$\sqrt{n}(\hat{c'} - c') \sim N\left(0, \frac{\mathrm{E}_{\epsilon_3}[\psi^2]}{(\mathrm{E}_{\epsilon_3}[\psi'])^2}\boldsymbol{V}_{[2,2]}\right), \quad \sqrt{n}(\hat{b} - b) \sim N\left(0, \frac{\mathrm{E}_{\epsilon_3}[\psi^2]}{(\mathrm{E}_{\epsilon_3}[\psi'])^2}\boldsymbol{V}_{[3,3]}\right).$$

Finally, based on the M-estimators in (4)-(6), we can define two new estimators of the indirect effect: one is the difference estimator $\hat{c} - \hat{c'}$ and the other is the product estimator $\hat{a}\hat{b}$.

2.2. Solution to M-regression

For a general loss $\rho(\cdot)$, noting that the M-estimator may not have an explicit expression, a numerical solution is often required. To present our

https://doi.org/10.1017/psy.2024.28 Published online by Cambridge University Press

algorithm, we will focus only on (4) since the same algorithm can be extended to solve (5) and (6) as well. Differentiating the objective function $\sum_{i=1}^{n} \rho(Y_i - \beta_1 - cX_i)$ with respect to β_1, c and setting the partial derivatives to be zero, it yields a system of two estimating equations as

$$\sum_{i=1}^{n} \psi(Y_i - \beta_1 - cX_i) = 0,$$
$$\sum_{i=1}^{n} \psi(Y_i - \beta_1 - cX_i)X_i = 0.$$

Further by introducing the weight function $w(e) = \psi(e)/e$, the estimating equations can be rewritten as

$$\sum_{i=1}^{n} w_i \times (Y_i - \beta_1 - cX_i) = 0,$$
$$\sum_{i=1}^{n} w_i \times (Y_i - \beta_1 - cX_i)X_i = 0,$$

where $w_i = w(Y_i - \beta_1 - cX_i)$. Solving these two equations is equivalent to minimizing

$$\sum_{i=1}^{n} w_i \times (Y_i - \beta_1 - cX_i)^2,$$

which is a weighted LS problem. Moreover, an iteratively reweighted least-squares (IRLS) algorithm can be appropriate to obtain the numerical solution of the regression coefficients, because the weights depend on the regression coefficients, and the regression coefficients in turn depend on the weights (Holland and Welsch, 1977). To also handle the multiple-minima problem,

https://doi.org/10.1017/psy.2024.28 Published online by Cambridge University Press

in case it has, we choose several different points in the parameter space as the initial estimates, in such a way to get a higher confidence to obtain the true global minimum (Green, 1984). More specifically, the IRLS algorithm for our problem is as follows.

Algorithm 1: Iteratively Reweighted Least-Squares

---

1. Choose some initial estimates $\theta^{(0)} = (\beta_1^{(0)}, c^{(0)})^T$, including those from the LS or LAD methods.
2. For each iteration $t \geq 1$, calculate the residuals $e_i^{(t-1)} = Y_i - \beta_1^{(t-1)}$ $-c^{(t-1)} X_i$ and the associated weights $w_i^{(t-1)} = w(e_i^{(t-1)})$.
3. Obtain the weighted LS estimates

$$\theta^{(t)} = (\tilde{\boldsymbol{X}}^T \boldsymbol{W}^{(t-1)} \tilde{\boldsymbol{X}})^{-1} \tilde{\boldsymbol{X}}^T \boldsymbol{W}^{(t-1)} Y,$$

where $\boldsymbol{W}^{(t-1)} = \text{diag}\{w_i^{(t-1)}\}$.

4. Repeat steps 2 and 3 until $\theta^{(t)}$ satisfies $\|\theta^{(t)} - \theta^{(t-1)}\|_2 < 10^{-5}$.

2.3. Error Conditions for Model Consistency

When is the product of parameters ab equal to the difference in parameters $c - c'$ in population? This is an important question in mediation analysis since it uncovers the relationship between the indirect, direct and total effects (Yuan and MacKinnon, 2014; Wang et al., 2023; Wang and Yu, 2023).

Note that the three regression equations, (1)-(3), are interrelated in the simple mediation model. By substituting (2) into (3), we have

$$
\begin{aligned}
Y_i &= \beta_3 + c' X_i + b(\beta_2 + aX_i + \epsilon_{2,i}) + \epsilon_{3,i} \\
&= (\beta_3 + b\beta_2) + (c' + ab)X_i + \epsilon_i,
\end{aligned} \tag{7}
$$

where $\epsilon_i = b\epsilon_{2,i} + \epsilon_{3,i}$. Assume that $\epsilon_{2,i}$ and $\epsilon_{3,i}$ are independent and symmetrically distributed with median 0, then ϵ_i is also symmetric with $\mathrm{Med}[\epsilon_i] = 0$ (see Proposition 1 in Wang and Yu (2023)). In addition, let $\epsilon_{1,i}$ also be symmetrically distributed with $\mathrm{Med}[\epsilon_{1,i}] = 0$. Then by (1) and (7),

$$
\begin{aligned}
&\mathrm{Med}[Y_i|X_i] = \beta_1 + cX_i + \mathrm{Med}[\epsilon_{1,i}|X_i], \\
&\mathrm{Med}[Y_i|X_i] = (\beta_3 + b\beta_2) + (c' + ab)X_i + \mathrm{Med}[\epsilon_i|X_i].
\end{aligned}
$$

Noting also that the random errors are independent of the corresponding regressors as assumed in Section 2.1, we have $\mathrm{Med}[\epsilon_{1,i}|X_i] = \mathrm{Med}[\epsilon_{1,i}] = 0$ and $\mathrm{Med}[\epsilon_i|X_i] = \mathrm{Med}[\epsilon_i] = 0$, and moreover,

$$
\beta_1 + cX_i \equiv (\beta_3 + b\beta_2) + (c' + ab)X_i, \quad i = 1, \dots, n,
$$

which further yields that $\beta_1 = \beta_3 + b\beta_2$ and $c = c' + ab$. Finally, by comparing (1) and (7), we also have $\epsilon_i = \epsilon_{1,i}$. For convenience, we summarize the above result in Theorem 1.

Theorem 1. *In the simple mediation model, given the independence of the errors and the corresponding regressors, we further assume that the errors*

https://doi.org/10.1017/psy.2024.28 Published online by Cambridge University Press

are independent and symmetrically distributed with a unique median 0 *for* $j = 1, 2, 3$*. Then we have* $ab = c - c'$*, which builds an equality between the indirect effect, direct effect and total effect.*

Remark 1. Many error distributions satisfy the error assumption in Theorem 1. For instance, when $\epsilon_{2,i}$ and $\epsilon_{3,i}$ are independent and normally distributed, Yuan and MacKinnon (2014) discussed the model consistency. Wang and Yu (2023) further discussed the consistency conditions for the LAD loss and obtained the similar equality as in Theorem 1.

2.4. Inference Based on Confidence Interval

There are two estimators for the indirect effect: $\hat{c} - \hat{c'}$ and $\hat{a}\hat{b}$. Unlike the equivalence of the two LS estimators (MacKinnon et al., 1995; Wang et al., 2023), the two M-estimators of the indirect effect for a general loss are not the same in general, that is, $\hat{a}\hat{b} \neq \hat{c} - \hat{c'}$. Simulation studies show that the product estimator is often more efficient than the difference estimator (see Appendix A). Interestingly, the same conclusion can also be seen when the LAD loss is applied (Wang and Yu, 2023). In view of this, we thus consider the null hypothesis $H_0 : ab = 0$. To test whether $ab = 0$, there are two common methods in the literature including the parameter method (Sobel, 1982) and the nonparametric resampling method (MacKinnon et al., 2004; Preacher and Selig, 2012).

