Psychometrika (2026), **91**, 445–472
doi:10.1017/psy.2026.10091

PSYCHOMETRIC SOCIETY

THEORY AND METHODS

Robust Estimation of Polyserial Correlation Coefficients: A Density Power Divergence Approach

Max Welz

Department of Psychology, University of Zurich, Switzerland
Email: max.welz@uzh.ch

(Received 17 October 2025; revised 27 January 2026; accepted 2 February 2026)

Abstract

The association between a continuous and an ordinal variable is commonly modeled through the polyserial correlation model. However, this model, which is based on a partially-latent normality assumption, may be misspecified in practice, due to, for example (but not limited to), outliers or careless responses. The typically used maximum likelihood (ML) estimator is highly susceptible to such misspecification: One single observation not generated by partially-latent normality can suffice to produce arbitrarily poor estimates. As a remedy, we propose a novel estimator of the polyserial correlation model designed to be robust against the adverse effects of observations discrepant to that model. The estimator leverages *density power divergence estimation* to achieve robustness by implicitly downweighting such observations; the ensuing weights constitute a useful tool for pinpointing potential sources of model misspecification. The proposed estimator generalizes ML and is consistent as well as asymptotically Gaussian. As price for robustness, some efficiency must be sacrificed, but substantial robustness can be gained while maintaining more than 98% of ML efficiency. We demonstrate our estimator's robustness and practical usefulness in simulation experiments and an empirical application in personality psychology where our estimator helps identify outliers. Finally, the proposed methodology is implemented in free open-source software.

Keywords: careless responding; mixed data; model misspecification; polyserial correlation; robust estimation

1. Introduction

Empirical research in the psychological, social, and health sciences often features data that contain both continuous and ordered categorical (ordinal) random variables. Examples of continuous variables are household income, blood pressure, or time spent doing an activity of interest, while examples of ordinal variables are responses to rating scales (measuring, e.g., personality traits, wellbeing, or overall health), relationship status (with categories, such as *single*, *in a relationship*, and *married*), or grouped measurements of continuous variables like one's income group. The association between a continuous and an ordinal variable is typically modeled by means of *polyserial correlation* (Pearson, 1909). Polyserial correlation is a key building block in the analysis of mixed data, particularly structural equation models (SEMs). For instance, the popular R package `lavaan` (Rosseel, 2012) for SEM analyses by default uses polyserial correlation for SEMs involving both continuous and ordinal variables.

The polyserial correlation model postulates the existence of a latent continuous variable that underlies and governs the observed ordinal variable through an unobserved discretization process. The correlation between the observed continuous variable and the latent variable is called *polyserial correlation*, whereas the correlation between the observed continuous and the observed ordinal variable is called *point* polyserial correlation, where the latter can be computed from the former. For identifica-

© The Author(s), 2026. Published by Cambridge University Press on behalf of Psychometric Society.
This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/psy.2026.10091 Published online by Cambridge University Press

tion, the polyserial correlation model assumes that the observed continuous variable and the latent continuous variable are jointly normally distributed, thereby making it a partially-latent normality model. Estimation is typically conducted by means of maximum likelihood (ML; Cox, 1974; Olsson et al., 1982). However, the validity of the partially-latent normality assumption is often questionable in practice (e.g., Barbiero, 2025; Bedrick, 1995; Demirtas & Hedeker, 2016). Disquietingly, though, central statistical properties of polyserial correlation crucially depend on this very assumption to hold true: Violations of partially-latent normality have been shown to have potentially devastating effects on identification (e.g., Moss & Grønneberg, 2023) and ML estimation (e.g., Bedrick, 1995) of polyserial correlation. Nonnormality of (partially) latent variables can also introduce serious biases in SEM analyses conducted with software that assumes such normality (e.g., Foldnes & Grønneberg, 2020, 2022), such as `lavaan` (Rosseel, 2012), `LISREL` (Jöreskog & Sörbom, 1996), and `Mplus` (Muthén & Muthén, 2026).

Motivated by the susceptibility of polyserial correlation to partially-latent nonnormality, the contributions of this article are twofold. First, we study estimation of the polyserial correlation model when a (possibly empty) subset of the observed data are of low-quality and therefore may have *not* been generated by a normal distribution, such as (but not limited to) outliers in the continuous variable and/or careless responses in the ordinal variable. Consequently, such observations are uninformative for estimating the polyserial correlation model. This situation is called *partial model misspecification* because the assumption of partially-normality might be violated for parts of the data, but is satisfied for the remaining data. We demonstrate that already one single uninformative observation can suffice for ML estimation of polyserial correlation to yield arbitrary results and, furthermore, that the accuracy of ML estimation is highly susceptible to even minor misspecification of partially-latent normality. Partial misspecification stems from classic literature on robust statistics (e.g., Huber & Ronchetti, 2009) where it is known as *Huber contamination model*, owing to Huber (1964).

Second, in wake of the non-robustness of ML estimation to uninformative observations generated by partially-latent nonnormality, we propose an alternative estimator that is designed to be robust against such partial misspecification. The proposed methodology applies *density power divergence* (DPD) estimation (Basu et al., 1998) to the polyserial correlation model and achieves robustness by implicitly downweighting observations that cannot be sufficiently well fitted by that model. The ensuing weights are a useful tool for pinpointing potential sources of (partial) model misspecification. As an additional methodological contribution, we devise a simple rescaling of the weights to ensure that they are contained in the unit interval. Overall, to the best of our knowledge, the proposed methodology is the first contamination-robust approach to polyserial correlation.

In line with the partial misspecification framework, the robust estimator allows the model to be misspecified for an unknown fraction of uninformative observations, but makes *no assumption* on *how* and *where* partial misspecification occurs (which may be absent altogether). Consequently, partial misspecification can manifest through an unlimited and unrestricted variety of ways, such as (but not limited to) outliers or careless responses. Conversely, if the polyserial correlation model is correctly specified for all observations in a sample, then the robust estimator is, just like ML, consistent for the true parameter vector, and, therefore, generalizes ML estimation.

Studying their respective behavior under model misspecification, we show that both ML and the robust estimator still converge in probability, but to different limits. Crucially, the robust estimator converges to a parameter vector that is closer to the true parameter vector than ML, thereby gaining its robustness. Moreover, the robust estimator and ML remain asymptotically normally distributed, allowing for statistical inference both under correct and incorrect specification of the polyserial correlation model.

In robust statistics, there is a well-established fundamental tension between robustness and efficiency for estimation procedures for data involving continuous variables (e.g., Hampel et al., 1986; Huber & Ronchetti, 2009). Our estimator is no exception: The price to pay for its enhanced robustness manifests in the form of diminished efficiency compared to ML. However, we show that sacrificing as little as 2% of efficiency suffices to obtain a substantial gain in robustness, so the efficiency loss is only

https://doi.org/10.1017/psy.2026.10091 Published online by Cambridge University Press

comparatively minor. An additional price of our robust estimator is increased computational intensity. Nevertheless, using our implementation, the robust estimator usually executes in less than 2 seconds on a regular laptop, so the overall computational burden should remain small in practical applications. This implementation of our proposed methodology is publicly and freely available as part of the R package `robcat` (for "ROBust CATegorical data analysis"; Welz et al., 2026a) on CRAN (the Comprehensive R Archive Network) at https://CRAN.R-project.org/package=robcat.

This article is organized as follows. Section 2 reviews related literature, while Section 3 summarizes the polyserial correlation model and its ML estimation. Section 4 describes the partial misspecification framework adopted in this article. Section 5 introduces our robust estimator, and Section 6 derives its theoretical and computational properties. Section 7 carries out simulation studies to compare the performance of the robust estimator to ML in a variety of settings. Section 8 provides an empirical application on data from personality psychology. Section 9 discusses and concludes.

2. Literature

Potential violations of the normality assumption that underlies the polyserial correlation model have been studied in previous literature, though the misspecification framework used therein is fundamentally different from the partial misspecification framework adopted in this article. Specifically, previous literature focuses on *distributional misspecification*, where the polyserial correlation model is misspecified (usually through nonnormality) for the *entire* observed sample. In contrast, in partial misspecification, the model is only misspecified for a (possibly empty) subset of the sample. We explain the differences between partial and distributional misspecification in more detail in Section 4.2.

Focusing on distributional misspecification. Bedrick (1995) shows that the accuracy of normality-based ML estimation of polyserial correlation crucially depends on whether or not the marginal distribution of the latent variable that underlies the observed ordinal variable is normal. If that distribution is not normal, but, for instance, an exponential or t-distribution, ML estimates of polyserial correlation may be attenuated (Brogden, 1949; Kraemer, 1981; Lord, 1963). For a dichotomous ordinal variable, Demirtas and Vardar-Acar (2017) and Demirtas and Hedeker (2016) devise an algorithm that relates polyserial correlation to point polyserial correlation when the underlying joint distribution (which they assume to be known) is not bivariate normal. For a given potentially nonnormal joint distribution and a dichotomous ordinal variable, Cheng and Liu (2016) derive a general expression for the maximum point polyserial coefficient (in population). Barbiero (2025) generalizes the results of Cheng and Liu (2016) to a polytomous ordinal variable.

Moss and Grønneberg (2023, Section 2) and Grønneberg et al. (2020, Section 2.5) use partial identification analyses to study distributional misspecification of the polyserial model from a theoretical perspective. Specifically, given known marginal distributions of the observed continuous and latent continuous variables but keeping their joint distribution unspecified, they derive partial identification sets for the partially-latent correlation. They show that the partial identification sets tend to be uninformatively wide. Consequently, polyserial correlation is very sensitive to partially-latent joint nonnormality because different nonnormal distributions can yield widely different correlations. To reduce the width of the partial identification sets, one must make restrictive assumptions on the partially-latent joint distribution, such as Gaussian-like characteristics or equipping it with a parametric structure.

In a broader context, violations of normality assumptions have been studied extensively in the psychometric literature, especially with respect to the distribution of latent variables (e.g., Asparouhov & Muthén, 2016; Lyhagen & Ornstein, 2023; Monroe, 2018; Moss & Grønneberg, 2023; Roscino & Pollice, 2006; Yuan et al., 2004, and the references therein). A particular focus of recent literature has been the *polychoric* correlation model of Pearson and Pearson (1922), which models through a latent bivariate normality distribution the association of two latent variables that govern two observed ordinal variables (see Olsson, 1979 for a modern exposition). Foldnes and Grønneberg (2020) and Jin and Yang-Wallentin (2017) show that ML estimation of polychoric correlation is highly susceptible to

https://doi.org/10.1017/psy.2026.10091 Published online by Cambridge University Press

distributional misspecification, leading to possibly large biases in SEM analyses based on polychoric correlation (Foldnes & Grønneberg, 2022; Grønneberg & Foldnes, 2022).[1] Welz et al. (2026b) are concerned with partial misspecification of the polychoric correlation model, where latent normality is violated due to a (possibly empty) subset of data points that were generated by an unspecified and unknown nonnormal process. We use the same (partial) misspecification framework in this article. Welz et al. (2026b) show that having in a sample already about 5% of such observations, commonly referred to as *contamination*, can suffice for a substantial estimation bias. Contamination might arise due to, for instance but not limited to, careless responding to polytomous items, which has been identified as a major threat to the validity of psychometric analyses (e.g., Alfons & Welz, 2024; Arias et al., 2020; Bowling et al., 2016; Huang et al., 2015; Meade & Craig, 2012; Ward & Meade, 2023, and references therein). As a remedy, Welz et al. (2026b) propose a fully efficient contamination-robust estimator of the polychoric correlation model that makes no assumptions on the prevalence and type of contamination (which is possibly absent altogether). Their estimator exploits the theory of *C*-estimation (Welz, 2024), being a general framework of robust estimation with categorical data.

In the context of item response theory (IRT), Itaya and Hayashi (2025) focus on the seminal model by Rasch (1960), whose estimation may be compromised by contamination due to, for instance, careless responding or random guessing. They therefore robustify marginal ML estimation thereof by using a minimum DPD estimator (Basu et al., 1998) and design a majorization–minimization algorithm for this purpose. In this article, we also utilize DPD estimation theory to robustify the (joint) ML estimation of the polyserial correlation model.

3. Polyserial correlation

This section defines (point) polyserial correlation and then reviews ML estimation thereof.

3.1. The polyserial correlation model

Suppose we observe a continuous real-valued random variable X with unknown population mean $\mu = \mu_X \in \mathbb{R}$ and unknown population variance $\sigma^2 = \sigma_X^2 > 0$. In addition, suppose that we also observe a polytomous ordinal random variable Y that takes values in some finite set $\mathcal{Y}$ of known cardinality r. Without loss of generality, we assume throughout this article that $\mathcal{Y} = \{1, 2, \ldots, r\}$. Further assume that there exists a latent continuous random variable η governing the ordinal variable Y through the unobserved discretization process

$$Y = \begin{cases} 1 & \text{if } \eta < \tau_1, \\ 2 & \text{if } \tau_1 \leq \eta < \tau_2, \\ 3 & \text{if } \tau_2 \leq \eta < \tau_3, \\ \vdots & \\ r & \text{if } \tau_{r-1} \leq \eta, \end{cases} \tag{3.1}$$

where $-\infty < \tau_1 < \tau_2 < \cdots < \tau_{r-1} < +\infty$ are fixed but unknown *threshold* parameters. In practice, Y often denotes the responses to a Likert-type rating item with r response categories.