To move forward, our first method is to perform a robust Sobel test.

https://doi.org/10.1017/psy.2024.28 Published online by Cambridge University Press

Given the robust estimates $\hat{a}$ and $\hat{b}$, we define the robust test statistic as

$$Z = \frac{\hat{a}\hat{b}}{\widehat{\mathrm{SE}}_{Sobel}},$$

where $\widehat{\mathrm{SE}}_{Sobel} = \sqrt{\hat{a}^2 \times \widehat{\mathrm{SE}}_b^2 + \hat{b}^2 \times \widehat{\mathrm{SE}}_a^2}$, and $\widehat{\mathrm{SE}}_a$ and $\widehat{\mathrm{SE}}_b$ are the standard errors (SEs) of $\hat{a}$ and $\hat{b}$, respectively. Following Theorem 1, the two SEs can be estimated by

$$\widehat{\mathrm{SE}}_a = \left(\frac{n^{-1}\sum_{i=1}^n \psi^2(M_i - \hat{\beta}_2 - \hat{a}X_i)[(\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{X}})^{-1}]_{[2,2]}}{[n^{-1}\sum_{i=1}^n \psi'(M_i - \hat{\beta}_2 - \hat{a}X_i)]^2} \right)^{1/2},$$

$$\widehat{\mathrm{SE}}_b = \left(\frac{n^{-1}\sum_{i=1}^n \psi^2(Y_i - \hat{\beta}_3 - \hat{c}'X_i - \hat{b}M_i)[(\check{\boldsymbol{X}}^T\check{\boldsymbol{X}})^{-1}]_{[3,3]}}{[n^{-1}\sum_{i=1}^n \psi'(Y_i - \hat{\beta}_3 - \hat{c}'X_i - \hat{b}M_i)]^2} \right)^{1/2}.$$

Moreover, the normal-based $(1-\alpha)\%$ CI of ab can be constructed as

$$[\hat{a}\hat{b} - z_{1-\alpha/2}\widehat{\mathrm{SE}}_{Sobel},\ \hat{a}\hat{b} + z_{1-\alpha/2}\widehat{\mathrm{SE}}_{Sobel}],$$

where α is the significance level, and $z_{1-\alpha/2}$ represents the $(1-\alpha/2)$ quantile of the standard normal distribution. Note however that, when a and b are small, the sampling distribution of $\hat{a}\hat{b}$ may not be normal (MacKinnon et al., 2004; Wang et al., 2023). Thus to obtain an accurate CI, critical values of the distribution of $\hat{a}\hat{b}$ can be obtained by Mote Carlo simulation study (Meeker et al., 1981; Meeker and Escobar, 1994). In fact, one can easily obtain these critical values via inputting $\hat{a}$, $\hat{b}$, $\widehat{\mathrm{SE}}_a$ and $\widehat{\mathrm{SE}}_b$ into an R procedure `medci()` which was introduced by Tofighi and MacKinnon (2011).

Our second method to construct CI is the bootstrap method based on resampling. The bootstrap method is nonparametric and robust in the sense that it does not need to estimate the SEs. First, we repeatedly resample the original dataset with replacement (Efron and Tibshirani, 1993); second, we estimate the indirect effect for each bootstrap sample using our proposed Huber method; third, we construct the CI by the percentile bootstrap (PRCT) as $[q_{\alpha/2}, q_{1-\alpha/2}]$, where $q_{\alpha/2}$ is the $\alpha/2$ quantile of the empirical distribution of the indirect effect. To adjust and remove the potential estimation bias, the bias-corrected and accelerated bootstrap (BCa) is an important variation (Efron, 1987; Efron and Tibshirani, 1993). In general, the BCa method can yield a more accurate CI than the PRCT method when the true parameter value is not the median of the distribution of the bootstrap estimates (MacKinnon et al., 2004).

3. Robust and Efficient Estimation via Huber Loss

From a likelihood perspective, the best loss function would be the negative log-likelihood function (Schrader and Hettmansperger, 1980). Nevertheless, since the likelihood function is often unknown, one needs to specify an appropriate loss function in real applications. In this section, we study the robust and efficient estimation using the Huber loss with the optimal choice of tuning parameter. Note that our methodology is general and can also be extended to other loss functions.

1 *3.1. Huber Loss*

The Huber loss, as defined in Huber (1964), is given as

$$\rho_H(e) = \begin{cases} \frac{1}{2}e^2, & \text{if} \quad |e| \le k, \\ k|e| - \frac{1}{2}k^2, & \text{if} \quad |e| > k, \end{cases}$$

$$\psi_H(e) = \begin{cases} e, & \text{if} \quad |e| \le k, \\ k \times \operatorname{sgn}(e), & \text{if} \quad |e| > k, \end{cases}$$

where $k > 0$ is the tuning parameter. A smaller value of k produces more resistance to outliers, but at the expense of lower efficiency when the error is normal. For instance, by letting $k = 1.345\sigma$ with σ being the standard deviation of the error, it will yield a 95% efficiency for the normal errors, which is also resistant to outliers with a breakdown point of 5.8%. Moreover, the standard deviation σ can be estimated robustly by the median absolute deviation (MAD) as

$$\hat{\sigma}_{MAD} = \text{Med}\{|e_i|\}/0.6745.$$

For any error ϵ, we denote $\tau = \sigma_\psi^2 / B_\psi^2$ as the asymptotic variance of the Huber estimator (Huber, 1964), where $\sigma_\psi^2 = \mathrm{E}[\psi^2(\epsilon)]$ and $B_\psi = \mathrm{E}[\psi'(\epsilon)]$. We then minimize the τ value to determine the optimal $\rho(\cdot)$. For the Huber loss

https://doi.org/10.1017/psy.2024.28 Published online by Cambridge University Press

with a given k, we have

$$B_\psi(k) = \int_{-k}^{k} \mathrm{d}F(\epsilon),$$
$$\sigma_\psi^2(k) = \int_{-k}^{k} \epsilon^2 \mathrm{d}F(\epsilon) + k^2(1 - B_\psi(k)),$$

where $F(\cdot)$ is the cumulative distribution function of ϵ.

Remark 2. As $k \to \infty$, the Huber loss becomes the LS loss so that $\tau_{LS} = \sigma^2$, where σ^2 is the variance of the error distribution. As $k \to 0$, the Huber loss becomes the LAD loss so that $\tau_{LAD} = 1/(4f(0)^2)$, where $f(0)$ is the density value of the error distribution at 0. Based on the observational data, the optimal tuning constant can be selected to obtain the smallest estimation variance. From this viewpoint, the Huber estimator is more efficient than its competitors when dealing with the unknown and complex error distributions.

3.2. Optimal Tuning Constant

As is known, the tuning parameter k of the Huber loss can have a great impact on the estimation efficiency. When the error is normally distributed without contamination, the best choice of k is ∞. On the other hand, when the error follows a heavy-tailed distribution such as the t distribution, then k tends to be a small value close to 0.

We adopt a numerical method proposed by Wang et al. (2007) to select the optimal tuning constant, which minimizes the asymptotic variance of the estimator. For the Huber loss, the optimal k minimizes the efficiency factor

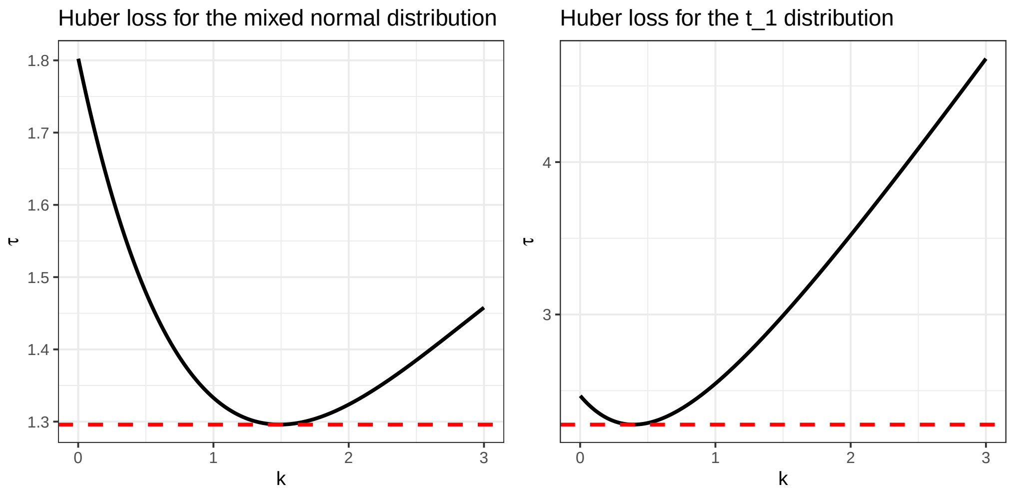

Figure 2: $\tau(k)$ is plotted for $0.9N(0,1)+0.1N(0,3^2)$ (left) and t_1 (right). The corresponding red lines are $\tau(1.489)=1.296$ and $\tau(0.395)=2.278$, respectively.

τ with a three-step procedure as follows. First, we compute $\tau(k)$ for a range of k values, i.e., $0 \le k \le K$ by 0.001, where K is a positive number, e.g. $K=4$. Second, we select the optimal k as

$$k_{opt} = \arg\min_{0<k\le K} \tau(k).$$

Lastly, we compute the minimum value $\tau(k_{opt})$. In Appendix B, we provide an R procedure to obtain the optimal tuning constant with a known error distribution.

For ease of reference, we also list the optimal k_{opt} and $\tau(k_{opt})$ in Table 1 for some error distributions, including the standard normal distribution $N(0,1)$, the Laplace distribution Laplace$(0,1)$, the mixed normal distribution

Table 1: Optimal k and $\tau(k)$ for various error distributions and loss functions.

Distribution	k_H	$\tau_H(k)$	τ_{LS}	τ_{LAD}
$N(0,1)$	∞	1	1	1.571
Laplace$(0,1)$	0	1	2	1
$0.9N(0,1)+0.1N(0,3^2)$	1.489	1.296	1.800	1.803
$0.9N(0,1)+0.1N(0,10^2)$	1.222	1.432	10.900	1.897
t_1	0.395	2.278	∞	2.467
t_2	0.692	1.722	∞	2

$0.9N(0,1)+0.1N(0,\sigma^2)$ with $\sigma=3$ or 10, and the t distribution with 1 or 2 degrees of freedom. In general, the Huber loss with the optimal tuning parameter k is more efficient than the LS and LAD losses, since the less τ is, the more efficient the loss is. Moreover, to intuitively reflect the variation trend of $\tau(k)$ as k varies, we also plot the $\tau(k)$ function for a normal mixed and t_1 distributions in Figure 2. It is evident that the value of $\tau(k)$ varies dramatically along with the k value.

3.3. Nonparametric Selection of Tuning Constant

Following (4) and letting $e_i = Y_i - \hat{\beta}_1 - \hat{c}X_i$ be the residuals, we propose to estimate τ nonparametrically by

$$\hat{\tau}(k) = \frac{\hat{\sigma}^2_{\psi}(k)}{\hat{B}^2_{\psi}(k)}, \tag{8}$$

https://doi.org/10.1017/psy.2024.28 Published online by Cambridge University Press

where $\hat{\sigma}_\psi^2(k) = n^{-1}\sum_{i=1}^n \psi^2(e_i)$ and $\hat{B}_\psi(k) = n^{-1}\sum_{i=1}^n \psi'(e_i)$. More specifically for the Huber loss, we have

$$\hat{B}_\psi(k) = \frac{1}{n}\sum_{i=1}^n \mathrm{I}(|e_i| \le k),$$
$$\hat{\sigma}_\psi^2(k) = \frac{1}{n}\sum_{i=1}^n \left\{ e_i^2 \mathrm{I}(|e_i| \le k) + k^2 \mathrm{I}(|e_i| > k) \right\},$$

where I is the $0-1$ indicator function.

We propose a data-driven procedure that determines the optimal $\hat{k}$ by minimizing $\hat{\tau}(k)$, which is, in fact, similar to Wang et al. (2007) for a linear regression model with a scale parameter σ. Our new procedure is summarized in Algorithm 2.

Algorithm 2: Nonparametric Selection of Tuning Constant

1. Select the initial estimates $(\hat{\beta}_1, \hat{c})^T$, e.g. the LAD estimates.
2. Compute $\hat{\tau}(k)$ for a range of k values satisfying $0.2 \le k \le 3\hat{\sigma}_{MAD}$ by 0.01, and then choose the optimal k as

$$\hat{k}_{opt} = \arg \min_{0.2 \le k \le 3\hat{\sigma}_{MAD}} \hat{\tau}(k).$$

3. Obtain the robust estimates of the regression parameters using the IRLS in Algorithm 1 with $k = \hat{k}_{opt}$.

Note that in the algorithm, we have specified the maximum allowable k as $3\hat{\sigma}_{MAD}$, which is often treated as sufficient since the probability that

https://doi.org/10.1017/psy.2024.28 Published online by Cambridge University Press

the errors fall within the interval $[-3\hat{\sigma}_{MAD}, 3\hat{\sigma}_{MAD}]$ is as large as 99.73% for the normal errors. To further investigate the performance of the proposed method on the selection of tuning constant k, we conduct a simulation study and report the results in Table B of the Appendix. When the sample size is large, the selected tuning constant k is very close to the theoretical one. Moreover, we note that the standard deviation of the tuning constant decreases dramatically as the sample size increases. These findings coincide with the conclusion in Wang et al. (2007).

4. Simulation Studies

Two simulation studies are carried out to evaluate the performance of the proposed method. Simulation A compares the efficiency of the three estimators based on the LS, LAD and Huber losses under various designs, and Simulation B evaluates their type I error rate and power. For the simulation settings, we follow Yuan and MacKinnon (2014) and Wang and Yu (2023) and set $\beta_2 = \beta_3 = 0$, $c' = 1$, and $a = b = 0.14, 0.39, 0.59$. Moreover, the sample size is set at $n = 50, 200, 1000$, corresponding to the small, medium and large samples, and four error distributions will be considered including $N(0, 1)$, Laplace$(0, 1)$, $0.9N(0, 1) + 0.1N(0, 10^2)$, and t_2.

For each simulated dataset, we estimate the regression parameters based on the LS, LAD and Huber losses, and apply the product $\hat{a}\hat{b}$ to estimate the indirect effect. Then with 1000 simulations for each setting, we compute the

https://doi.org/10.1017/psy.2024.28 Published online by Cambridge University Press

mean square error (MSE) to assess the estimation accuracy as follows:

$$\mathrm{MSE}[\hat{a}\hat{b}] = \frac{1}{1000}\sum_{i=1}^{1000}(\hat{a}\hat{b} - ab)^2.$$

Moreover, we apply the type I error rate and the statistical power to assess the performance of the LS, LAD and Huber estimators for testing $H_0 : ab = 0$. We use the robust Sobel test (Sobel Z), the percentile bootstrap (PRCT), and the BCa methods to construct the CIs. The type I error rate denotes the probability of incorrectly rejecting the null hypothesis when it is actually true, whereas the statistical power refers to the probability correctly rejecting the null hypothesis when the alternative hypothesis is true. A good testing procedure should control the type I error rate and, meanwhile, it also maximizes the power as much as possible. In practice, the empirical type I error rate (or power) is calculated as the proportion of CIs that do not contain zero when the indirect effect does not exist (or exists).

4.1. Efficiency of the LS, LAD and Huber Estimators

The MSE($\times 10^3$) and standard deviation (SD$\times 10^3$) of the LS, LAD and Huber estimators are presented in Table 2 for various designs. Comparing the MSE of the three estimators, we have two main findings. First, the MSE and SD of the three estimators decrease as the sample size increases. Second, the MSE and SD of the Huber estimator are always the smallest or close to the smallest. When the error follows $N(0, 1)$ (or Laplace$(0, 1)$), the LS (or LAD) estimator provides the optimal estimation. In these two

https://doi.org/10.1017/psy.2024.28 Published online by Cambridge University Press

cases, the Huber estimator performs very close to the performance of the optimal estimator. While for $0.9N(0,1)+0.1N(0,10^2)$ and t_2, the MSE of the Huber estimator is the smallest among the three estimators. To conclude, the Huber estimator is more efficient than the LAD estimator when the error distribution is normal, and is more robust than the LS estimator when the error distribution is non-normal.

4.2. Type I Error Rate and Power

We now apply the Sobel Z, PRCT and BCa methods to construct the 95% CI. Note that the medium effect sizes ($a=b=0.39$) will yield a high power even when the sample size is moderate ($n=200$). Thus to save space, we omit the simulation for the large effect size.

Table 3 report the type I error rates of the three estimators under various designs. When the sample size is large, i.e. $n=1000$, we note that the type I error rates of the LS, LAD and Huber estimators are all controlled in most cases. One exception is the LS estimator with the CIs constructed by the BCa method, which was also observed by Fritz et al. (2012) with an explanation that the increased type I error rate is a function of an interaction between the nonzero effect size and the sample size. Another notable situation is that the type I error rate of the Huber loss Sobel test is slightly too high for the mixed normal and t_2 under the small and moderate sample sizes. Possible reasons can be, e.g., the standard error used for the Sobel test $\widehat{\mathrm{SE}}_{Sobel}$ is affected by the Optimizer's curse (Smith and Winkler, 2006), and/or there is

Table 2: MSE ($\times 10^3$) and SD ($\times 10^3$ labeled below MSE) for the LS, LAD and Huber estimators.