[1] It should be noted that ML estimation of polychoric correlation was originally believed to be fairly robust against distributional misspecification of latent normality (e.g., Coenders et al., 1997; Flora & Curran, 2004; Li, 2016), an assessment based on simulation experiments using the method of Vale and Maurelli (1983) for generating nonnormal bivariate data that are then discretized to be ordinal. However, Grønneberg and Foldnes (2019) show that ordinal data generated in this way are indistinguishable from ordinal data generated by latent normality. Using a method that ensures proper violation of latent normality (the VITA method of Grønneberg & Foldnes, 2017) reveals strong susceptibility of ML estimation of polychoric correlation to such violations (e.g., Foldnes & Grønneberg, 2020, 2022).

https://doi.org/10.1017/psy.2026.10091 Published online by Cambridge University Press

The primary object of interest is the fixed but unknown population correlation between the observed X and the latent η, denoted by

$$\rho = \mathbb{C}\mathrm{or}\left[X,\, \eta\right].$$

To identify the correlation coefficient ρ, one usually assumes that X and η are jointly normally distributed according to

$$\begin{pmatrix} X \\ \eta \end{pmatrix} \sim \mathrm{N}_2\left(\begin{pmatrix} \mu \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix}\right). \tag{3.2}$$

The bivariate normality model (3.2) implies the marginal normality properties $X \sim \mathrm{N}(\mu, \sigma^2)$ and $\eta \sim \mathrm{N}(0,1)$. It also identifies the correlation coefficient $\rho \in (-1,1)$ through the familiar identity $\mathbb{C}\mathrm{or}\left[X,\, \eta\right] = \mathbb{C}\mathrm{ov}\left[X,\, \eta\right] \big/ \sqrt{\mathbb{V}\mathrm{ar}\left[X\right] \mathbb{V}\mathrm{ar}\left[\eta\right]} = \rho$. Note that since the variable η is unobserved, its population mean and variance are not jointly identifiable, which is why they are fixed to 0 and 1, respectively.

Combining the discretization model (3.1) with the bivariate normality model (3.2) yields the *polyserial correlation model* (Pearson, 1913), or, in short, polyserial model. In this model, the correlation parameter $\rho = \mathbb{C}\mathrm{or}\left[X,\, \eta\right]$ is referred to as the *polyserial correlation coefficient*. If Y is dichotomous, the polyserial model reduces to the *biserial correlation model* of Pearson (1909). The theoretical properties of biserial correlation are studied by Tate (1955a, 1955b) as well as Jaspen (1946), and those of polyserial correlation by Olsson et al. (1982).

The polyserial model is subject to $d = r + 2$ parameters, namely, the mean and variance parameters (μ, σ^2) of the observed X, the polyserial correlation coefficient ρ from the normality model (3.2), as well as the $r - 1$ thresholds from the discretization process (3.1) of the latent η. These parameters are jointly collected in a d-dimensional parameter vector

$$\boldsymbol{\theta} = \left(\rho, \mu, \sigma^2, \boldsymbol{\tau}^\top\right)^\top,$$

where the vector $\boldsymbol{\tau} = \left(\tau_1, \ldots, \tau_{r-1}\right)^\top$ contains the $r - 1$ thresholds.

Under the polyserial model evaluated at a parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, the density of the observed-latent pair (X, η) of continuous variables is given by

$$p_{X\eta}\left(x, v; \boldsymbol{\theta}\right) = \phi_2\left(\begin{pmatrix} x \\ v \end{pmatrix}; \begin{pmatrix} \mu \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix}\right), \qquad x, v \in \mathbb{R},$$

where $\phi_2(\cdot; \boldsymbol{m}, \boldsymbol{S})$ is the density of the bivariate normal distribution with population mean $\boldsymbol{m}$ and covariance matrix $\boldsymbol{S}$. Further, denote by $P_{X\eta}(\cdot, \cdot; \boldsymbol{\theta})$ the distribution function corresponding to the normal density $p_{X\eta}(\cdot, \cdot; \boldsymbol{\theta})$.

For the observed variables (X, Y), the joint density at a realization $x \in \mathbb{R}$ and a response $y \in \mathcal{Y} = \{1, \ldots, r\}$ of the ordinal Y under the polyserial model at parameter $\boldsymbol{\theta}$ reads

$$p_{XY}\left(x, y; \boldsymbol{\theta}\right) = \int_{\tau_{y-1}}^{\tau_y} p_{X\eta}\left(x, v; \boldsymbol{\theta}\right) \mathrm{d}v, \tag{3.3}$$

where we adopt the conventions $\tau_0 = -\infty$ and $\tau_r = +\infty$.[2] The joint distribution function of the observed (X, Y) under the polyserial model can now be expressed as

$$P_{XY}\left(x, y; \boldsymbol{\theta}\right) = \mathbb{P}_{\boldsymbol{\theta}}\left[X \leq x, Y \leq y\right] = \int_{-\infty}^{x} \sum_{w \leq y} p_{XY}\left(u, w; \boldsymbol{\theta}\right) \mathrm{d}u = \int_{-\infty}^{x} \int_{-\infty}^{\tau_y} p_{X\eta}\left(u, v; \boldsymbol{\theta}\right) \mathrm{d}v \mathrm{d}u, \tag{3.4}$$

[2]Being a random vector comprising a continuous and discrete variable, (X, Y) technically does not have a joint density in a regular sense, that is, a Randon–Nikodým derivative with respect to the two-dimensional Lebesgue measure. The formally correct way to refer to p_{XY} in (3.3) would be that it is the Randon–Nikodým derivative of the distribution P_{XY} in (3.4) with respect to the product of the one-dimensional Lebesgue measure and the counting measure. For the sake of brevity, however, we refer to p_{XY} simply as joint density in this article, acknowledging an abuse of terminology.

https://doi.org/10.1017/psy.2026.10091 Published online by Cambridge University Press

where the third equality follows from (3.3) in conjunction with the linearity of the integral operator. As such, $P_{XY}(x,y;\boldsymbol{\theta})$ is the distribution function associated with density $p_{XY}(x,y;\boldsymbol{\theta})$, so we refer to it as the *polyserial model distribution*.

Our expressions for the polyserial model distribution and density are different but equivalent to the more commonly used expressions in Olsson et al. (1982), which are provided in Section A of the Supplementary Material.

3.2. Point polyserial correlation

In addition to the correlation between the observed X and the latent η, one might also be interested in the correlation between X and the observed ordinal Y. The correlation between the observed X and Y is known as *point polyserial correlation*. In order to identify the desired point polyserial correlation $\mathbb{C}\mathrm{or}[X,\,Y]$, one needs to assign a numerical interpretation to the r answer categories of Y, that is, introduce a *scoring system*. Given a scoring system, the point polyserial correlation coefficient $\widetilde{\rho} = \mathbb{C}\mathrm{or}[X,\,Y]$ is identified by the polyserial model and can be estimated by using estimates of the polyserial model parameters $\boldsymbol{\theta}$. We provide details in Section A of the Supplementary Material.

3.3. Maximum likelihood estimation

Suppose we observe a sample $\{(X_i,Y_i)\}_{i=1}^N$ of N independent copies of (X,Y) generated by the polyserial model at some *true* parameter vector $\boldsymbol{\theta}_* = \left(\rho_*,\mu_*,\sigma_*^2,\boldsymbol{\tau}_*^\top\right)^\top$. The statistical problem is to estimate the true $\boldsymbol{\theta}_*$ from the observed sample, which is traditionally achieved by the ML estimator proposed by Cox (1974) and Olsson et al. (1982).

The ML estimator (MLE) of $\boldsymbol{\theta}_*$ is defined as the log-likelihood maximizer

$$\widehat{\boldsymbol{\theta}}_N^{\mathrm{MLE}} = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\{ \sum_{i=1}^N \log\left(p_{XY}(X_i,Y_i;\boldsymbol{\theta})\right) \right\}, \tag{3.5}$$

where the parameter space

$$\boldsymbol{\Theta} = \left\{ \left(\rho,\mu,\sigma^2,\boldsymbol{\tau}^\top\right)^\top \;\middle|\; \rho\in(-1,1),\ \mu\in\mathbb{R},\ \sigma>0,\ -\infty<\tau_1<\cdots<\tau_{r-1}<+\infty \right\} \tag{3.6}$$

is the set of legal parameters $\boldsymbol{\theta}$ the MLE maximizes over. As such, $\boldsymbol{\Theta}$ rules out degenerate cases, such as $\rho=\pm1$, non-positive standard deviations (SDs), or non-monotonic thresholds. Assuming that the polyserial model is correctly specified for the data at hand, Cox (1974) and Olsson et al. (1982) show that the MLE is consistent for the true $\boldsymbol{\theta}_*$, asymptotically normally distributed, and fully efficient.

As a computationally attractive alternative to ML estimation, Olsson et al. (1982) propose a two-step (TS) estimation procedure. In the first step, one computes as estimators of μ, σ^2, and $\tau_k, k=1,\ldots,r-1$, the sample statistics

$$\widehat{\mu}_{\mathrm{TS}} = \frac{1}{N}\sum_{i=1}^N X_i, \quad \widehat{\sigma}^2_{\mathrm{TS}} = \frac{1}{N-1}\sum_{i=1}^N (X_i-\widehat{\mu}_{\mathrm{TS}})^2, \quad \widehat{\tau}_{k,\mathrm{TS}} = \Phi^{-1}\left(\frac{1}{N}\sum_{j=1}^k N_{Y,j}\right), \tag{3.7}$$

respectively, where $N_{Y,j} = \sum_{i=1}^N \mathbb{1}\{Y_i=j\}$ denotes the empirical marginal frequency of the j-th response option of Y. In the second step, one substitutes for these sample statistics in the log-likelihood in (3.5) and maximizes the ensuing log-likelihood with respect to the remaining parameter, ρ, via conditional ML (conditional on the sample statistics). The main advantage of the TS approach is reduced computing time because one needs to numerically solve the maximization problem (3.5) only with respect to one parameter, ρ, rather than all d parameters in $\boldsymbol{\theta}$. A drawback is diminished efficiency as compared to (joint) ML, where all d parameters are estimated simultaneously. By means of simulation experiments, Olsson et al. (1982) find that if the polyserial model is correctly specified, ML and the TS estimator produce similar results, with differences decreasing with an increasing number of response options

https://doi.org/10.1017/psy.2026.10091 Published online by Cambridge University Press

for Y. The TS approach is the default estimator for polyserial correlation in the SEM software package `lavaan` (Rosseel, 2012).

While it is in principle possible to robustify the TS approach against contamination by using the same estimation theory as in this article, doing so would result in substantial theoretical and computational drawbacks. We discuss this in detail in Section C.4 of the Supplementary Material.

Alternative estimators of polyserial correlation have been proposed by Bedrick and Breslin (1996), Lord (1963), and Brogden (1949), with Bedrick (1990, 1992), Koopman (1983) and Kraemer (1981) studying the theoretical properties of the latter two. We discuss these approaches in more detail in Section 4.2, after conceptualizing misspecification of the polyserial model.

4. Model misspecification

In order to study the effects of partial model misspecification, we first define this concept and then explain how it differs from distributional misspecification.

4.1. Partial misspecification of the polyserial model

The polyserial model is misspecified if at least one observation for the continuous-ordinal variable pair (X, Y) has not been generated by the bivariate partially-latent normality model in (3.2). Akin to Welz et al. (2026b), we consider a *partial misspecification* framework where only a fraction $(1-\varepsilon)$ of observations in a given sample are generated by an underlying normal distribution $P_{X\eta}$ with true parameter $\boldsymbol{\theta}_*$, whereas a fixed but unknown fraction ε of observations are generated from some different but unspecified underlying distribution $H_{X\eta}$.[3] Since $H_{X\eta}$ is unspecified, its correlation structure may differ from $P_{X\eta}$ so that observations (X, Y) generated by the underlying $H_{X\eta}$ (after discretization) may be uninformative for the true polyserial correlation coefficient.

Formally, the polyserial model is said to be *partially* misspecified if the unknown sampling distribution of the observed-latent variable pair (X, η) is given by

$$F_{\varepsilon,X\eta}(x,v) = (1-\varepsilon)P_{X\eta}(x,v;\boldsymbol{\theta}_*) + \varepsilon H_{X\eta}(x,v) \tag{4.1}$$

for $x, v \in \mathbb{R}$. Correspondingly, the implied unknown sampling distribution of the observed variables (X, Y) reads

$$F_{\varepsilon,XY}(x,y) = (1-\varepsilon)P_{XY}(x,y;\boldsymbol{\theta}_*) + \varepsilon H_{XY}(x,y) \tag{4.2}$$

for $x \in \mathbb{R}, y \in \mathcal{Y}$, where

$$H_{XY}(x,y) = \int_{-\infty}^{\tau_{\varepsilon,y}} H_{X\eta}(x,\mathrm{d}v)$$

is the distribution of the observations for which the polyserial model is (partially) misspecified. The unknown and unspecified discretization thresholds $-\infty = \tau_{\varepsilon,0} < \tau_{\varepsilon,1} < \cdots < \tau_{\varepsilon,r-1} < \tau_{\varepsilon,r} = +\infty$ need not equal the true discretization thresholds $\tau_{*,1} < \cdots < \tau_{*,r-1}$ of the polyserial model. Furthermore, denote by $f_{\varepsilon,XY}$ the unknown density corresponding to the sampling distribution $F_{\varepsilon,XY}$.

Misspecification models of the type in (4.1) are standard in the robust statistics literature, where they are known as *Huber contamination models*, owing to pioneering work of Huber (1964). Following Welz et al. (2026b), we therefore adopt terminology from robust statistics and call ε the *contamination fraction*, the uninformative $H_{X\eta}$ the *contamination distribution* (or simply *contamination*), and $F_{\varepsilon,X\eta}$ the *contaminated* distribution. In the case of a zero-valued contamination fraction ($\varepsilon = 0$), there is no misspecification so that the polyserial model is correctly specified. Indeed, if $\varepsilon = 0$, then $F_{0,X\eta}(\cdot,\cdot) =$

[3] Being a population quantity and not a sample quantity, it is strictly speaking imprecise to call ε a fraction. It would be more accurate to call ε a (contamination) population probability with which one samples from the contamination distribution $H_{X\eta}$ instead of the model distribution $P_{X\eta}$. However, to be consistent with robust statistics literature, we refer to ε as a fraction throughout this article.

https://doi.org/10.1017/psy.2026.10091 Published online by Cambridge University Press

$P_{X\eta}(\cdot,\cdot;\boldsymbol{\theta}_*)$ and $F_{0,XY}(\cdot,\cdot) = P_{XY}(\cdot,\cdot;\boldsymbol{\theta}_*)$, so the partial misspecification framework in (4.1) nests the correctly specified polyserial model.