	$a=b=0.14$			$a=b=0.39$			$a=b=0.59$		
n	LS	LAD	Huber	LS	LAD	Huber	LS	LAD	Huber
				$N(0,1)$					
50	**1.19**	2.22	1.69	**13.9**	22.25	18.2	**6.29**	10.26	8.34
	(2.91)	(4.57)	(3.73)	(10.44)	(15.65)	(13.34)	(21.22)	(31.61)	(27.07)
200	**0.24**	0.39	0.28	**3.83**	5.82	4.42	**1.69**	2.58	1.95
	(0.46)	(0.71)	(0.53)	(2.54)	(3.76)	(2.90)	(5.56)	(8.20)	(6.36)
1000	**0.04**	0.07	**0.04**	**0.74**	1.14	0.75	**0.32**	0.50	0.33
	(0.06)	(0.11)	(0.06)	(0.45)	(0.76)	(0.47)	(1.04)	(1.73)	(1.08)
				Laplace$(0,1)$					
50	1.34	0.84	**0.82**	6.76	4.92	**4.71**	14.65	11.03	**10.47**
	(3.29)	(1.75)	(1.80)	(11.93)	(7.55)	(7.43)	(23.80)	(16.33)	(15.72)
200	0.24	0.15	**0.14**	1.70	1.08	**1.07**	3.89	2.46	**2.43**
	(0.38)	(0.24)	(0.24)	(2.36)	(1.56)	(1.52)	(5.35)	(3.56)	(3.43)
1000	0.04	**0.02**	**0.02**	0.34	**0.17**	0.18	0.78	**0.39**	0.42
	(0.07)	(0.03)	(0.03)	(0.49)	(0.24)	(0.26)	(1.14)	(0.55)	(0.59)
				$0.9N(0,1)+0.1N(0,10^2)$					
50	1929.95	21.02	**18.14**	3130.94	73.84	**69.83**	4928.99	152.67	**147.54**
	(4461.83)	(61.24)	(57.06)	(6524.01)	(169.61)	(165.16)	(9212.73)	(315.85)	(308.93)
200	568.45	2.76	**2.53**	1586.29	13.78	**13.42**	3100.87	29.97	**29.39**
	(1386.04)	(9.65)	(8.06)	(3237.90)	(29.33)	(27.12)	(5592.51)	(55.75)	(52.94)
1000	28.44	0.32	**0.31**	154.49	2.34	**2.29**	340.70	5.33	**5.23**
	(71.51)	(0.53)	(0.49)	(264.92)	(3.34)	(3.23)	(531.80)	(7.54)	(7.36)
				t_2					
50	23.61	1.75	**1.52**	75.38	9.82	**8.84**	144.62	21.76	**19.70**
	(379.83)	(3.94)	(3.38)	(931.66)	(15.07)	(13.73)	(1574.78)	(31.82)	(29.63)
200	1.33	0.30	**0.26**	8.73	2.15	**1.90**	19.82	4.89	**4.32**
	(3.64)	(0.53)	(0.41)	(22.65)	(3.07)	(2.61)	(51.53)	(6.80)	(5.89)
1000	0.65	0.05	**0.04**	3.04	0.36	**0.31**	6.59	0.82	**0.71**
	(9.24)	(0.06)	(0.06)	(14.08)	(0.48)	(0.43)	(26.75)	(1.09)	(0.97)

Note that the bold font indicates the samllest MSE among the three estimators under one set of experimental conditions.

a potential gap between the optimal tuning constant and the one determined by Algorithm 2 in the small sample size. In Appendix E, we have also conducted another simulation study to assess their effect on the standard error used for the Sobel test. The results indicate that the $\widehat{\mathrm{SE}}_{Sobel}$ is indeed

https://doi.org/10.1017/psy.2024.28 Published online by Cambridge University Press

influenced by the optimizer's curse, whereas its effect will diminish as the sample size increases. At the same time, the Huber estimator with the fixed $k = 1.345$ performs better than the Huber estimator with the selected tuning constant (Huber-SEL) in the case of small sample size. Observing this, when the Huber-SEL estimator fails to yield satisfactory results, we suggest to take a moderate tuning constant, i.e. $k = 1.345$, as an alternative.

Following the same designs, we report the power of the three estimators in Table 4. For the normal errors, it is evident that the LS estimator not only controls the type I error rate but also achieves the highest power among the three estimators. Nevertheless, for the non-normal errors, the LS estimator is notably lacking in statistical power especially for the mixed normal distribution (e.g. $a = b = 0.14$, $0.9N(0,1) + 0.1N(0,10^2)$, and $n = 1000$). In addition, despite that the LAD estimator is the most robust method with respect to the outliers, it however suffers from the efficiency loss and consequently yields a lower power (e.g. $a = b = 0.14$, $N(0,1)$, and $n = 1000$). In contrast, the Huber estimator makes a trade-off between the efficiency and robustness, in which its power is close to the largest and, meanwhile, it also controls the type I error rate below 5% regardless of the error distribution.

5. Real Data Analysis

In this section, we conduct two real data analyses to illustrate the usefulness of the proposed method. Both the studies show that our newly method can provide a more efficient estimation than the existing competitors for me-

https://doi.org/10.1017/psy.2024.28 Published online by Cambridge University Press

Table 3: Type I error rates (%) of the LS, LAD and Huber estimators for various designs.

		Sobel Z			PRCT			BCa		
	n	LS	LAD	Huber	LS	LAD	Huber	LS	LAD	Huber
$a=0, b=0.14$					$N(0,1)$					
	50	0.0	0.0	0.3	0.3	0.0	0.2	1.5	2.1	1.4
	200	0.3	0.3	0.5	1.9	0.5	1.0	3.7	3.0	3.4
	1000	1.4	1.2	1.5	3.1	2.0	2.9	5.7	5.2	5.4
					Laplace$(0,1)$					
	50	0.2	0.2	1.5	0.7	0.1	0.3	1.7	2.3	1.8
	200	0.2	0.9	1.6	1.8	1.8	1.7	5.0	4.7	4.4
	1000	1.9	1.8	2.7	4.1	2.9	3.0	5.9	4.6	4.6
					$0.9N(0,1)+0.1N(0,10^2)$					
	50	0.0	1.5	**8.2**	0.9	0.3	0.3	4.7	2.4	2.1
	200	1.4	2.2	5.9	0.2	0.1	0.0	2.5	1.9	1.6
	1000	2.2	2.5	3.5	1.3	0.7	0.4	3.3	2.5	1.5
					t_2					
	50	0.0	0.5	2.9	1.0	0.3	0.6	4.4	3.7	2.2
	200	0.8	0.8	3.8	2.0	1.6	2.1	**7.8**	5.0	4.8
	1000	2.6	3.6	4.1	4.3	3.4	3.9	6.7	4.6	5.0
$a=0, b=0.39$					$N(0,1)$					
	50	1.7	0.8	4.2	3.9	1.3	3.0	**7.0**	6.7	6.3
	200	3.2	2.7	4.4	5.2	3.2	4.3	**7.0**	**7.1**	6.2
	1000	5.3	4.9	5.4	5.5	4.6	5.0	5.4	6.0	5.0
					Laplace$(0,1)$					
	50	1.2	1.2	**6.9**	4.4	1.5	3.5	**8.6**	6.7	6.0
	200	3.3	2.6	4.7	5.1	2.6	3.5	**7.4**	5.4	5.0
	1000	4.5	3.1	5.0	4.7	3.7	4.4	5.3	4.1	4.6
					$0.9N(0,1)+0.1N(0,10^2)$					
	50	0.0	0.9	**9.0**	1.4	0.3	0.4	**7.8**	3.3	2.6
	200	5.6	3.7	**7.9**	2.3	0.6	0.9	**8.5**	3.9	4.1
	1000	3.3	3.4	3.9	4.7	2.5	2.3	**8.0**	5.6	4.8
					t_2					
	50	0.7	1.0	**8.0**	4.2	2.2	3.4	**9.2**	6.8	5.7
	200	2.1	4.8	**8.8**	5.1	4.0	5.6	**10.4**	**7.3**	**7.4**
	1000	3.9	3.1	5.3	5.6	3.8	4.1	**7.2**	5.4	3.7

Note that the bold font indicates the excessive type I error rate which exceeds 6.8% since with 1000 independent simulation runs, the type I error rate of a test with level 0.05 is expected lie in the interval $[2.3\%, 6.8\%]$ with probability 0.99, using the normal approximation.

https://doi.org/10.1017/psy.2024.28 Published online by Cambridge University Press

Table 4: Power (%) of the LS, LAD and Huber estimators for various designs.

	n	Sobel Z LS	Sobel Z LAD	Sobel Z Huber	PRCT LS	PRCT LAD	PRCT Huber	BCa LS	BCa LAD	BCa Huber
						$N(0,1)$				
$a=b=0.14$	50	0.2	0.2	**1.4**	**2.5**	0.6	1.1	**4.9**	4.8	4.3
	200	9.9	4.4	**12.7**	**22.8**	8.3	18.6	**33.5**	17.7	27.8
	1000	**95.0**	67.4	94.8	**97.4**	80.8	96.9	**98.1**	84.8	**98.1**
						Laplace$(0,1)$				
	50	0.2	0.6	**4.4**	**3.4**	1.7	3.3	8.1	**8.7**	6.4
	200	8.5	22.7	**37.8**	23.7	33.9	**41.1**	34.9	44.6	**51.4**
	1000	94.0	99.9	**100.0**	96.8	**100.0**	**100.0**	97.5	**99.9**	**99.9**
						$0.9N(0,1)+0.1N(0,10^2)$				
	50	12.3	5.9	**28.4**	**2.4**	0.8	1.3	**6.1**	3.7	3.4
	200	1.3	15.3	**37.7**	0.8	1.8	**2.5**	3.3	**5.9**	**5.9**
	1000	3.9	66.5	**89.2**	1.4	16.5	**18.3**	2.7	23.8	**25.1**
						t_2				
	50	0.5	0.5	**5.2**	**1.9**	0.7	1.4	**5.3**	4.3	3.8
	200	2.9	12.4	**25.4**	9.1	17.8	**21.7**	15.9	25.4	**29.7**
	1000	20.3	80.1	**88.5**	36.3	81.2	**87.3**	40.6	83.1	**88.0**
						$N(0,1)$				
$a=b=0.39$	50	20.0	7.6	**32.0**	**35.2**	10.9	25.2	**47.8**	23.2	38.7
	200	**100.0**	95.7	**100.0**	**100.0**	98.3	**100.0**	**100.0**	97.5	**100.0**
	1000	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
						Laplace$(0,1)$				
	50	45.2	47.4	**84.9**	60.2	54.8	**67.1**	68.9	59.7	**75.3**
	200	99.8	**100.0**	**100.0**	99.8	99.9	**100.0**	99.7	99.7	**100.0**
	1000	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
						$0.9N(0,1)+0.1N(0,10^2)$				
	50	0.0	24.1	**74.7**	1.8	7.7	**9.6**	6.3	16.3	**18.7**
	200	0.7	73.5	**98.9**	2.5	30.5	**31.3**	8.6	37.6	**36.2**
	1000	11.8	99.4	**100.0**	9.1	**90.8**	90.2	12.9	**90.9**	90.3
						t_2				
	50	6.5	17.3	**50.7**	20.0	23.5	**30.1**	27.9	33.1	**36.7**
	200	51.5	95.0	**99.3**	64.1	96.7	**98.6**	66.4	95.4	**98.0**
	1000	93.3	99.9	**100.0**	93.8	**100.0**	**100.0**	92.9	**100.0**	**100.0**

Note that the bold font indicates the maximal empirical power among the three estimators under one set of experimental conditions.

diation analysis. To promote the practical application, we have also made the R code publicly available on GitHub at https://github.com/pxj66/REMA.git.

5.1. Pathways to Desistance Study

Our first study is to uncover the causal mechanisms between mental health and violent offending among serious adolescent offenders (Kim et al., 2024). In criminology, one possible mechanism is that individuals with mental health issues may be more likely to experience victimization, and this, in turn, may lead to their committing a serious crime. Our data comes from the Pathways to Desistance (PTD) study, which consists of 1354 serious juvenile offenders in two sites, including the Maricopa County in Arizona (N=654) and Philadephia County in Pennsylvania (N=700), over the years from 2000 to 2010 (Mulvey et al., 2013). Focusing on the data of baseline interviews, our study contains a total of 1195 respondents after the data cleansing.

Consider the linear mediation model,

$$
\begin{aligned}
Expvic_i &= \beta_2 + aHealth_i + \boldsymbol{\delta}_1^T \boldsymbol{Z}_i + \varepsilon_{2,i} \\
Offend_i &= \beta_3 + c' Health_i + bExpvic_i + \boldsymbol{\delta}_2^T \boldsymbol{Z}_i + \varepsilon_{3,i}
\end{aligned}
$$

where *Health* (mental health) is the independent variable, *Expvic* (experienced victimization) is the mediating variable, *Offend* (violent effending) is the response variable. In addition, $\boldsymbol{Z}$ denotes the matrix of other controlled variables including age, gender, enthnicity, family structure, parental

https://doi.org/10.1017/psy.2024.28 Published online by Cambridge University Press

warmth, alchhol, marijuana, gang membership, parental hostility, and unsupervised routine activities. We summarize the type and the measure of these variables in Appendix F.

Table 5: Skewness and kurtosis of two regression residuals and the Kolmogorov-Smirnov test for the pathways to desistance study.

	Skewness	Kurtosis	KS test (p-value)
$m - x$	0.3965	2.6953	5.787×10^{-4}
$y - m, x$	0.9056	4.9128	3.111×10^{-6}
Normal	0	3	

To assess the normality assumption for the errors, we compute the skewness and kurtosis of the residuals of y after regressing on x and m and the residuals of m after regressing on x, and then report them in Table 5. These values, together with the Kolmogorov-Smirnov (KS) test, clearly suggest a violation of the normality assumption. In view of this, we thus apply our new method to this dataset and also compare it with the existing methods for mediation analysis. Table 6 reports the indirect effects and the 95% CIs constructed by the Sobel Z, PRCT and BCa methods. From the results, we note that the three estimators produce similar and statistically significant indirect effects, whereas the Huber estimator yields the shortest CI.

5.2. Action Planning Study

Our second study is to investigate the relationship between action planning and physical activity. In psychology, it is known that the action planning can promote the physical activity, yet the underlying mechanism between them is often unclear. To explore it, an illustrative study has recently been

https://doi.org/10.1017/psy.2024.28 Published online by Cambridge University Press

Table 6: The indirect effect estimates and their 95% CIs based on the LS, LAD and Huber estimators for the pathways to desistance study.

		95% CI		
Method	$\hat{a}\hat{b}$	Sobel Z	PRCT	BCa
LS	0.0133	[0.0087, 0.0179]	[0.0087, 0.0188]	[0.0091, 0.0193]
		0.0091	0.0101	0.0102
LAD	0.0113	[0.0067, 0.0160]	[0.0051, 0.0180]	[0.0058, 0.0188]
		0.0092	0.0129	0.0130
Huber	0.0118	[0.0077, 0.0159]	[0.0076, 0.0170]	[0.0076, 0.0171]
		0.0082	0.0094	0.0095

conducted to investigate the action planning promoting the physical activity mediated by the automaticity (Maltagliati et al., 2023), in which a total of 135 participants over 18 years from the tertiary industry were recruited. Participants were asked to wear an accelerometer Actigraph GT3X+, which records their physical activity behaviors and the time of these activities on a notebook for a total of seven days. More specifically in their study, the action planning is the independent variable, measured by four-item Likert scales ranging from 1 (completely disagree) to 6 (full agree). And the automaticity is the mediating variable, measured by four-item of Self-Reported Habit Index ranging from 1 (strongly disagree) to 7 (strongly agree).

Consider the linear mediation model,

$$Auto_i = \beta_2 + aPlan_i + \delta_1 Sex_i + \delta_2 BMI_i + \delta_3 Ill_i + \varepsilon_{2,i}, \tag{9}$$

$$PA_i = \beta_3 + c'Plan_i + \delta_4 Sex_i + \delta_5 BMI_i + \delta_6 Ill_i + bAuto_i + \varepsilon_{3,i}, \tag{10}$$

where *Auto*, *Plan*, *Sex*, *BMI*, *Ill*, and *PA* represent the automaticity, action

https://doi.org/10.1017/psy.2024.28 Published online by Cambridge University Press

plan of exercise, gender, body mass index, illness, and physical activity of the respondent, respectively.

Table 7: Skewness and kurtosis of two regression residuals and the Kolmogorov-Smirnov test in action planning study.

	Skewness	Kurtosis	KS test (p-value)
$m-x$	2.6095	-0.1855	0.3582
$y-m,x$	9.7279	2.0349	0.0047
Normal	0	3	

To assess the normality assumption for the errors, we also compute the skewness and kurtosis of the two residuals, and then report them in Table 7. These values, together with the KS test, suggest a serious violation of the normality assumption for the $y-m,x$ regression residuals. Based on this, we also apply the proposed method to the dataset and then report the result in Table 8. First of all, the three methods produce positive indirect effects from 0.6600 to 0.7594. While for the CIs, only the LAD method shows insignificant outcome in the PRCT CI. At the same time, the Huber loss also yields the shortest CI among the three methods.

Table 8: The indirect effect estimates and their 95% CIs based on the LS, LAD and Huber losses for the action planning study.

		95% CI		
Method	$\hat{a}\hat{b}$	Sobel Z	PRCT	BCa
LS	0.7594	[0.2351, 1.3927]	[0.2714, 1.3755]	[0.3258, 1.4967]
		1.1576	1.1041	1.1709
LAD	0.6619	[0.1796, 1.2521]	[-0.1183, 1.3755]	[0.1487, 1.9636]
		1.0725	1.4938	1.8149
Huber	0.6600	[0.4199, 0.9347]	[0.0676, 1.0417]	[0.1470, 1.1820]
		0.5148	0.9741	1.035

https://doi.org/10.1017/psy.2024.28 Published online by Cambridge University Press

6. Discussion

This article proposed a novel M-regression for mediation analysis that minimizes the Huber loss function with the optimal tuning constant. The Huber loss can produce a more robust estimator compared to the LS loss when facing outliers and non-normal data, and on the other hand, it can produce a more efficient estimator compared to the LAD loss. Moreover, since the M-estimator may not have an explicit expression for a general loss function, we further proposed an IRLS algorithm for obtaining the numerical solutions. Under some mild conditions on the error distribution, the consistency of the mediation model was also established. Lastly, simulation studies and real data analysis showed that the Huber estimator has a better performance than the LS and LAD estimators.