Neither the contamination fraction ε nor the contamination distribution $H_{X\eta}$ in (4.1) is assumed to be known. Thus, both quantities are left completely unspecified in practice, which *"means that we are not making any assumptions on the degree, magnitude, or type of contamination (which is possibly absent altogether)"* (Welz et al., 2026b). Hence, the polyserial model may be misspecified due to an unlimited variety of reasons, for instance, but not limited to outliers or nonnormality in the continuous X, and/or careless responding or item misunderstanding in the ordinal Y. Because ε and $H_{X\eta}$ are unspecified in practice, the polyserial model distribution $P_{XY}(\cdot,\cdot;\boldsymbol{\theta}_*)$ remains the distribution of interest: Our only aim is to estimate the parameter $\boldsymbol{\theta}_*$ of the polyserial model while reducing the impact of potential contamination in the observed data. Consequently, the contaminated distributions $F_{\varepsilon,XY}$ and $F_{\varepsilon,X\eta}$ are never estimated. They are purely theoretical objects to study the properties of estimators of the polyserial model when that model is partially misspecified due to data contamination.

While no assumption is made on the specific value of the contamination fraction ε, we impose the restriction $\varepsilon \in [0,0.5)$, which is common in robust statistics (e.g., Hampel et al., 1986, p. 67). This restriction ensures proper identification: If half or more of the data points were contaminated, it would no longer be possible to distinguish between contamination and observations generated by the polyserial model, at least not without further assumptions. Under such additional assumptions, also values of $\varepsilon \geq 0.5$ could be considered. We refer to Welz et al. (2026b) for a more detailed discussion.

Figure 1 visualizes a simulated example data set generated by the contaminated distribution $F_{\varepsilon,X\eta}$ in (4.1) with contamination fraction $\varepsilon = 0.15$. In this example, the contamination distribution $H_{X\eta}$, whose random draws are orange dots, is a shifted bivariate t-distribution with noncentrality parameter set to $(10,-2)^\top$, scale matrix $\mathrm{diag}(0.25,0.25)$, and 10 degrees of freedom. The remaining realizations (gray dots) are generated by the bivariate normal distribution $P_{X\eta}$ with true parameters $\mu_* = 0$, $\sigma_*^2 = 1$, and $\rho_* = 0.5$. Here, contamination manifests through somewhat larger values in the X-dimension and inflation of the first two response options in the Y-dimension (after discretization), resulting in a (partially) misspecified polyserial model. Applying the MLE in (3.5) to the observed data for (X,Y) in Figure 1 yields a sign-flipped polyserial correlation estimate of -0.522, representing a substantial bias with respect to the true value of $\rho_* = 0.5$. In contrast, the robust estimator—which will be introduced in Section 5 and uses the exact same information as ML—estimates a correlation of 0.498, which is accurate for the true value of $\rho_* = 0.5$.

We stress that there exist nonnormal joint distributions of (X,η) that, after discretizing the latent η variable with fixed thresholds, result in the same density for the observed (X,Y) as if the partially-latent (X,η) were jointly normal. This phenomenon is known as *discretize equivalence* of latent variables (Foldnes & Grønneberg, 2019). It follows that there exist contamination distributions $H_{X\eta}$ and contamination fractions $\varepsilon > 0$ in (4.1) under which the contaminated density $f_{\varepsilon,XY}$ of the observed (X,Y) is exactly equal to the polyserial model density in (3.3) at the true parameter, that is, $f_{\varepsilon,XY}(x,y) = p_{XY}(x,y;\boldsymbol{\theta}_*)$ for all $(x,y) \in \mathbb{R} \times \mathcal{Y}$. Hence, in this scenario, the polyserial model is misspecified, but the misspecification does not have adverse effects because the true polyserial model density of (X,Y) remains unaffected. To avoid cumbersome notation in our analysis, we follow Welz et al. (2026b) and assume *consequential* misspecification throughout this article, that is, $f_{\varepsilon,XY}(x,y) \neq p_{XY}(x,y;\boldsymbol{\theta}_*)$ for at least one response $y \in \mathcal{Y}$ whenever $\varepsilon > 0$, given $x \in \mathbb{R}$. Nevertheless, *"it is silently understood that misspecification need not be consequential"* (Welz et al., 2026b), in which case there is no practical problem and both ML and robust estimator are consistent for the true $\boldsymbol{\theta}_*$.

Before we define our robust estimator, we briefly juxtapose the partial misspecification framework to distributional misspecification.

4.2. Distributional misspecification

The polyserial model is said to be *distributionally* misspecified if it is misspecified for *all* observations in a sample. This contrasts the partial misspecification framework in (4.1) where the model is only

https://doi.org/10.1017/psy.2026.10091 Published online by Cambridge University Press

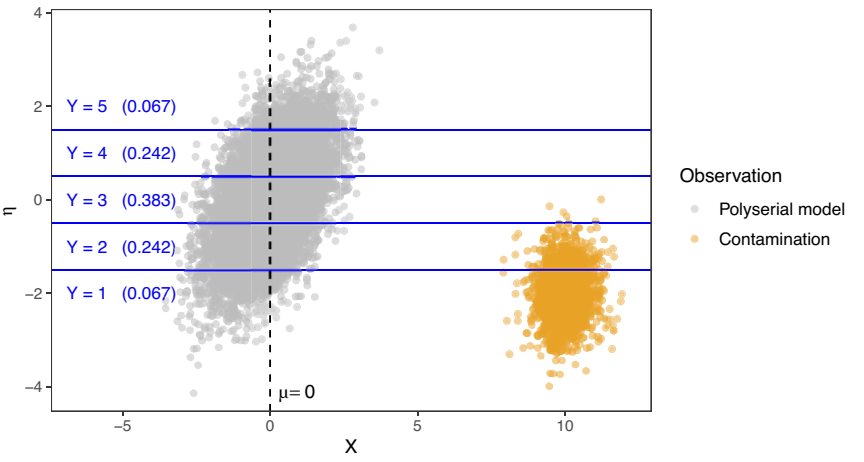


Figure 1. Simulated data where the polyserial model is misspecified for a fraction $\varepsilon = 0.15$ of the $N = 10{,}000$ points. The gray dots are draws of (X, η) from the polyserial model with true parameters $\rho = 0.5, \mu = 0, \sigma^2 = 1$, while the orange dots are draws from a contamination distribution $H_{X\eta}$, being a bivariate t-distribution here with noncentrality parameter $(10, -2)^\top$, scale matrix $\text{diag}(0.25, 0.25)$, and 10 degrees of freedom. The horizontal lines mark the thresholds that discretize the latent η to the observed ordinal Y with five response options. The numbers in parentheses indicate the population marginal probability of the respective response option under the true polyserial model.

misspecified for a (possibly zero-valued) fraction of the observations. Let $G = G_{X,\eta}$ denote the unknown joint distribution of X and the latent η, which, under distributional misspecification, is nonnormal for all data points. In distributional misspecification, the object of interest is the correlation coefficient between X and η under the distribution G, denoted $\rho_G = \mathbb{C}\text{or}_G[X, \eta]$, instead of the correlation coefficient under bivariate normality (which would be the polyserial correlation coefficient). Although our robust estimator is designed for partial misspecification rather than distributional misspecification, it can in certain situations also offer a robustness gain under distributional misspecification. We discuss this in more detail in Section 7.3.

5. Robust estimation of polyserial correlation

It is well-known that ML estimation is highly susceptible to partial model misspecification due to data contamination (e.g., Hampel et al., 1986; Huber, 1964; Huber & Ronchetti, 2009), so it might be desirable to construct alternative estimators that are more robust against contamination. However, attempting to robustify estimation of the polyserial model against contamination bears the inherent challenge of how to handle the mixed-type structure of the data with one continuous and one ordinal variable.

While there exist general frameworks for robust estimation with exclusively continuous variables and exclusively categorical variables, none of them are applicable to mixed variables. On the one hand, robust M-estimation as proposed by Huber (1964) and related methods (e.g., Huber & Ronchetti, 2009) are intended for continuous random variables, and, broadly speaking, achieve robustness by downweighting observations with extreme values. For instance, Alfons et al. (2022) use a variation of M-estimation, MM-estimation, to robustify mediation analyses against outliers, heavy-tailedness, or skewness of the observed distribution. Such estimators are not applicable to mixed data because the ordinal variable cannot take extreme values due to being categorical, nor does it admit a direct numerical interpretation to begin with. On the other hand, the theory of robust C-estimation (Welz, 2024) is designed for exclusively categorical variables. C-estimation downweighs categories whose empirical frequency disagrees with their corresponding theoretical frequency under a postulated model. Since one variable in the polyserial model is continuous, it does not have discrete categories, so C-estimators

https://doi.org/10.1017/psy.2026.10091 Published online by Cambridge University Press

cannot be applied. However, it turns out that the fundamental idea of downweighting data points that cannot be modeled sufficiently well by a postulated model can be exploited to achieve robustness through minimum DPD estimation (Basu et al., 1998). We explain in this section how minimum DPD estimation can be applied to the polyserial model.

For further reference, define the Kullback–Leibler (KL) divergence (Kullback & Leibler, 1951) between two bivariate densities g_1 and g_2, each defined on $\mathbb{R}^2$ (or a common subset thereof), by

$$\mathrm{KL}\left(g_1 \,\|\, g_2\right) = \int \int g_1(s,t) \log\left(\frac{g_1(s,t)}{g_2(s,t)}\right) \mathrm{d}t\mathrm{d}s, \tag{5.1}$$

where the integrals are taken over the densities' domain. If the second dimension of the densities corresponds to a discrete random variable, that is, the first variable is continuous, but the second one is discrete, replace the inner integral in (5.1) by a summation over the second variable's domain.[4]

Throughout this section, suppose that one has access to a random sample of N independent continuous-ordinal variable pairs, $(X_i, Y_i), i = 1, \ldots, N$, following the unknown sampling distribution $F_{\varepsilon,XY}$ in (4.2). Hence, the polyserial model is possibly misspecified for an unknown fraction ε of the observed sample.

5.1. Maximum likelihood revisited

The observed sample can be uniquely characterized by a particular empirical density function

$$\widehat{f}_N(x,y) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}\{X_i = x, Y_i = y\} \tag{5.2}$$

for $x \in \mathbb{R}, y \in \mathcal{Y}$, where the indicator function $\mathbb{1}\{E\}$ takes value 1 if an event E is true, and 0 otherwise. As such, the empirical density $\widehat{f}_N$ only takes nonzero values when evaluated at points in the observed sample $\{(X_i, Y_i)\}_{i=1}^N$. The MLE $\widehat{\boldsymbol{\theta}}_N^{\mathrm{MLE}}$ in (3.5) can be expressed as the minimizer of the KL divergence between the empirical density $\widehat{f}_N$ and model density $p_{XY}(\cdot,\cdot;\boldsymbol{\theta})$, which is given by

$$\mathrm{KL}\left(\widehat{f}_N \,\|\, p_{XY}(\cdot,\cdot;\boldsymbol{\theta})\right) = \frac{1}{N} \sum_{i=1}^{N} \left(\log\left(\widehat{f}_N(X_i, Y_i)\right) - \log\left(p_{XY}(X_i, Y_i; \boldsymbol{\theta})\right) \right) \tag{5.3}$$

and follows by definition of $\widehat{f}_N$ and the KL divergence in (5.1) as well as the convention $0\log(0) = 0$.[5] We refer to White (1982) for a detailed exposition of the connection between KL divergence and ML.

As alluded to earlier, ML estimation tends to be highly susceptible to contamination in the observed sample. It may therefore be desirable to consider alternatives that are designed to more robust against contamination. It turns out that the idea of minimizing a given divergence between the empirical density $\widehat{f}_N$ and the model density $p_{XY}(\cdot,\cdot;\boldsymbol{\theta})$ can be exploited to construct contamination-robust estimators by choosing alternative divergences to the KL divergence. This is exactly what *minimum power divergence* estimation (Basu et al., 1998) does.

5.2. Minimum power divergence

Basu et al. (1998) propose a class of divergences between two densities that generalizes the KL divergence, but is less affected by data contamination. Suppose g_1 and g_2 are bivariate densities supported

[4] To explicitly accommodate discrete distributions in the definition of the KL divergence, we could have adapted the Riemann-integral-based definition in (5.1) to a definition based on Lebesgue integration with respect to a general measure that can either be the Lebesgue measure (for continuous distributions) or the counting measure (for discrete distributions). However, we believe that is more instructive to refrain from introducing measure-theoretic concepts in this article.

[5] Alternatively, the MLE (and also the proposed robust estimator) can be equivalently defined using statistical divergence functionals, as done in Basu et al. (1998).

https://doi.org/10.1017/psy.2026.10091 Published online by Cambridge University Press

on $\mathbb{R}^2$ or a common subset thereof. For a fixed tuning constant $\alpha > 0$, the divergence of Basu et al. (1998) between g_1 and g_2 is defined by

$$\mathrm{D}_\alpha\left(g_1 \,\|\, g_2\right) = \int\int \left(g_2^{1+\alpha}(s,t) - \left(1+\frac{1}{\alpha}\right) g_1(s,t) g_2^{\alpha}(s,t) + \frac{1}{\alpha} g_1^{1+\alpha}(s,t) \right) \mathrm{d}t\mathrm{d}s,$$

where each integral is taken over the densities' domain. Note that the divergence D_α is not defined at $\alpha = 0$. To overcome this issue, Basu et al. (1998) define D_0 as the limit of D_α as $\alpha \downarrow 0$, that is,

$$\mathrm{D}_0\left(g_1 \,\|\, g_2\right) = \lim_{\alpha \downarrow 0} \mathrm{D}_\alpha\left(g_1 \,\|\, g_2\right) = \int\int g_1(s,t) \log\left(\frac{g_1(s,t)}{g_2(s,t)}\right) \mathrm{d}t\mathrm{d}s = \mathrm{KL}\left(g_1 \,\|\, g_2\right), \tag{5.4}$$

where the second equality follows from the identity $z \mapsto \log z = \lim_{\alpha \downarrow 0} \alpha^{-1}(z^\alpha - 1)$ and the third equality from the definition of the KL divergence in (5.1). Basu et al. (1998) call the divergence $\mathrm{D}_\alpha, \alpha \geq 0$, the DPD. It follows that the DPD generalizes the KL divergence. We explain in a moment the intuition behind the DPD.