In the literature, there are two methods commonly used to improve the estimation efficiency. The first method is the M-regression by selecting an optimal loss function from the loss function family. Besides the Huber loss that is among the most commonly used, other popular loss functions include, but not limited to, the Hampel loss (Hampel et al., 1986), the generalized Gauss-weight and linear quadratic losses (Koller and Stahel, 2011), and other general losses (Tukey, 1977; Barron, 2019). When the error distribution is skewed, it is appropriate to adopt the asymmetric Huber and Tukey's biweight losses for enhancing the estimation efficiency. In the field of microeconomics, the M-regression is done by solving the estimating equations which can be incorporated in the generalized method of moments (GMM), as also introduced in Chapter 6 of Cameron and Trivedi (2005). By making some additional moment conditions, one can obtain more efficient estimators. The second method is to combine the information of quantiles for improving the estimation efficiency, i.e., the composite quantile regression (Zou and Yuan, 2008), the weighted quantile average regression (Zhao and Xiao, 2014), and the combination of difference and robust methods (Wang et al., 2019). Hence as a further direction, it can be of interest to investigate whether the estimation efficiency and power of our new method can be further improved.

https://doi.org/10.1017/psy.2024.28 Published online by Cambridge University Press

Allison, P. D., 1995. The impact of random predictors on comparisons of coefficients between models: Comment on Clogg, Petkova, and Haritou. American Journal of Sociology 100, 1294–1305.

Baron, R. M., Kenny, D. A., 1986. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology 51, 1173–1182.

Barron, J. T., 2019. A general and adaptive robust loss function. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4331–4339.

Cameron, A. C., Trivedi, P. K., 2005. Microeconometrics: Methods and Applications. Cambridge University Press.

Efron, B., 1987. Better bootstrap confidence intervals. Journal of the American Statistical Association 82, 171–185.

Efron, B., Tibshirani, R. J., 1993. An Introduction to the Bootstrap. New York: Chapman & Hall.

Field, A. P., Wilcox, R. R., 2017. Robust statistical methods: A primer for clinical psychology and experimental psychopathology researchers. Behaviour Research and Therapy 98, 19–38.

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

Fritz, M. S., Taylor, A. B., MacKinnon, D. P., 2012. Explanation of two anomalous results in statistical mediation analysis. Multivariate Behavioral Research 47, 61–87.

Frölich, M., Huber, M., 2017. Direct and indirect treatment effects-causal chains and mediation analysis with instrumental variables. Journal of the Royal Statistical Society: Series B 79, 1645–1666.

Green, P. J., 1984. Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. Journal of the Royal Statistical Society: Series B 46, 149–192.

Hampel, F., Ronchetti, E., Rousseeuw, P., Stahel, W., 1986. Robust Statistics: The Approach Based on Influence Function. New York: Wiley.

Hansen, B., 2022. Econometrics. Princeton University Press.

Hayes, A. F., 2023. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach, 3rd Edition. New York: Guiford Press.

Holland, P. W., Welsch, R. E., 1977. Robust regression using iteratively reweighted least-squares. Communications in Statistics - Theory and Methods 6, 813–827.

https://doi.org/10.1017/psy.2024.28 Published online by Cambridge University Press

Huber, P. J., 1964. Robust estimation of a location parameter. The Annals of Mathematical Statistics 35, 73–101.

Huber, P. J., Ronchetti, E. M., 2009. Robust Statistics, 2nd Edition. New York: Wiley.

Kenny, D., Kashy, D., Bolger, N., 1998. Data analysis in social psychology. In: Gilbert, D. T., Fiske, S. T., Lindzey, G. (Eds.), The Handbook of Social Psychology. Boston: McGraw-Hill, pp. 233–265.

Kim, J., Harris, M. N., Lee, Y., 2024. The relationships between mental health and violent offending among serious adolescent offenders: An examination of the mediating role of experienced and witnessed victimization. Crime & Delinquency 70 (10), 2622–2646.

Koller, M., Stahel, W. A., 2011. Sharpening Wald-type inference in robust regression for small samples. Computational Statistics & Data Analysis 55, 2504–2515.

Lachowicz, M. J., Preacher, K. J., Kelley, K., 2018. A novel measure of effect size for mediation analysis. Psychological Methods 23, 244–261.

Lockhart, G., MacKinnon, D. P., Ohlrich, V., 2011. Mediation analysis in psychosomatic medicine research. Psychosomatic Medicine 73, 29–43.

MacKinnon, D. P., 2008. Introduction to Statistical Mediation Analysis. New York: Taylor & Francis Group.

MacKinnon, D. P., Lockwood, C. M. Williams, J., 2004. Confidence limits for the indirect effect: Distribution of the product and resampling methods. Multivariate Behavioral Research 39, 99–128.

MacKinnon, D. P., Lockwood, C., Hoffman, J., 1998. A new method to test for mediation. The Annual Meeting of the Society for Prevention Research, Park City, USA.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., Sheets, V., 2002. A comparison of methods to test mediation and other intervening variable effects. Psychological Methods 7, 83–104.

MacKinnon, D. P., Warsi, G., Dwyer, J. H., 1995. A simulation study of mediated effect measures. Multivariate Behavioral Research 30, 41–62.

Maltagliati, S., Sarrazin, P., Isoard-Gautheur, S., Pelletier, L., Rocchi, M., Cheval, B., 2023. Automaticity mediates the association between action planning and physical activity, especially when autonomous motivation is high. Psychology & Health, in press. https://doi.org/10.1080/08870446.2023.2188886.

Meeker, W. Q., Cornwell, L. W., Aroian, L. A., 1981. Selected Table in Mathematical Statistics (Volume VII). The Product of Two Normally Distributed Random Variables. American Mathematical Society.

Meeker, W. Q., Escobar, L. A., 1994. An algorithm to compute the cdf of the product of two normal random variables. Communications in Statistics - Simulation and Computation 23 (1), 271–280.

Micceri, T., 1989. The unicorn, the normal curve, and other improbable creatures. Psychological Bulletin 105, 156–166.

Mulvey, E. P., Schubert, C. A., Piquero, A., 2013. Pathways to desistance - final technical report. Tech. rep., National Institute of Justice. URL https://www.ojp.gov/library/publications/pathways-desistance-final-technical-report

Newland, R. P., Crnic, K. A., Cox, M. J., Mills-Koonce, W. R., Investigators, F. L. P. K., 2013. The family model stress and maternal psychological symptoms: Mediated pathways form economic hardship to parenting. Journal of Family Psychology 27, 96–105.

Ogden, C. L., Carroll, M. D., Curtin, L. R., Lamb, M. M., Flegal, K., 2010. Prevalence of high body mass index in U.S. children and adolescents, 2007-2008. Journal of the American Medical Association 303, 242–249.

Preacher, K. J., 2015. Advances in mediation analysis: A survey and synthesis of new developments. Annual Review of Psychology 66, 825–852.

Preacher, K. J., Hayes, A. F., 2008. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behavior Research Methods 40, 879–891.

Preacher, K. J., Selig, J. P., 2012. Advantages of Monte Carlo confidence intervals for indirect effects. Communication Methods and Measures 6, 77–98.

Richiardi, L., Bellocco, R., Zugna, D., 2013. Mediation analysis in epidemiology: Methods, interpretation and bias. International Journal of Epidemiology 42, 1511–1519.

Rucker, D. D., Preacher, K. J., Tormala, Z. L., Petty, R. E., 2011. Mediation analysis in social psychology: Current practices and new recommendations. Social and Personality Psychology Compass 5, 359–371.

Schrader, R. M., Hettmansperger, T. P., 1980. Robust analysis of variance based upon a likelihood ratio criterion. Biometrika 67, 93–101.

Serfling, R. J., 2001. Approximation Theorems of Mathematical Statistics. New York: John Wiley & Sons.

Smith, J. E., Winkler, R. L., 2006. The optimizer's curse: Skepticism and postdecision surprise in decision analysis. Management Science 52, 311–322.

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

Tofighi, D., MacKinnon, D., 2011. RMediation: An R package for mediation analysis confidence intervals. Behavior Research Methods 43, 692–700.

Tukey, J. W., 1977. Exploratory Data Analysis. Reading, MA: Addison-Wesley.

VanderWeele, T. J., 2015. Explanation in Causal Inference: Methods for Mediation and Interaction. Oxford: Oxford University Press.

VanderWeele, T. J., Tchetgen, E. J. T., 2017. Mediation analysis with time varying exposures and mediators. Journal of the Royal Statistical Society: Series B 79, 917–938.

Wang, W. W., Yu, P., 2023. Nonequivalence of two least-absolute-deviation estimators for mediation effect. TEST 32, 370–387.

Wang, W. W., Yu, P., Lin, L., Tong, T., 2019. Robust estimation of derivatives using locally weighted least absolute deviation regression. Journal of Machine Learning Research 20 (60), 1–49.

Wang, W. W., Yu, P., Zhou, Y., Tong, T., Liu, Z., 2023. Equivalence of two least-squares estimators for mediation effect. Current Psychology 42, 7364–7375.

Wang, Y. G., Lin, X., Zhu, M., Bai, Z., 2007. Robust estimation using the Huber function with a data-dependent tuning constant. Journal of Computational and Graphical Statistics 16, 468–481.

Yuan, Y., MacKinnon, D. P., 2014. Robust mediation analysis based on median regression. Psychological Methods 19, 1–20.

https://doi.org/10.1017/psy.2024.28 Published online by Cambridge University Press

Zhao, Z., Xiao, Z., 2014. Efficient regression via optimally combining quantile information. Econometric Theory 30, 1272–1314.