As before, if the second dimension of the two densities g_1, g_2 corresponds to a discrete random variable, that is, the first variable in g_j is continuous, but the second one is discrete, replace the inner integral in the DPD $\mathrm{D}_\alpha, \alpha \geq 0$, by a summation over the second variable's domain.

5.3. Proposed robust estimator

For $\alpha > 0$, the DPD between the empirical density $\widehat{f}_N$ in (5.2) and the polyserial model's density $p_{XY}(\cdot,\cdot;\boldsymbol{\theta})$ reads

$$\mathrm{D}_\alpha\left(\widehat{f}_N \,\|\, p_{XY}(\cdot,\cdot;\boldsymbol{\theta})\right) = \int_{\mathbb{R}} \sum_{y \in \mathcal{Y}} p_{XY}^{1+\alpha}(x,y;\boldsymbol{\theta})\, \mathrm{d}x - (1+\alpha^{-1}) \frac{1}{N} \sum_{i=1}^{N} p_{XY}^{\alpha}(X_i, Y_i; \boldsymbol{\theta}) + \alpha^{-1}, \tag{5.5}$$

which follows from writing out $\widehat{f}_N$. For the choice $\alpha = 0$, the DPD $\mathrm{D}_0\left(\widehat{f}_N \,\|\, p_{XY}(\cdot,\cdot;\boldsymbol{\theta})\right)$ reduces to the KL divergence in (5.3) by (5.4).

For a pre-specified tuning constant $\alpha \geq 0$, our proposed estimator minimizes the divergence D_α between $\widehat{f}_N$ and $p_{XY}(\cdot,\cdot;\boldsymbol{\theta})$ with respect to the model parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Specifically, our proposed estimator is given by

$$\widehat{\boldsymbol{\theta}}_N = \arg\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \mathrm{D}_\alpha\left(\widehat{f}_N \,\|\, p_{XY}(\cdot,\cdot;\boldsymbol{\theta})\right). \tag{5.6}$$

In particular, if $\alpha = 0$, the estimator $\widehat{\boldsymbol{\theta}}_N$ coincides with the MLE $\widehat{\boldsymbol{\theta}}_N^{\mathrm{MLE}}$ (see Section 5.1). Since it minimizes a DPD, estimator $\widehat{\boldsymbol{\theta}}_N$ constitutes a *minimum DPD estimator* (Basu et al., 1998). As such, a limit theory for $\widehat{\boldsymbol{\theta}}_N$ is readily available from Basu et al. (1998), who derive the theoretical properties of such estimators for general models. Before we turn to the estimator's limit theory, we first provide some intuition as to *why* minimum DPD estimators are more robust than ML whenever $\alpha > 0$.

An estimator defined as the minimum of a (differentiable) loss function can be equivalently characterized as a root of the loss' gradient; a characterization known as *estimating equation*. The estimating equation of the MLE $(\alpha = 0)$ can be written as

$$\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{u}_{\boldsymbol{\theta}}(X_i, Y_i) = 0,$$

which is satisfied for $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_N^{\mathrm{MLE}}$, and where the d-vector

$$\boldsymbol{u}_{\boldsymbol{\theta}}(x,y) = \frac{\partial}{\partial \boldsymbol{\theta}} \log p_{XY}(x,y;\boldsymbol{\theta}), \qquad x \in \mathbb{R}, y \in \mathcal{Y},$$

denotes the log-likelihood *score* function. A closed-form expression of $\boldsymbol{u}_{\boldsymbol{\theta}}(x,y)$ is provided in Section A of the Supplementary Material. Analogously, the estimating equation of the proposed estimator $\widehat{\boldsymbol{\theta}}_N$

https://doi.org/10.1017/psy.2026.10091 Published online by Cambridge University Press

in (5.6) can be shown to read

$$\frac{1}{N}\sum_{i=1}^{N} p_{XY}^{\alpha}(X_i, Y_i; \boldsymbol{\theta})\, \boldsymbol{u}_{\boldsymbol{\theta}}(X_i, Y_i) - \boldsymbol{c}_{\alpha}(\boldsymbol{\theta}) = 0, \tag{5.7}$$

which is satisfied for $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_N$ and where $\boldsymbol{c}_{\alpha}(\boldsymbol{\theta}) = \int_{\mathbb{R}} \sum_{y \in \mathcal{Y}} p_{XY}^{1+\alpha}(x, y; \boldsymbol{\theta})\, \boldsymbol{u}_{\boldsymbol{\theta}}(x, y)\, \mathrm{d}x$ is a correction factor independent of observed data. Observe that for $\alpha = 0$, the estimating equation (5.7) equals that of the MLE,[6] in which all observations in the sample $\{(X_i, Y_i)\}_{i=1}^{N}$ are weighted equally. Conversely, when $\alpha > 0$, the term

$$\frac{1}{N}\sum_{i=1}^{N} p_{XY}^{\alpha}(X_i, Y_i; \boldsymbol{\theta})\, \boldsymbol{u}_{\boldsymbol{\theta}}(X_i, Y_i)$$

in estimating equation (5.7) *"provides a relative-to-the-model downweighting for [contaminated][7] observations"* (Basu et al., 1998, p. 551) with individual-specific weights

$$\widetilde{w}_{i,\alpha}(\boldsymbol{\theta}) = p_{XY}^{\alpha}(X_i, Y_i; \boldsymbol{\theta}). \tag{5.8}$$

A weight $\widetilde{w}_{i,\alpha}(\boldsymbol{\theta})$ is bounded from below by 0 and takes values close to 0 for observations that the polyserial model cannot fit well, such as contaminated data points. As such, *"observations that are widely discrepant to the [polyserial] model will get nearly zero weights"* (Basu et al., 1998, pp. 551–552). Larger choices of $\alpha > 0$ lead to more stringent downweighting. For instance, a contaminated observation with an ordinal response Y_i that strongly disagrees with the latent threshold structure implied by the model density p_{XY} would receive a weight close to 0.

If contamination is absent $(\varepsilon = 0)$, it can be shown that the robust estimator $\widehat{\boldsymbol{\theta}}_N$ and the MLE $\widehat{\boldsymbol{\theta}}_N^{\mathrm{MLE}}$ in (3.5) are asymptotically equal to one another for all choices $\alpha \geq 0$. This property is implied by the property of *Fisher consistency* that minimum DPD estimators possess (implied by Theorem 1 in Basu et al., 1998, see page 551 therein).

Being expressible as a root of a summation over the sample where each summand is a certain function evaluated at a single observation (Eq. (5.7)), minimum DPD estimators are a special case of M-estimation (see Basu et al., 1998, Section 3.1). However, unlike classic M-estimation as proposed in Huber (1964), minimum DPD estimation does not require the data to admit a numerical interpretation due to operationalizing "outlyingness" through discrepant model fit rather than extreme values. Consequently, it can be directly applied to models for mixed data, such as the polyserial model.

5.4. Rescaling of weights

By construction, for a given tuning constant $\alpha > 0$, the robust estimator's raw weights $\widetilde{w}_{i,\alpha}(\boldsymbol{\theta})$ in (5.8) are bounded from below by 0 and from above by some data-independent finite value $M_{\alpha}(\boldsymbol{\theta})$ only depending on α and $\boldsymbol{\theta}$, which is defined by $M_{\alpha}(\boldsymbol{\theta}) = \sup\{p_{XY}^{\alpha}(x, y; \boldsymbol{\theta}) : x \in \mathbb{R}, y \in \mathcal{Y}\}$. The fact that the upper bound $M_{\alpha}(\boldsymbol{\theta})$ varies for different values of $\boldsymbol{\theta}$ and α makes it difficult to compare raw weights across different estimates and specifications. Since the raw weights are intended as a tool for detecting contaminated observations, it would be beneficial to have such a comparability. In the following, we describe a simple novel procedure to rescale the weights to always be bounded from above by 1. Rescale the raw weights by their upper bound $M_{\alpha}(\boldsymbol{\theta})$ to obtain

$$w_{i,\alpha}(\boldsymbol{\theta}) = \widetilde{w}_{i,\alpha}(\boldsymbol{\theta}) \big/ M_{\alpha}(\boldsymbol{\theta}), \qquad i = 1, \ldots, N. \tag{5.9}$$

[6] By elementary likelihood theory, $\boldsymbol{c}_0(\boldsymbol{\theta}) = \int_{\mathbb{R}} \sum_{y \in \mathcal{Y}} p_{XY}(x, y; \boldsymbol{\theta})\, \boldsymbol{u}_{\boldsymbol{\theta}}(x, y)\, \mathrm{d}x = \mathbb{E}_{P_{XY}}[\boldsymbol{u}_{\boldsymbol{\theta}}(X, Y)] = 0$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

[7] The original quote of Basu et al. (1998, p. 551) refers to contaminated observations as *outlying* observations. In their terminology, an outlier is any data point that was not generated by the postulated model. However, since one often associates outliers with extreme values rather than discordant model fit, we believe that the term "outlying" is potentially misleading, so we prefer "contaminated."

https://doi.org/10.1017/psy.2026.10091 Published online by Cambridge University Press

We henceforth refer to the rescaled weights $w_{i,\alpha}(\boldsymbol{\theta})$ simply as *weights*. The rescaled weights are contained in the interval $[0,1]$ and take values close to 0 for poorly fitting observations and value 1 for perfectly fitting observations. This rescaling is without loss of generality because the estimating equation (5.7) can equivalently be expressed as

$$\frac{1}{N}\sum_{i=1}^{N} w_{i,\alpha}(\boldsymbol{\theta})\boldsymbol{u}_{\boldsymbol{\theta}}(X_i,Y_i) - \boldsymbol{c}_{\alpha}(\boldsymbol{\theta})/M_{\alpha}(\boldsymbol{\theta}) = 0.$$

Section B of the Supplementary Material describes an algorithm for computing the upper bound $M_{\alpha}(\boldsymbol{\theta})$. The usefulness of the rescaled weights as a tool for identifying contaminated data points will be demonstrated in an empirical application in Section 8, where they successfully detect observations discrepant from the model.

6. Statistical and computational properties

This section is devoted to the properties of minimum DPD estimators for estimating polyserial models. We first discuss their estimands under contamination, then their asymptotic properties and efficiency properties, and then computational aspects as well as our software implementation.

6.1. Estimand

It is instructive to discuss what the proposed minimum DPD estimator $\widehat{\boldsymbol{\theta}}_N$ in (5.6) actually estimates. Recall that $f_{\varepsilon,XY}$ denotes the density of the sampling distribution $F_{\varepsilon,XY}$ in (4.2) that the observed sample follows. Just like $F_{\varepsilon,XY}$, the population density $f_{\varepsilon,XY}$ is completely unspecified and unknown in practice, including contamination fraction ε. For a fixed tuning constant $\alpha \geq 0$, the estimand $\boldsymbol{\theta}_0$ of estimator $\widehat{\boldsymbol{\theta}}_N$ is the parameter that minimizes the DPD between the population density and the polyserial model density, that is,

$$\boldsymbol{\theta}_0 = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \mathrm{D}_{\alpha}\left(f_{\varepsilon,XY} \,\middle\|\, p_{XY}(\cdot,\cdot;\boldsymbol{\theta})\right).$$

Observe that $\mathrm{D}_{\alpha}\left(f_{\varepsilon,XY} \,\middle\|\, p_{XY}(\cdot,\cdot;\boldsymbol{\theta})\right)$ is simply the population analog of the empirical divergence $\mathrm{D}_{\alpha}\left(\widehat{f}_N \,\middle\|\, p_{XY}(\cdot,\cdot;\boldsymbol{\theta})\right)$ that is minimized by the estimator $\widehat{\boldsymbol{\theta}}_N$ in (5.6).

In the absence of contamination ($\varepsilon = 0$), one has that $f_{0,XY}(\cdot,\cdot) = p_{XY}(\cdot,\cdot;\boldsymbol{\theta}_*)$ by (4.2), and it follows from Theorem 1 in Basu et al. (1998) that $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_*$, so that the population divergence is zero-valued. Subsequently, just like the MLE, $\widehat{\boldsymbol{\theta}}_N$ is unbiased for the true $\boldsymbol{\theta}_*$ (in population) if the model is correctly specified. This property is known as *Fisher consistency*. We stress that Fisher consistency holds true for *all* choices of $\alpha \geq 0$. In other words, if contamination is absent, then the proposed estimator is unbiased (in population) no matter the choice of tuning constant α.

On the other hand, if contamination is present ($\varepsilon > 0$), then the estimand $\boldsymbol{\theta}_0$ will generally differ from the true $\boldsymbol{\theta}_*$. How much they differ depends on the contamination fraction ε and contamination type $H_{X\eta}$ in the contaminated distribution (4.1), as well as the choice of tuning constant $\alpha \geq 0$. Roughly speaking, holding the unknown contaminated distribution $F_{\varepsilon,X\eta}$ constant, the larger α, the closer $\boldsymbol{\theta}_0$ tends to be to the true $\boldsymbol{\theta}_*$.[8] In other words, if the polyserial model is misspecified, larger values of α asymptotically tend to lead to less bias for the true $\boldsymbol{\theta}_*$.

[8] We stress that this sentence is a heuristic rather than a formal result. Since no assumption is made on the contamination distribution $H_{X\eta}$, there may exist contaminated distributions $F_{\varepsilon,X\eta}$ for which higher values of α do *not* uniformly reduce bias. We refer to Hampel et al. (1986, Chapter 8.2d) for a discussion on the fundamental difficulty of deriving bias-reduction results of robust procedures in contamination models such as (4.1).

https://doi.org/10.1017/psy.2026.10091 Published online by Cambridge University Press

6.2. Asymptotic analysis

We are now ready to study the asymptotic properties of the proposed estimator $\widehat{\boldsymbol{\theta}}_N$ in (5.6). Under mild standard regularity conditions, it can be shown that $\widehat{\boldsymbol{\theta}}_N$ is asymptotically consistent for estimand $\boldsymbol{\theta}_0$, that is,

$$\widehat{\boldsymbol{\theta}}_N \xrightarrow{\mathbb{P}} \boldsymbol{\theta}_0,$$

as $N \to \infty$, and is furthermore asymptotically Gaussian,

$$\sqrt{N}\left(\widehat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_0\right) \xrightarrow{\mathrm{d}} \mathrm{N}_d\Big(0, \boldsymbol{\Sigma}\left(\boldsymbol{\theta}_0\right)\Big),$$

as $N \to \infty$, where "$\xrightarrow{\mathbb{P}}$" and "$\xrightarrow{\mathrm{d}}$" denote convergence in probability and distribution, respectively. These two stochastic convergence results are rigorously established in Theorem A.1 in the Supplementary Material. The theorem follows immediately from a more general result in Basu et al. (1998). The asymptotic covariance matrix $\boldsymbol{\Sigma}\left(\boldsymbol{\theta}_0\right)$ has a closed-form sandwich-type construction

$$\boldsymbol{\Sigma}\left(\boldsymbol{\theta}_0\right) = \boldsymbol{J}\left(\boldsymbol{\theta}_0\right)^{-1} \boldsymbol{K}\left(\boldsymbol{\theta}_0\right) \boldsymbol{J}\left(\boldsymbol{\theta}_0\right)^{-1}, \tag{6.1}$$

where the $d \times d$ full-rank matrices $\boldsymbol{\theta} \mapsto \boldsymbol{J}\left(\boldsymbol{\theta}\right), \boldsymbol{K}\left(\boldsymbol{\theta}\right)$ are defined in Section A of the Supplementary Material. Neither $\boldsymbol{J}\left(\boldsymbol{\theta}\right)$ nor $\boldsymbol{K}\left(\boldsymbol{\theta}\right)$ are observed in practice, so neither is $\boldsymbol{\Sigma}\left(\boldsymbol{\theta}_0\right)$, but each of these matrices can be consistently estimated, as we will explain momentarily. Owing to this asymptotic normality result, one can construct standard errors and confidence intervals for the estimand $\boldsymbol{\theta}_0$. We stress that the estimator remains asymptotically normal even when the polyserial model is misspecified. Similar results for inference in general misspecified models with sandwich-type covariance constructions have been derived by, for example, White (1982) and Huber (1967) for the MLE, Basu et al. (1998) for robust minimum DPD estimators (of which our proposed estimator is a special case), and Welz (2024) for a class of robust estimators for general categorical data.

Being a function of the unobserved estimand $\boldsymbol{\theta}_0$ as well as additional unobserved population quantities, the population covariance matrix $\boldsymbol{\Sigma}\left(\boldsymbol{\theta}_0\right)$ in (6.1) is also unobserved in practice. Nevertheless, we can consistently estimate it by constructing certain estimators $\widehat{\boldsymbol{J}}_N\left(\boldsymbol{\theta}\right)$ and $\widehat{\boldsymbol{K}}_N\left(\boldsymbol{\theta}\right)$, which are pointwise consistent for $\boldsymbol{J}\left(\boldsymbol{\theta}\right)$ and $\boldsymbol{K}\left(\boldsymbol{\theta}\right)$, respectively, for a given parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Since both estimators are continuous in $\boldsymbol{\theta}$, it follows from the continuous mapping theorem that the sample analog of the sandwich-type construction in (6.1)

$$\widehat{\boldsymbol{\Sigma}}_N\left(\boldsymbol{\theta}\right) = \widehat{\boldsymbol{J}}_N\left(\boldsymbol{\theta}\right)^{-1} \widehat{\boldsymbol{K}}_N\left(\boldsymbol{\theta}\right) \widehat{\boldsymbol{J}}_N\left(\boldsymbol{\theta}\right)^{-1}$$

is pointwise consistent for $\boldsymbol{\Sigma}\left(\boldsymbol{\theta}\right)$. Combining this result with the convergence result $\widehat{\boldsymbol{\theta}}_N \xrightarrow{\mathbb{P}} \boldsymbol{\theta}_0$, the plug-in estimator $\widehat{\boldsymbol{\Sigma}}_N\left(\widehat{\boldsymbol{\theta}}_N\right)$ is consistent for the asymptotic covariance matrix $\boldsymbol{\Sigma}\left(\boldsymbol{\theta}_0\right)$. We refer to Section A of the Supplementary Material for the definition of $\widehat{\boldsymbol{J}}_N\left(\boldsymbol{\theta}\right)$ and $\widehat{\boldsymbol{K}}_N\left(\boldsymbol{\theta}\right)$ as well as details.

6.3. Efficiency

For the tuning constant $\alpha = 0$, which corresponds to the MLE $\widehat{\boldsymbol{\theta}}_N^{\mathrm{MLE}}$, it can be shown (see Section A.5 of the Supplementary Material) that if the polyserial model is correctly specified ($\varepsilon = 0$), the asymptotic covariance matrix in (6.1) reduces to the inverted *Fisher information* of the polychoric model, $\boldsymbol{I}_{\boldsymbol{\theta}_0}^{-1}$, where

$$\boldsymbol{I}_{\boldsymbol{\theta}} = \int_{\mathbb{R}} \sum_{y \in \mathcal{Y}} p_{XY}\left(x, y; \boldsymbol{\theta}\right) \boldsymbol{u}_{\boldsymbol{\theta}}\left(x, y\right) \boldsymbol{u}_{\boldsymbol{\theta}}\left(x, y\right)^{\top} \mathrm{d}x$$

denotes the Fisher information of the polyserial model at $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Hence, (6.1) nests the well-known result that the MLE's asymptotic covariance matrix is equal to the inverted Fisher information matrix. It is furthermore well-known that ML estimation is fully efficient, meaning that no unbiased estimator has a smaller variance (e.g., Theorem 3.10 in Lehmann & Casella, 1998).

https://doi.org/10.1017/psy.2026.10091 Published online by Cambridge University Press

Table 1. Relative efficiency for estimating the polyserial correlation coefficient ρ_* when Y has $r = 5$ response options, for various choices of the tuning constant α at a true parameter vector $\boldsymbol{\theta}_* = \left(\rho_*, \mu_*, \sigma_*^2, \tau_{*,1}, \tau_{*,2}, \tau_{*,3}, \tau_{*,4}\right)^\top = (0.5, 0, 1, -1.5, -0.5, 0.5, 1.5)^\top$

α	0 (MLE)	0.1	0.25	0.5	0.75	1
rel. efficiency	1.000	0.983	0.916	0.762	0.612	0.488

In contrast, for strictly positive tuning constants $\alpha > 0$, the asymptotic covariance matrix of the corresponding estimator $\widehat{\boldsymbol{\theta}}_N$ in (6.1) is generally *not* equal to the inverse Fisher information matrix at the polyserial model. Thus, recalling the full efficiency property of the MLE, our robust estimator is *less* efficient than the MLE. The efficiency loss is due to downweighting of observations with low probability under the polyserial model, which happens even when the model is correctly specified. Consequently, while the robust estimator and MLE are both consistent and unbiased for the true parameter value as long as contamination is absent, the former has a larger estimation variance.

To quantify the efficiency loss, we calculate at the population level the relative efficiency of our robust estimator compared to the MLE when the polyserial model is correctly specified ($\varepsilon = 0$). Recall that in this zero-contamination case, the estimand $\boldsymbol{\theta}_0$ is equal to the true parameter value $\boldsymbol{\theta}_*$ for all $\alpha \geq 0$ due to the property of Fisher consistency. For a given true value $\boldsymbol{\theta}_*$, the *relative efficiency* (e.g., Van der Vaart, 1998, Chapter 8.2) of our estimator for the true polyserial correlation ρ_* with respect to the fully efficient MLE is given by

$$\frac{\mathbb{V}\mathrm{ar}\left[\widehat{\rho}_N^{\mathrm{MLE}}\right]}{\mathbb{V}\mathrm{ar}\left[\widehat{\rho}_N\right]} = \left(\boldsymbol{I}_{\boldsymbol{\theta}_*}^{-1}\right)_{1,1} \Big/ \left(\boldsymbol{\Sigma}\left(\boldsymbol{\theta}_*\right)\right)_{1,1},$$

where the operator $(\cdot)_{1,1}$ picks out the top left element of a matrix. The efficiency loss of minimum DPD estimators can also be expressed as a function of α for a given model (see Basu et al., 1998, Section 4.2).

Table 1 lists the relative efficiencies of various choices of the tuning constant α for an ordinal Y with $r = 5$ response options and true parameter vector $\boldsymbol{\theta}_* = \left(\rho_*, \mu_*, \sigma_*^2, \tau_{*,1}, \tau_{*,2}, \tau_{*,3}, \tau_{*,4}\right)^\top = (0.5, 0, 1, -1.5, -0.5, 0.5, 1.5)^\top$. Up to about $\alpha = 0.25$, the relative efficiency of the robust estimator stays above 90%, and then decreases roughly linearly to slightly below 50% for $\alpha = 1$. Choices of $\alpha > 1$, while yielding even more robustness, would result in a comparatively poor relative efficiency of considerably less than 50% and are therefore not considered in this article. The efficiency results for other choices of the true parameter vector $\boldsymbol{\theta}_*$ are very similar and are therefore deferred to Section C of the Supplementary Material. Our software implementation, discussed in more detail in Section 6.4, provides functionality to compute the relative efficiency at arbitrary user-specified parameter vectors and tuning constants α.

A loss of efficiency is a common property of many robust estimators involving continuous random variables, and, as demonstrated in this section, our proposed estimator is no exception.[9] As such, our estimator is based on a compromise between efficiency and robustness. Borrowing an insurance metaphor from Anscombe (1960), we *"sacrifice some efficiency at the model, in order to insure against accidents caused by deviations from the model"* (Huber & Ronchetti, 2009, p. 5). However, it turns out that the efficiency loss is relatively minor for reasonably small positive tuning constants like $\alpha = 0.1$ (Table 1), which, as the simulation experiments in Section 7 will reveal, suffices to gain substantial robustness over ML estimation.

[9] In contrast, for exclusively categorical variables, it is possible to construct robust estimators that are fully efficient. See Welz (2024) for general models of categorical data, and Welz et al. (2026b) for the special case of the polychoric correlation model.

https://doi.org/10.1017/psy.2026.10091 Published online by Cambridge University Press

6.4. Implementation

We provide an open-source implementation of our robust estimator as part of the package `robcat` (for "ROBust CATegorical data analysis"; Welz et al., 2026a) for the statistical programming environment R (R Core Team, 2024). The package is freely available from CRAN (the Comprehensive R Archive Network) at https://CRAN.R-project.org/package=robcat. We used this package for obtaining all numerical results in this article.

In principle, every suitable method for numerical optimization can be used to minimize the robust estimator's minimization problem in (5.6). In our experience, unconstrained optimization via the BFGS algorithm (e.g., Nocedal & Wright, 2006, Section 6.1) works well. Yet, additional stability could be gained from making explicit the constraints in the parameter space $\boldsymbol{\Theta}$ in (3.6), for which standard algorithms for constrained optimization can be used, such as the simplex method of Nelder and Mead (1965). In our implementation, we adopt a similar default behavior as the implementation of robust polychoric correlation estimation (Welz et al., 2026b) in package `robcat`. Specifically, the implementation first tries unconstrained optimization with the BFGS algorithm. If instability or numerical nonconvergence are encountered or a constraint is violated, it instead uses the algorithm of Nelder and Mead (1965) for constrained optimization.[10] Nevertheless, numerous other optimization algorithms are supported, and users can freely choose their preferred option.

As for the choice of tuning constant $\alpha \geq 0$ in the DPD function (5.5), higher values lead to greater robustness but less efficiency, whereas values closer to 0 are less robust but more efficient. We shall see in the simulation experiments in the next section that the choice $\alpha = 0.5$ constitutes a good compromise between robustness against contamination up to about $\varepsilon = 0.3$ while being reasonably efficient with about 76% of the MLE's efficiency (Table 1). Therefore, $\alpha = 0.5$ is the default choice in our implementation. While this is the default choice, we acknowledge that there might be situations in which choices that lead to more (or less) robustness are more appropriate, so we advise users to decide on α on a case-by-case basis.

Moreover, we do not recommend the TS estimation procedure in (3.7) for robust estimation. Recall from (3.7) that the TS estimators for μ_* and σ_*^2 are given by the sample mean and sample variance, respectively, of the $X_i, i = 1, \ldots, N$, while the estimators for the thresholds are computed from the empirical cumulative frequencies of the $Y_i, i = 1, \ldots, N$. However, if there is contamination in the sample, such as outliers in the data for X and careless responses in the data for Y, all of these three sample statistics might become heavily biased. This bias is then possibly inherited by the estimate of the correlation coefficient in the second stage. To avoid this problem, our robust estimator estimates all model parameters simultaneously.

7. Simulation experiments

This section conducts a number of simulation experiments to evaluate the robustness of the robust minimum DPD estimator against partial misspecification of the polyserial model. Section 7.1 describes the simulation design, Section 7.2 describes the results, and Section 7.3 briefly summarizes additional simulations that are featured in detail in the Supplementary Material.