Zou, H., Yuan, M., 2008. Composite quantile regression and the oracle model selection theory. The Annals of Statistics 36, 1108–1126.

Zu, J., Yuan, Y., 2010. Local influence and robust procedures for mediation analysis. Multivariate Behavioral Research 45, 1–44.

https://doi.org/10.1017/psy.2024.28 Published online by Cambridge University Press

Appendix A. Comparing the Product and Difference Estimators

To compare the efficiency of the product and difference estimators, we follow the same simulation design as that for Simulation A in Section 4. Table A shows the MSE ($\times 10^3$) and SD ($\times 10^3$) of the product and difference estimators based on the Huber loss. It is evident that the MSE and SD of the product estimator are smaller than those of the difference estimator. We hence recommend to adopt the product estimator for the subsequent hypothesis testing.

Appendix B. An R Procedure for Selecting of the Tuning Constant

From the likelihood perspective, the optimal loss function is given as LS (or LAD) when the error distribution is Normal (or Laplace). Incorporating the relationship of the Huber loss with the LS and LAD losses, the optimal tuning constant is ∞ (or 0) for the Normal (or Laplace) distribution. For other error distributions, the optimal tuning constant minimizes the asymptotic variance of the Huber estimator. More specifically, we can compute $\hat{\tau}(k) = \hat{\sigma}_\psi^2/\hat{B}_\psi^2$ with a sequence of k, and then the optimal k, which correspondings to the minimum value of $\hat{\tau}$, can be located. In what follows, we provide the R code for two examples, one for the mixed normal distribution $0.9N(0,1) + 0.1N(0,3^2)$ and the other for the t_1 distribution.

https://doi.org/10.1017/psy.2024.28 Published online by Cambridge University Press

```r
1 # 0.9N(0,1) + 0.1N(0,9)
2 df1 <- function(x){
3   0.9 / sqrt(2*pi) * exp(-x^2/2) + 0.1 / sqrt(2*pi*9) * exp(-x^2/(2*9))
4 }
5
6 df2 <- function(x){
7   x^2 * (0.9 / sqrt(2*pi) * exp(-x^2/2) + 0.1 / sqrt(2*pi*9) * exp(-x^2/(2*9)))
8 }
9
10 i <- 0
11 tau <- numeric(0)
12 for (k in seq(0, 4, 0.001)) {
13   i <- i+1
14   B <- integrate(df1, -Inf, k)$value - integrate(df1, -Inf, -k)$value
15   Sig2 <- integrate(df2, -k, k)$value + k^2 * (1 - B)
16   tau[i] <- Sig2 / B^2
17 }
18
19 k <- seq(0, 4, 0.001)
20 plot(k, tau, type = "l")
21 k[which.min(tau)]
22
23 # t1
24 df1 <- function(x){ 1 / (pi * (1 + x^2)) }
25 df2 <- function(x){ x^2 / (pi * (1 + x^2)) }
26 # Run lines 10-21 again.
```


https://doi.org/10.1017/psy.2024.28 Published online by Cambridge University Press

Table A: MSE ($\times 10^3$) and SD ($\times 10^3$) for the product and difference estimators based on the Huber loss.

n	$a=b=0.14$ MSE$_P$	$a=b=0.14$ MSE$_D$	$a=b=0.39$ MSE$_P$	$a=b=0.39$ MSE$_D$	$a=b=0.59$ MSE$_P$	$a=b=0.59$ MSE$_D$
			$N(0,1)$			
50	1.52	2.79	7.52	10.21	16.41	20.27
	(3.58)	(6.59)	(12.61)	(17.26)	(25.41)	(30.80)
200	0.27	0.49	1.85	2.45	4.17	4.88
	(0.54)	(1.04)	(2.87)	(3.66)	(6.22)	(6.86)
1000	0.04	0.05	0.32	0.33	0.74	0.75
	(0.06)	(0.08)	(0.46)	(0.46)	(1.05)	(1.05)
			Laplace$(0,1)$			
50	0.81	2.14	4.65	8.44	10.31	16.38
	(1.73)	(4.50)	(7.21)	(13.03)	(15.18)	(25.10)
200	0.14	0.40	1.06	2.06	2.41	4.06
	(0.24)	(0.64)	(1.51)	(2.97)	(3.40)	(5.78)
1000	0.02	0.07	0.18	0.39	0.41	0.78
	(0.03)	(0.11)	(0.25)	(0.52)	(0.58)	(1.04)
			$0.9N(0,1)+0.1N(0,10^2)$			
50	18.20	22.04	71.08	72.33	150.04	150.36
	(58.27)	(57.54)	(169.64)	(149.64)	(317.91)	(287.81)
200	2.49	3.52	13.45	14.77	29.50	30.84
	(7.72)	(9.05)	(26.57)	(28.12)	(52.20)	(52.54)
1000	0.31	0.40	2.29	2.56	5.24	5.57
	(0.48)	(0.67)	(3.21)	(3.63)	(7.31)	(7.85)
			t_2			
50	1.52	6.53	8.74	20.79	19.41	39.02
	(3.29)	(12.91)	(13.76)	(31.75)	(29.78)	(58.49)
200	0.25	1.31	1.86	5.16	4.25	9.19
	(0.39)	(2.17)	(2.52)	(7.20)	(5.70)	(12.76)
1000	0.04	0.23	0.31	0.91	0.70	1.59
	(0.06)	(0.34)	(0.43)	(1.31)	(0.97)	(2.33)

Appendix C. Selection of the Tuning Constant

To evaluate the performance of Algorithm 2, we follow the same simulation design as that for Simulation A in Section 4. Table B presents the Mean, SD and Median of the selected tuning constant. As the sample size increases, the k values are very close to those from the theoretical results.

https://doi.org/10.1017/psy.2024.28 Published online by Cambridge University Press

This shows that Algorithm 2 provides a good performance for selecting the tuning constant for practical use. These findings also coincide with the conclusion in Wang et al. (2007). Note that k_1 and k_2 correspond to the chosen tuning constant from Equations (2) and (3), respectively. In practice, when the value of k is small, the value of the efficiency factor τ is very unstable. So we set the k value ranging from 0.2 to $3\hat{\sigma}_{MAD}$ by 0.01.

Table B: The values of Mean, SD and Median for the selected tuning constant.

	k_1			k_2			
n	Mean	SD	Median	Mean	SD	Median	Optimal
			$N(0,1)$				
50	1.001	0.715	0.840	0.980	0.703	0.810	
200	1.539	0.841	1.660	1.529	0.836	1.630	∞
1000	2.380	0.509	2.470	2.367	0.515	2.460	
			Laplace$(0,1)$				
50	0.426	0.296	0.310	0.441	0.303	0.330	
200	0.321	0.163	0.260	0.327	0.165	0.260	0
1000	0.253	0.075	0.220	0.256	0.075	0.230	
			$0.9N(0,1)+0.1N(0,10^2)$				
50	0.671	0.466	0.510	0.770	0.590	0.570	
200	0.749	0.447	0.700	0.783	0.477	0.730	1.222
1000	0.931	0.365	1.005	0.951	0.365	1.030	
			t_2				
50	0.612	0.470	0.430	0.611	0.466	0.430	
200	0.558	0.359	0.440	0.562	0.358	0.450	0.692
1000	0.558	0.279	0.530	0.551	0.274	0.520	

Appendix D. Asymptotic Relative Efficiency of the Huber Estimator

We first prove that the Huber loss with $k = 1.345$ produces a 95% efficiency for the normal errors. We focus on Equation (1): $Y = \beta_1 + cX + \epsilon_1$, where $\epsilon_1 \sim N(0, \sigma^2)$. When $k = k_0$, the efficiency factor of the Huber estimator is computed by $\tau_H = \sigma_\psi^2 / B_\psi^2$, where

$$\begin{aligned}
B_\psi(k_0) &= \int_{-k_0}^{k_0} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{x^2}{2\sigma^2}\right\} dx = \Phi\left(\frac{k_0}{\sigma}\right) - \Phi\left(-\frac{k_0}{\sigma}\right), \\
\sigma_\psi^2(k_0) &= \int_{-k_0}^{k_0} \frac{x^2}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{x^2}{2\sigma^2}\right\} dx + k_0^2\big[1 - B_\psi(k_0)\big] \\
&= \int_{-k_0/\sigma}^{k_0/\sigma} \frac{\sigma^2 x^2}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} dx + k_0^2\big[1 - B_\psi(k_0)\big] \\
&= \sigma^2 \left\{ G\left(\frac{k_0}{\sigma}\right) - G\left(-\frac{k_0}{\sigma}\right) + \Phi\left(\frac{k_0}{\sigma}\right) - \Phi\left(-\frac{k_0}{\sigma}\right) \right\} + k_0^2\big[1 - B_\psi(k_0)\big],
\end{aligned}$$

with $G(x) = -x(\sqrt{2\pi})^{-1} \exp\{-x^2/2\}$ and $\Phi(x)$ being the cumulative distribution function of the standard normal distribution. Then

$$\tau_H = \frac{\sigma_\psi^2(1.345\sigma)}{B_\psi^2(1.345\sigma)} = \frac{0.7101645\sigma^2}{0.6746565} = 1.052361\sigma^2$$

and so

$$\frac{\tau_{LS}}{\tau_H} = \frac{\sigma^2}{1.05236\sigma^2} = 0.9500003.$$

This shows that the asymptotic relative efficiency of the Huber estimator relatived to the LS estimator is 95% (Serfling, 2001).

At the same time, using Equation (4.52) on page 84 in Huber and Ronchet-

https://doi.org/10.1017/psy.2024.28 Published online by Cambridge University Press

ti (2009), we have

$$\frac{2\phi(k)}{k} - 2\Phi(-k) = \frac{\varepsilon}{1-\varepsilon},$$

where $\phi = \Phi'$ is the probability density function of the standard normal distribution. This implies that, when $k = 1.345$, the Huber estimator is resistant to outliers with a breakdown point of $\varepsilon = 5.8\%$.

Appendix E. Simulation Study for Huber Loss Sobel Test

We conduct a new simulation to investigate the effect of the optimizer's curse on the standard error used for the Sobel test, which is denoted by $\widehat{\mathrm{SE}}_{Sobel}$. To achieve this, we consider various tuning constants as alternatives, specifying the number of alternatives as 6, 30, and 291. These numbers correspond to step lengths of 0.5, 0.1, and 0.01, respectively, within the tuning constants' value range of $[0.1, 3]$. Concentrating on the type I error rates, we set the sample size to be 50, 200, 1000, or 2000. We also employ the same true values for the regression parameters and the error distributions as those specified in Section 4. Under 1000 simulated experiments, we then compute the mean standard error used for the Sobel test ($\widehat{\mathrm{SE}}_{Sobel}$) for the Huber loss with the selected tuning constant (Huber-SEL) under the different alternatives, the Huber loss with the fixed tuning constant (Huber-FIX), and the Huber loss with the optimal tuning constant (Huber-OPT).

Table E shows the mean standard error used for the Sobel test ($\widehat{\mathrm{SE}}_{Sobel}$) of the Huber-SEL, Huber-FIX and Huber-OPT estimators under various designs. First of all, the simulation results reveal that the Huber-SEL estimator

https://doi.org/10.1017/psy.2024.28 Published online by Cambridge University Press

is indeed affected by the optimizer's curse, yet this effect diminishes as the sample size increases. For example, let us look at the change in $\widehat{\mathrm{SE}}_{Sobel}$ between different number of alternatives. With $a = 0, b = 0.14$, $N(0, 1)$, and $n = 50$, the $\widehat{\mathrm{SE}}_{Sobel}$ of the Huber-SEL estimator exhibits a gradual decline away from the $\widehat{\mathrm{SE}}_{Sobel}$ of the Huber-OPT estimator, i.e. 31.50, with the values shifting from 27.90 to 27.66, and then to 26.95, as the number of alternatives increases. But when the sample size is 1000 or 2000, these values are all close to the optimal value. Secondly, we found that the Huber estimator with the fixed $k = 1.345$ performs better than the Huber-SEL estimator in the case of small sample sizes. However, as the sample size increases, the Huber-SEL estimator is more close to the Huber-OPT estimator than the Huber-FIX estimator. Combined with the conclusions drawn from Table B, two plausible explanations for the poor performance of the Huber loss Sobel tests in $n = 50$ or $n = 200$ are that there is a potential gap between the optimal tuning constant and the one determined by Algorithm 2 and the $\widehat{\mathrm{SE}}_{Sobel}$ is influenced by the optimizer's curse. For practical applications, when the Huber-SEL estimator fails to yield satisfactory results, we suggest to take a moderate tuning constant, i.e. $k = 1.345$, as an alternative.

Table E: Mean standard error ($\times 10^3$) used for the Sobel test for the Huber-SEL, Huber-FIX and Huber-OPT estimators under various designs.

		Huber-SEL			Huber-FIX			Huber-OPT
	n	$A = 6$	$A = 30$	$A = 291$	$k = 0.8$	$k = 1.345$	$k = 2.2$	$k = k^*$
$a = 0, b = 0.14$					$N(0, 1)$			
	50	27.90	27.66	26.95	35.09	32.92	31.55	31.50
	200	11.12	10.97	11.10	12.52	11.84	11.27	11.22
	1000	4.47	4.54	4.56	4.90	4.66	4.58	4.58
	2000	3.18	3.17	3.16	3.43	3.25	3.19	3.14
					$Laplace(0, 1)$			
	50	17.59	16.86	16.62	24.56	26.30	29.24	18.80
	200	7.47	7.30	7.09	9.07	9.58	10.58	7.71
	1000	3.23	3.18	3.15	3.76	4.01	4.34	3.26
	2000	2.29	2.27	2.22	2.61	2.79	2.99	2.27
					$0.9N(0, 1) + 0.1N(0, 10^2)$			
	50	19.44	18.70	17.26	26.54	26.52	28.01	25.65
	200	11.13	10.90	10.57	12.18	12.17	13.13	12.12
	1000	5.32	5.27	5.23	5.45	5.42	5.83	5.40
	2000	3.80	3.77	3.76	3.85	3.83	4.13	3.82
					t_2			
	50	24.22	23.27	23.11	33.95	34.51	40.41	34.45
	200	12.13	11.87	11.33	13.65	14.23	15.57	13.51
	1000	5.69	5.63	5.55	5.86	6.09	6.60	5.86
	2000	4.07	4.02	3.98	4.13	4.27	4.62	4.11
$a = 0, b = 0.39$					$N(0, 1)$			
	50	49.63	49.15	47.46	63.74	61.20	59.88	59.24
	200	27.04	26.62	26.77	30.44	28.93	28.10	27.94
	1000	12.31	12.34	12.35	13.29	12.70	12.45	12.43
	2000	8.75	8.73	8.72	9.40	8.95	8.77	8.70
					$Laplace(0, 1)$			
	50	36.00	34.75	33.65	49.91	52.23	55.87	39.55
	200	19.45	18.92	18.41	23.25	24.69	26.49	20.03
	1000	8.87	8.76	8.66	10.23	10.94	11.76	8.96
	2000	6.33	6.26	6.17	7.20	7.73	8.25	6.30
					$0.9N(0, 1) + 0.1N(0, 10^2)$			
	50	51.66	49.63	46.52	69.66	70.56	74.90	68.72
	200	30.96	30.37	29.44	33.91	33.90	36.57	33.73
	1000	14.82	14.69	14.58	15.18	15.08	16.25	15.03
	2000	10.58	10.51	10.48	10.74	10.66	11.50	10.64
					t_2			
	50	54.39	52.37	50.59	75.99	78.89	86.58	76.38
	200	32.62	31.70	30.80	36.56	38.11	41.37	36.39
	1000	15.73	15.56	15.35	16.24	16.83	18.24	16.22
	2000	11.29	11.16	11.07	11.46	11.88	12.86	11.45

https://doi.org/10.1017/psy.2024.28 Published online by Cambridge University Press

Appendix F. Measures of Interesting Variables in the PTD Study

Table F: Measures of interesting variables in the pathways to desistance study.

Interesting Variable	Data type	Measures
Key variable		
Violent Offending	Continuous	The proportion of 11 violent offenses committed during the last 6 months. The example items included beating up somebody badly needing a doctor, being in a fight, and killing someone. The higher the value, the greater variety of offenses the youth engaged in.
Mental Health	Continuous	Brief Symptom Inventory consists of 9 subscales. The larger the value, the worse the mental health of the respondents.
Experienced Victimization	Continuous	A total 6 items and example questions included "In the past 6 months, have you been chased where you thought you might be seriously hurt?". The larger value, the more victimizations are experienced.
Control variables		
Age	Continous	14-19.
Ethnicity	Discrete	White, Black, Hispanic, Other.
Gender	0-1	Male = 1, Female =0
Family Structure	Discrete	Single Biological Parent live with the youth, Two Biological Parent with the youth, Other
Gang Membership	0-1	The status of the participant's gang membership for last 6 months.
Parental Monitoring	Continuous	Parental Monitoring inventory (9 items) range from "never" to "always".
Parental Warmth	Continuous	Responses ranged from 1 (never) to 4 (always), with higher scores representing more parental warmth.
Parental Hostility	Continuous	There are a total of 42 items, 21 items for maternal and paternal respectively, and responses ranged from 1 (never) to 4 (always), with higher scores indicating greater hostility.
Unsupervised routine activities	Continous	The higher score indicates more unsupervised routine activities.
Alcohol	Continous	The frequency of alcohol drink consumption in the recall period.
Marijuana	Continuous	The frequency of using marijuana in the recall period.