7.1. Design

We consider a polyserial model under which the random vector (X, η) is jointly normally distributed according to (3.2) with true correlation parameter $\rho_* = 0.5$ and the true first two moments of the observed X are $\mu_* = \mathbb{E}[X] = 0$ as well as $\sigma_*^2 = \mathbb{V}\mathrm{ar}[X] = 1$. The latent variable η also has zero mean

[10] In our experience, the higher the contamination fraction, the more likely it is that unconstrained optimization fails, so constrained algorithms must be used instead. As an example, we refer to the convergence statistics for the simulation in Section 7, reported in Section C of the Supplementary Material.

https://doi.org/10.1017/psy.2026.10091 Published online by Cambridge University Press

and unit variance and governs the observed ordinal Y through the discretization process in (3.1) with true threshold parameters $\tau_{*,1} = -1.5, \tau_{*,2} = -0.5, \tau_{*,3} = 0.5, \text{and} \tau_{*,4} = 1.5,$ so that Y has $r = 5$ response categories. Using integer scoring $\mathcal{Y} = \{1,2,\ldots,5\}$, the corresponding true point polyserial correlation coefficient in this setting amounts to $\widetilde{\rho}_* = 0.477$ (see Section 3.2).

To simulate contamination, we replace a fraction ε of the data for (X,η) with draws from a particular contamination distribution $H_{X\eta}$, which is not modeled by our robust estimator nor ever assumed to be known. Specifically, the contamination distribution $H_{X\eta}$ in this simulation is set to a shifted bivariate t-distribution with noncentrality parameter $(10,-2)^\top$, scale matrix $\text{diag}(0.25, 0.25)$, and 10 degrees of freedom. To obtain contaminated data for the ordinal Y, we discretize this contamination distribution's realizations in the η-dimension according to the same thresholds $\tau_{*,1},\ldots,\tau_{*,4}$ as the uncontaminated realizations from the polyserial model. The ensuing contaminated observations are mean-shifted to the right in the continuous X-dimension and primarily inflate the first and second response options in the ordinal Y-dimension. For instance, the contaminated data in Figure 1 were generated by this process (with contamination fraction $\varepsilon = 0.15$) and correspondingly exhibit these features. In the robust statistics literature, such contaminating data are an example of *negative leverage points* because they drag positive correlational estimates toward zero or even negative values, thereby creating *negative* leverage.

We sample $N = 500$ observations $(X_i, Y_i), i = 1,\ldots,N$, from this process with contamination fraction $\varepsilon \in \{0, 0.002, 0.01, 0.05, 0.1, 0.15, 0.2, 0.3, 0.4, 0.49\}$. Note that $\varepsilon = 0.002$ corresponds to only one single contaminated data point for the sample size $N = 500$. For each simulated data set, we estimate the true parameter $\boldsymbol{\theta}_*$ by means of a minimum DPD estimator $\widehat{\boldsymbol{\theta}}_N$ with tuning constants $\alpha \in \{0, 0.1, 0.25, 0.5, 0.75, 1\}$. This procedure is repeated 5,000 times.

Recall that the tuning constant choice $\alpha = 0$ corresponds to the MLE, for which we use the usual Fisher-information-based asymptotic covariance matrix to compute standard errors instead of the sandwich-type construction in (6.1). Further, recall that larger values of α lead to a more robust estimator, at the cost of a loss in efficiency. In fact, the relative efficiencies in Table 1 were computed for the same parameter value that serves as true value $\boldsymbol{\theta}_*$ in this simulation. We do not consider α-values beyond 1 due to their poor efficiency properties of having substantially less than 50% of the MLE's efficiency (see Table 1). In addition to an estimate $\widehat{\rho}_N$ of the true polyserial correlation coefficient ρ_*, we also construct an estimate $\widehat{\widetilde{\rho}}_N$ of the true point polyserial correlation coefficient $\widetilde{\rho}_*$ from the individual parameter estimates in $\widehat{\boldsymbol{\theta}}_N$ (see Section 3.2 for details).

Let $\widehat{\rho}_N$ be a polyserial correlation estimate computed from a given simulated data set, and $\widehat{\text{SE}}(\widehat{\rho}_N)$ be its associated standard error estimate constructed from the limit theory in Theorem A.1 in the Supplementary Material. We evaluate performance by means of the following metrics:

- Bias of the polyserial correlation estimate, $\widehat{\rho}_N - \rho_*$, and the point polyserial correlation estimate, $\widehat{\widetilde{\rho}}_N - \widetilde{\rho}_*$, averaged over the 5,000 repetitions.
- Average approximate bias of the standard error estimate, defined as the difference between $\widehat{\text{SE}}(\widehat{\rho}_N)$ and the sample SD of the individual correlation estimates $\widehat{\rho}_N$ across the 5,000 repetitions, where the latter is a finite-sample approximation of the true population standard error. A rigorous definition is provided in Section C.2 of the Supplementary Material.
- Coverage at the significance level $\gamma = 0.05$, defined as the proportion (across repetitions) of confidence intervals $\left[\widehat{\rho}_N \mp q_{1-\gamma/2} \cdot \widehat{\text{SE}}(\widehat{\rho}_N)\right]$ that contain the true ρ_*, where $q_{1-\gamma/2}$ denotes the $(1-\gamma/2)$ quantile of the standard normal distribution.
- Average confidence interval length at significance level $\gamma = 0.05$, where the length of an individual confidence interval is given by $2 \cdot q_{1-\gamma/2} \cdot \widehat{\text{SE}}(\widehat{\rho}_N)$.

7.2. Results

We start evaluating the results by visualizing the bias of the polyserial and point polyserial estimates. Figure 2 illustrates by means of boxplots the biases across the 5,000 repetitions. An analogous plot for the parameter vector $\boldsymbol{\theta}$ is provided in Section C of the Supplementary Material; the results are

https://doi.org/10.1017/psy.2026.10091 Published online by Cambridge University Press

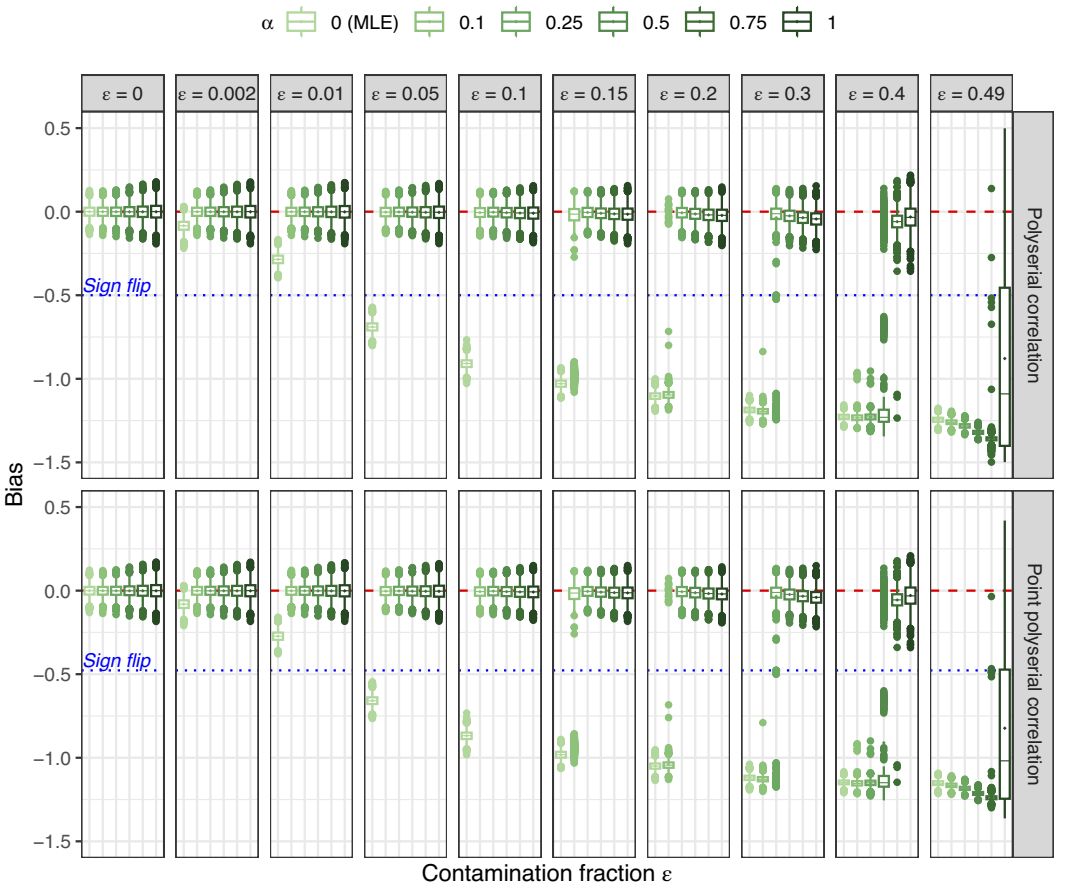


Figure 2. Boxplots of the bias of the considered estimators for the polyserial correlation coefficient, $\widehat{\rho}_N - \rho_*$ (top panel) and the point polyserial correlation coefficient with integer scoring, $\widehat{\widetilde{\rho}}_N - \widetilde{\rho}_*$ (bottom panel), for various contamination fractions in the misspecified polyserial models across 5,000 repetitions. Diamonds represent the respective average bias. The dotted lines at $\rho_* = -0.5$ and $-\widetilde{\rho}_* = -0.477$ indicate a sign flip in the respective estimate.

similar to those of the correlation estimates. Furthermore, as can be immediately seen from Figure 2, the results for polyserial and point polyserial correlation are very much alike, so we restrict the following discussion to polyserial correlation. In particular, Table 2 contains the performance measures for polyserial correlation.

In the absence of contamination ($\varepsilon = 0$), all estimators yield accurate results, but the robust estimators (those with $\alpha > 0$) exhibit more variation than the MLE ($\alpha = 0$), reflecting their diminished efficiency (e.g., Table 1). Furthermore, all estimators attain the nominal coverage level of 95%.

However, the MLE's performance rapidly deteriorates once contamination is introduced ($\varepsilon > 0$). For instance, one single bad data point ($\varepsilon = 0.002$ here) suffices for the MLE to exhibit a notable bias of about -0.084, resulting in a deteriorated coverage of less than 40%. At contamination $\varepsilon = 0.01$, its coverage even drops to 0%. At contamination level $\varepsilon = 0.05$, its point estimate experiences a sign flip: instead of yielding a positive estimate of the true value $\rho_* = 0.5$, its average estimate of -0.19 is negative. The ML estimate continues to deteriorate until it roughly stabilizes at about $\varepsilon = 0.2$ with an average bias of about -0.7.

Conversely, the minimum DPD estimators ($\alpha > 0$) turn out to be substantially more robust to contamination than ML. Up until and including the contamination fraction of $\varepsilon = 0.1$, all minimum DPD estimators are virtually unaffected and maintain the nominal coverage level. At $\varepsilon = 0.15$, the

https://doi.org/10.1017/psy.2026.10091 Published online by Cambridge University Press

Table 2. Performance measures for estimating polyserial correlation coefficients in the simulation in Section 7.1 at significance level $\gamma = 0.05$ (averaged across 5,000 repetitions)

		Point estimate			Standard error		Confidence interval	
Contamination	α	$\widehat{\rho}_N$	Bias	SD	$\widehat{\mathrm{SE}}(\widehat{\rho}_N)$	Bias	Coverage	Length
$\varepsilon = 0$	0 (MLE)	0.500	0.000	0.035	0.036	0.001	0.951	0.142
	0.1	0.500	0.000	0.036	0.036	0.000	0.948	0.141
	0.25	0.499	−0.001	0.037	0.037	0.000	0.949	0.147
	0.5	0.499	−0.001	0.041	0.041	0.000	0.950	0.162
	0.75	0.499	−0.001	0.046	0.046	0.001	0.951	0.182
	1	0.500	0.000	0.051	0.052	0.001	0.953	0.205
$\varepsilon = 0.002$	0 (MLE)	0.416	−0.084	0.035	0.039	0.003	0.395	0.152
	0.1	0.499	−0.001	0.036	0.036	0.000	0.948	0.141
	0.25	0.499	−0.001	0.037	0.037	0.000	0.950	0.147
	0.5	0.499	−0.001	0.041	0.041	0.000	0.949	0.162
	0.75	0.499	−0.001	0.046	0.046	0.001	0.951	0.182
	1	0.499	−0.001	0.051	0.052	0.001	0.953	0.205
$\varepsilon = 0.01$	0 (MLE)	0.215	−0.285	0.033	0.057	0.024	0.000	0.224
	0.1	0.499	−0.001	0.036	0.036	0.000	0.947	0.142
	0.25	0.499	−0.001	0.037	0.038	0.000	0.950	0.147
	0.5	0.499	−0.001	0.041	0.041	0.001	0.951	0.162
	0.75	0.499	−0.001	0.046	0.046	0.001	0.953	0.182
	1	0.499	−0.001	0.051	0.052	0.001	0.953	0.205
$\varepsilon = 0.05$	0 (MLE)	−0.190	−0.690	0.032	0.073	0.042	0.000	0.287
	0.1	0.497	−0.003	0.037	0.037	0.000	0.949	0.146
	0.25	0.498	−0.002	0.038	0.038	0.000	0.948	0.150
	0.5	0.497	−0.003	0.041	0.042	0.000	0.947	0.164
	0.75	0.496	−0.004	0.046	0.047	0.001	0.951	0.183
	1	0.496	−0.004	0.051	0.052	0.001	0.953	0.204
$\varepsilon = 0.1$	0 (MLE)	−0.409	−0.909	0.030	0.063	0.033	0.000	0.248
	0.1	0.494	−0.006	0.039	0.039	0.000	0.948	0.151
	0.25	0.497	−0.003	0.039	0.039	0.000	0.950	0.154
	0.5	0.494	−0.006	0.042	0.043	0.000	0.948	0.167
	0.75	0.492	−0.008	0.046	0.047	0.001	0.950	0.183
	1	0.491	−0.009	0.051	0.052	0.001	0.952	0.202
$\varepsilon = 0.15$	0 (MLE)	−0.529	−1.029	0.027	0.059	0.032	0.000	0.233
	0.1	0.369	−0.131	0.327	0.042	−0.285	0.830	0.164
	0.25	0.495	−0.005	0.040	0.040	0.000	0.948	0.157
	0.5	0.490	−0.010	0.043	0.043	0.000	0.948	0.170
	0.75	0.487	−0.013	0.046	0.047	0.001	0.948	0.184
	1	0.485	−0.015	0.050	0.051	0.001	0.949	0.201

(Continued)

https://doi.org/10.1017/psy.2026.10091 Published online by Cambridge University Press

Table 2. Continued

		Point estimate			Standard error		Confidence interval	
Contamination	α	$\widehat{\rho}_N$	Bias	SD	$\widehat{\mathrm{SE}}(\widehat{\rho}_N)$	Bias	Coverage	Length
$\varepsilon = 0.2$	0 (MLE)	−0.603	−1.103	0.024	0.060	0.035	0.000	0.234
	0.1	−0.591	−1.091	0.078	0.037	−0.041	0.004	0.145
	0.25	0.493	−0.007	0.041	0.041	0.000	0.947	0.162
	0.5	0.486	−0.014	0.044	0.044	0.000	0.945	0.172
	0.75	0.481	−0.019	0.047	0.047	0.000	0.941	0.185
	1	0.478	−0.022	0.050	0.051	0.001	0.939	0.199
$\varepsilon = 0.3$	0 (MLE)	−0.686	−1.186	0.020	0.063	0.043	0.000	0.247
	0.1	−0.694	−1.194	0.021	0.024	0.003	0.000	0.094
	0.25	0.464	−0.036	0.171	0.043	−0.128	0.926	0.170
	0.5	0.475	−0.025	0.045	0.046	0.000	0.923	0.179
	0.75	0.464	−0.036	0.047	0.048	0.001	0.901	0.187
	1	0.456	−0.044	0.049	0.050	0.001	0.882	0.196
$\varepsilon = 0.4$	0 (MLE)	−0.727	−1.227	0.018	0.065	0.047	0.000	0.257
	0.1	−0.732	−1.232	0.022	0.019	−0.003	0.000	0.073
	0.25	−0.729	−1.229	0.024	0.019	−0.005	0.000	0.074
	0.5	−0.475	−0.975	0.500	0.027	−0.473	0.191	0.106
	0.75	0.440	−0.060	0.058	0.056	−0.002	0.772	0.221
	1	0.468	−0.032	0.074	0.127	0.053	0.818	0.498
$\varepsilon = 0.49$	0 (MLE)	−0.745	−1.245	0.017	0.066	0.049	0.000	0.260
	0.1	−0.760	−1.260	0.016	0.016	0.000	0.000	0.064
	0.25	−0.782	−1.282	0.015	0.015	0.000	0.000	0.059
	0.5	−0.820	−1.320	0.014	0.014	0.000	0.000	0.053
	0.75	−0.858	−1.358	0.038	0.031	−0.007	0.001	0.118
	1	−0.379	−0.879	0.676	0.122	−0.554	0.007	0.478

coverage of the estimator with $\alpha = 0.1$ drops slightly to about 83%. At $\varepsilon = 0.2$, its coverage drops to nearly 0%, while all choices with $\alpha > 0.1$ remain stable with a coverage above 90%. Furthermore, all minimum DPD estimators up to this point accurately estimate the standard error of polyserial correlation (except $\alpha = 0.1$ at $\varepsilon > 0.15$, where its coverage is near 0). In the high-contamination setting of $\varepsilon = 0.3$, estimators with $\alpha > 0.1$ start exhibiting a minor but notable bias, while at $\varepsilon = 0.4$, only the choices $\alpha \in \{0.75, 1\}$ retain a reasonably good performance, while all other estimators have coverage rates of or near zero. At $\varepsilon = 0.49$, the highest considered contamination level, all estimators break down.[11]

This simulation demonstrates that ML estimation of the polyserial correlation coefficient is highly susceptible to already very low levels of contamination. Conversely, using our proposed estimator yields substantial gains in robustness. For instance, at tuning constant $\alpha = 0.1$, which only sacrifices less than

[11] All point estimators always numerically converged across all configurations. However, in extremely high-contamination settings, the covariance matrix estimator could occasionally not be computed due to singularity issues. Such non-existent standard error estimates were omitted in Figure 2 and Table 2. Nonconvergence statistics are provided in Section C of the Supplementary Material.

https://doi.org/10.1017/psy.2026.10091 Published online by Cambridge University Press

2% efficiency (Table 1), the robust estimator remains accurate up until a contamination level of about $\varepsilon = 0.1$. Robustness against higher contamination levels can be achieved by choosing higher values of α. In addition, the simulation also exposes the boundaries of the robust estimator: In extremely high-contamination settings of beyond 40% contamination, even the estimator with the highest considered tuning constant value of $\alpha = 1$ breaks down. However, it is questionable whether modeling with data of such extremely poor quality is meaningful to begin with.

7.3. Additional simulations

We perform three additional simulation studies, all described in detail in Section D of the Supplementary Material. These simulations are intended to explore the benefits and limitations of our proposed methodology.

The first additional simulation, described in Section D.1 of the Supplementary Material, has the same setup as the design in Section 7.1, but the contamination manifests through extreme outliers in both the X- and η-dimension. This simulation is primarily intended to study the behavior of the MLE and robust estimator under gross errors in the data. In brief, one single gross error observation in a sample suffices for the MLE to produce sign-flipped estimates that are extremely biased. In contrast, the robust estimator remains nearly unaffected by small to moderate contamination fractions with such gross errors, and only starts exhibiting considerable biases in high-contamination settings.

The second additional simulation, described in Section D.2 of the Supplementary Material, is motivated by the fact that the simulation design in Section 7.1 is primarily concerned with mean-shifted contamination. Mean-shifted contamination is characterized by the population mean of the contamination distribution $H_{X\eta}$ being markedly different than that of the true normal distribution $P_{X\eta}$. However, contamination may also manifest through changes in *correlation* rather than means. The second additional simulation therefore focuses on a *correlation-shifted* contamination distribution $H_{X\eta}$, which is equal to the true normal distribution $P_{X\eta}$, except for a sign-flipped correlation coefficient $-\rho_*$. In brief, for moderate polyserial correlation ρ_* (where $H_{X\eta}$ and $P_{X\eta}$ substantially overlap), our robust estimator does not yield an improvement over ML. Conversely, it does provide a notable gain in robustness for larger true correlations ρ_* (where $H_{X\eta}$ and $P_{X\eta}$ barely overlap).

Combined with the simulations on mean-shifted contamination in Section 7.1 and Section D.1 of the Supplementary Material, the results of the simulation on correlation-shifted contamination suggest that the potential for robustness gain depends on the overlap between the contamination distribution $H_{X\eta}$ and true normal distribution $P_{X\eta}$. If there is much overlap, the robust estimator cannot distinguish well between contamination and regular observations since it does not make any assumptions on the former, so it may not improve upon ML. In contrast, if $H_{X\eta}$ and $P_{X\eta}$ do not overlap much—like with mean-shifted contamination or correlation-shifted contamination with strong true correlation—the robust estimator can identify contaminated observations and subsequently downweigh their influence to achieve considerable robustness gains.

The third additional simulation, described in Section D.3 of the Supplementary Material, is concerned with distributional misspecification, where the polyserial model is misspecified for *all* observations in a sample (cf. Section 4.2). The simulation shows that the potential of robustness gain with our robust estimator depends on the case-specific characteristics of the nonnormal sampling distribution of (X,η) for which the polyserial model is distributionally misspecified. In brief, if the sampling distribution can be reasonably well approximated by a mixture of a normal distribution and some other unknown distribution to emulate the contamination model in (4.1), then the robust estimator can improve upon the MLE. If the sampling distribution does not admit such an approximation, the robust estimator may not yield an improvement but produce similar estimates as ML, which can be quite poor under distributional misspecification (e.g., Bedrick, 1995, and our simulations in Section D.3 of the Supplementary Material). Thus, we advice applied researchers to always test for partially-latent normality when ML and robust estimator yield similar estimates. While such similarity may be due to normality indeed holding true, it may also be due to distributional misspecification from a specific

https://doi.org/10.1017/psy.2026.10091 Published online by Cambridge University Press

nonnormal distribution for which the robust estimator cannot improve over ML. We provide details and more explicit guidance in Section D.3 of the Supplementary Material.

8. Empirical application

To demonstrate our proposed estimator in practice, we apply it to an empirical data set from personality psychology. The data are from an administration of the *Eysenck Personality Inventory* (Eysenck & Eysenck, 1964) to $N = 231$ undergraduate students at Northwestern University (collected by William Revelle) and are publicly available in the R package `psychTools` (Revelle, 2025). We restrict ourselves to the continuous variable `stateanx` and the ordinal variable `epilie` in this demonstration.

The continuous variable `stateanx` is the score of an unspecified scale measuring *state anxiety*, defined as a temporary emotional reaction to adverse events (e.g., Saviola et al., 2020).[12] Figure 3a provides a histogram of this variable in the data of Revelle (2025). The empirical distribution seems to be skewed to the right. In particular, there is one unusually large observation in its right tail with value 79 that seems to be separated from the rest of the data, so it may be viewed as an outlier (the sample mean is 39.8). Hence, due to the inflated right tail, the validity of the normality assumption seems questionable. Indeed, the nonparametric Shapiro–Wilk test rejects marginal normality of this variable with a test statistic of $W = 0.953$ and a p-value of < 0.001.

The ordinal variable `epilie` is the score of a lie scale and measures social desirability of a participant's responses: The higher the score, the less trustworthy their responses. Originally, this variable has 10 unique ordinal response options.[13] However, in the data of Revelle (2025), no participant has the highest score of 9 or lowest score of 0, and only two participants have the second-highest score of 8. To avoid numerical instability in the estimation of thresholds,[14] we therefore assign the two participants with score 8 the score 7, that is, we merge the seventh and eighth response categories, and treat the score 1 as the first response category. Thus, the variable `epilie` here takes values in $\mathcal{Y} = \{1, 2, \ldots, 7\}$ with a total of $r = 7$ response options. Figure 3b visualizes the frequency of each response option.

Suppose we are interested in the polyserial correlation coefficient between state anxiety (`stateanx`) and the ordinal lie score (`epilie`) in the data of Revelle (2025). We therefore fit the polyserial model to these data using ML and our proposed robust estimator. For the latter, we focus on the results with tuning constant $\alpha = 0.5$, being the default choice in our software, but also report those with other choices of α.

Table 3 summarizes the results of the two estimators. The correlation estimate is weaker for the robust estimator with a value of -0.107 than for the MLE with a value of -0.142. The largest differences occur in the moment estimates of the continuous variable `stateanx`: Both μ and σ^2 are estimated to be notably larger by ML (estimates of about 39.8 and 131.6, respectively) compared to the robust estimator (estimates of about 38 and 107.7, respectively). The threshold estimates are similar. Thus, overall, the robust estimates tend to be weaker in absolute magnitude than those of ML.

To obtain insights as to why the robust estimator's results differ from those of the MLE, we calculate the individual-specific weights in (5.9) using the robust parameter estimates. Figure 4 visualizes the sorted weights. While many observations can be fitted well with weights of nearly 1, there are also observations that receive a weight of close to 0, indicating that strong downweighting has taken place.

[12] Technically, being an aggregate score of ordinal measurements, the variable `stateanx` is strictly speaking not continuous. However, due to the granularity of the measurements (see Figure 3a), it may be viewed as being quasi-continuous.

[13] The Eysenck Personality Inventory constructs the variable `epilie` as the number of times that a given participant has responded "yes" in nine separate yes–no questions. Since higher scores are associated with higher levels of socially desirable responding, this variable can be interpreted as being ordinal.

[14] It may happen that a robust estimator attempts to effectively eliminate thresholds corresponding to (nearly) empty cells in a certain way that we describe in detail in Section D.3.2 of the Supplementary Material, thereby causing instability. This phenomenon has been noted before by Welz et al. (2026b).

https://doi.org/10.1017/psy.2026.10091 Published online by Cambridge University Press

Table 3. Parameter estimates and standard error estimates ($\widehat{\text{SE}}$) for the correlation between the continuous variable `stateanx` and the ordinal variable `epilie` (with $r = 7$ response options) in the data of Revelle (2025), using the robust estimator with $\alpha = 0.5$ and the MLE

	Robust		ML	
Parameter	Estimate	$\widehat{\text{SE}}$	Estimate	$\widehat{\text{SE}}$
ρ	−0.107	0.076	−0.142	0.067
μ	37.990	0.615	39.842	0.900
σ^2	107.730	10.280	131.575	14.485
τ_1	−1.359	0.120	−1.358	0.118
τ_2	−0.586	0.091	−0.604	0.088
τ_3	0.276	0.085	0.276	0.084
τ_4	0.828	0.096	0.841	0.094
τ_5	1.268	0.116	1.286	0.116
τ_6	1.849	0.152	1.771	0.152

Table 4. Weights and values of variables `stateanx` (continuous) and `epilie` (ordinal with $r = 7$) for observations in the data of Revelle (2025) whose robustly estimated weights are below 0.1 (using tuning constant $\alpha = 0.5$)

i	1	2	3	4	5	6
$w_{i,\alpha}(\widehat{\boldsymbol{\theta}}_N)$	0.020	0.049	0.068	0.082	0.093	0.096
`stateanx` (X_i)	79	69	61	68	70	67
`epilie` (Y_i)	3	5	7	1	3	4

Note: The observations are sorted according to the estimated weights.

We therefore investigate in detail all observations whose weight is reasonably close to zero, say, below 0.1. Table 4 lists the weights and values of the two variables for observations whose estimated weights are below 0.1. The smallest weight (valued 0.02) belongs to the observation with the outlying value of 79 in variable `stateanx` (Figure 3a). Thus, the robust estimator has identified this observation as outlying and subsequently downweighs it in the estimation procedure. In similar fashion, the other observations in Table 4 are those with relatively large `stateanx` values. It therefore seems that the robust estimator is downweighting observations whose `stateanx` values are in the heavy right tail (see Figure 3a). A notable exception is the observation that has a somewhat less high `stateanx` value of only 61. Nevertheless, this observation's lie score amounts to the maximum value of 7, indicating low trustworthiness of the given responses. The fact that this observation was strongly downweighted suggests that it could not sufficiently well modeled by the polyserial model, which might be due to its low trustworthiness.

As a sensitivity analysis, Figure 5 plots the polyserial correlation estimates computed for various choices of tuning constant α. All strictly positive choices of α yield estimates that are weaker in magnitude than the MLE, but stay around −0.1 up until approximately $\alpha = 0.6$. Beyond that tuning constant value, the estimates sharply converge to about 0, primarily reflecting more stringent downweighting of the heavy right tail of the continuous variable `stateanx`. Overall, though, for tuning constants $\alpha \leq 0.6$, we obtain the robust finding of a correlation of about −0.1, being weaker than the ML estimate of −0.14.

This empirical application demonstrated the practical usefulness of our proposed robust estimator by showing how it identifies observations that are likely discrepant to the polyserial model and

https://doi.org/10.1017/psy.2026.10091 Published online by Cambridge University Press

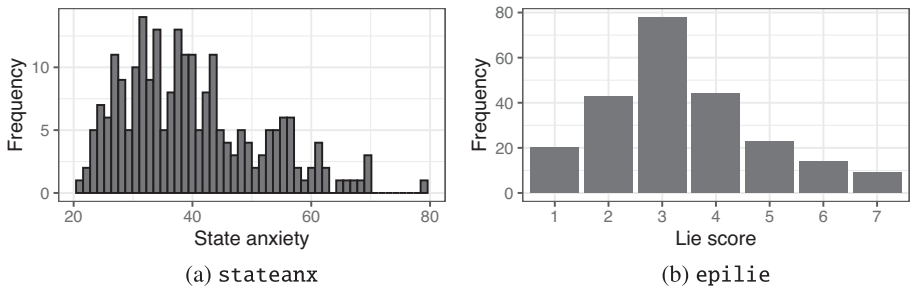


Figure 3. Histogram of the continuous variable `stateanx` (left) and response frequencies of the ordinal variable `epilie` (right) in the data of Revelle (2025), for $N = 231$ observations.

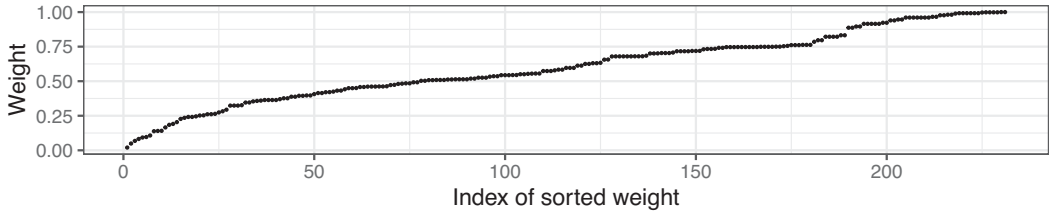


Figure 4. Weights $w_{i,\alpha}\left(\widehat{\theta}_N\right), i = 1, \ldots, N$, computed with the robust parameter estimates at tuning constants $\alpha = 0.5$, using the data in Revelle (2025) for the variables `stateanx` and `epilie`. For a clearer visualization, the $N = 231$ weights are sorted here.

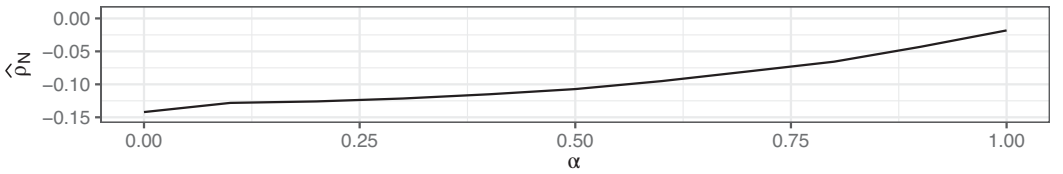


Figure 5. Estimates of the polyserial correlation coefficient for different values of tuning constant α ($\alpha = 0$ is MLE), computed on the variables `stateanx` and `epilie` in the data of Revelle (2025).

subsequently downweights their influence. It furthermore demonstrates how the estimated weights are an attractive tool for interpretation and exploratory data analysis.

9. Discussion and conclusion

Motivated by recent interest in the psychometric literature in the non-robustness of ML estimation of latent variable models under model misspecification, we study estimation of the polyserial correlation model, which models the association between a continuous variable and an ordinal variable. We consider a *partial* misspecification framework stemming from the robust statistics literature where the polyserial model is misspecified for an unknown, possibly zero-valued fraction of an observed sample. Crucially, this framework does not impose any assumptions on *how* the model is possibly misspecified, which might be due to, for instance but not limited to, outliers or careless responses. We show that one single observation not generated by the polyserial model suffices for the commonly used MLE to produce arbitrarily poor results.

As a remedy, we propose a novel estimator designed to be robust against partial misspecification of the polyserial model. The robust estimator leverages the theory of minimum DPD (Basu et al., 1998) estimation, which, as we show, is able to cope with mixed-type variables, unlike standard approaches for robust estimation with exclusively continuous or exclusively categorical variables. The proposed

https://doi.org/10.1017/psy.2026.10091 Published online by Cambridge University Press

minimum DPD estimator achieves robustness by implicitly downweighting observations that cannot be sufficiently well-fitted by the polyserial model. The ensuing weights are a useful analytical tool for identifying observations discrepant to the polyserial model, such as (but not limited to) outliers and careless responses. Our proposed methodology is implemented as part of the free open-source R package `robcat` (Welz et al., 2026a), which is publicly available at https://CRAN.R-project.org/package=robcat.

We show that the proposed robust estimator is consistent as well as asymptotically normally distributed, thereby enabling inference. The price to pay for enhanced robustness is diminished statistical efficiency of our estimator compared to ML. However, we show that substantial robustness can be gained with our estimator while maintaining more than 98% of ML efficiency.

We verify the estimator's robustness and theoretical properties by means of numerous simulation studies. An empirical application on a data set from personality psychology furthermore demonstrates its practical usefulness, where it identifies outlying data points.

A central practical consideration is the choice of tuning constant α, governing the robustness-efficiency tradeoff. While simulations suggest that $\alpha = 0.5$ is a good compromise with about 76% of ML efficiency, we acknowledge that a detailed investigation is an important avenue of further research. We recommend to always compare various choices of α, like in Figure 5, to assess estimation stability and severity of (partial) model misspecification. Alternatively, one may fix the efficiency loss and select the corresponding value of α as part of an ex-ante power analysis (cf. Basu et al., 1998, Section 5).

This article suggests a number of extensions. For instance, while our software implementation is reasonably fast with about 2 seconds for computing a robust estimate, substantial speedups could be achieved by rewriting the source code in C++ via `Rcpp` (Eddelbuettel, 2013). Moreover, the ability to robustly estimate polyserial correlation could be particularly useful in structural equation modeling with mixed data to robustify such analyses against partial model misspecification. We leave these avenues to further research. Overall, we believe that our robust estimator can contribute to the growing literature on making latent variable analyses less dependent on easily violated modeling assumptions.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/psy.2026.10091.

Data availability statement. Replication files are publicly available on GitHub at https://github.com/mwelz/robust-polyserial-replication.

Acknowledgements. I thank Andreas Alfons for valuable feedback. I further thank the editor as well as three anonymous reviewers, whose comments led to a substantially improved article.

Author contributions. This article is single-authored. The author is solely responsible for all aspects of the research, including conception, design, data collection, analysis, and manuscript writing.

Funding statement. This research received no specific grant funding from any funding agency, commercial, or not-for-profit sectors.

Competing interests. The author declares none.

Ethical standards. The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

References

Alfons, A., Ateş, N., & Groenen, P. (2022). A robust bootstrap test for mediation analysis. *Organizational Research Methods*, *25*(3), 591–617. https://doi.org/10.1177/1094428121999096

Alfons, A., & Welz, M. (2024). Open science perspectives on machine learning for the identification of careless responding: A new hope or phantom menace? *Social and Personality Psychology Compass*, *18*(2), e12941. https://doi.org/10.1111/spc3.12941

Anscombe, F. (1960). Rejection of outliers. *Technometrics*, *2*, 123–147. https://doi.org/10.1080/00401706.1960.10489888

Arias, V. B., Garrido, L., Jenaro, C., Martinez-Molina, A., & Arias, B. (2020). A little garbage in, lots of garbage out: Assessing the impact of careless responding in personality survey data. *Behavior Research Methods*, *52*(6), 2489–2505. https://doi.org/10.3758/s13428-020-01401-8

Asparouhov, T., & Muthén, B. (2016). Structural equation models and mixture models with continuous nonnormal skewed distributions. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(1), 1–19. https://doi.org/10.1080/10705511.2014.947375

Barbiero, A. (2025). Maximal point-polyserial correlation for non-normal random distributions. *British Journal of Mathematical and Statistical Psychology*, *78*(1), 341–377. https://doi.org/10.1111/bmsp.12362

Basu, A., Harris, I. R., Hjort, N. L., & Jones, M. C. (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika*, *85*(3), 549–559. https://doi.org/10.1093/biomet/85.3.549

Bedrick, E. J. (1990). On the large sample distributions of modified sample biserial correlation coefficients. *Psychometrika*, *55*(2), 217–228. https://doi.org/10.1007/BF02295284

Bedrick, E. J. (1992). A comparison of generalized and modified sample biserial correlation estimators. *Psychometrika*, *57*(2), 183–201. https://doi.org/10.1007/BF02294504

Bedrick, E. J. (1995). A note on the attenuation of correlation. *British Journal of Mathematical and Statistical Psychology*, *48*(2), 271–280. https://doi.org/10.1111/j.2044-8317.1995.tb01064.x

Bedrick, E. J., & Breslin, F. C. (1996). Estimating the polyserial correlation coefficient. *Psychometrika*, *61*(3), 427–443. https://doi.org/10.1007/BF02294548

Bowling, N. A., Huang, J. L., Bragg, C. B., Khazon, S., Liu, M., & Blackmore, C. E. (2016). Who cares and who is careless? Insufficient effort responding as a reflection of respondent personality. *Journal of Personality and Social Psychology*, *111*(2), 218–229. https://doi.org/10.1037/pspp0000085

Brogden, H. E. (1949). A new coefficient: Application to biserial correlation and to estimation of selective efficiency. *Psychometrika*, *14*(3), 169–182. https://doi.org/10.1007/BF02289151

Cheng, Y., & Liu, H. (2016). A short note on the maximal point-biserial correlation under non-normality. *British Journal of Mathematical and Statistical Psychology*, *69*(3), 344–351. https://doi.org/10.1111/bmsp.12075

Coenders, G., Satorra, A., & Saris, W. E. (1997). Alternative approaches to structural modeling of ordinal data: A Monte Carlo study. *Structural Equation Modeling: A Multidisciplinary Journal*, *4*(4), 261–282. https://doi.org/10.1080/10705519709540077

Cox, N. (1974). Estimation of the correlation between a continuous and a discrete variable. *Biometrics*, 171–178. https://doi.org/10.2307/2529626

Demirtas, H., & Hedeker, D. (2016). Computing the point-biserial correlation under any underlying continuous distribution. *Communications in Statistics—Simulation and Computation*, *45*(8), 2744–2751. https://doi.org/10.1080/03610918.2014.920883

Demirtas, H., & Vardar-Acar, C. (2017). Anatomy of correlational magnitude transformations in latency and discretization contexts in Monte-Carlo studies. In D.-G. Chen & J. Chen (Eds.), *Monte-Carlo simulation-based statistical modeling* (pp. 59–84). Springer.

Eddelbuettel, D. (2013). *Seamless R and C++ integration with Rcpp*. Springer. https://doi.org/10.1007/978-1-4614-6868-4

Eysenck, H. J., & Eysenck, S. B. G. (1964). *Manual of the Eysenck personality inventory*. University of London Press.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*(4), 466–491. https://doi.org/10.1037/1082-989X.9.4.466

Foldnes, N., & Grønneberg, S. (2019). On identification and non-normal simulation in ordinal covariance and item response models. *Psychometrika*, *84*(4), 1000–1017. https://doi.org/10.1007/s11336-019-09688-z

Foldnes, N., & Grønneberg, S. (2020). Pernicious polychorics: The impact and detection of underlying non-normality. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(4), 525–543. https://doi.org/10.1080/10705511.2019.1673168

Foldnes, N., & Grønneberg, S. (2022). The sensitivity of structural equation modeling with ordinal data to underlying non-normality and observed distributional forms. *Psychological Methods*, *27*(4), 541–567. https://doi.org/10.1037/met0000385

Grønneberg, S., & Foldnes, N. (2017). Covariance model simulation using regular vines. *Psychometrika*, *82*, 1035–1051. https://doi.org/10.1007/s11336-017-9569-6

Grønneberg, S., & Foldnes, N. (2019). A problem with discretizing Vale–Maurelli in simulation studies. *Psychometrika*, *84*(2), 554–561. https://doi.org/10.1007/s11336-019-09663-8

Grønneberg, S., & Foldnes, N. (2022). Factor analyzing ordinal items requires substantive knowledge of response marginals. *Psychological Methods*. [In press.]. https://doi.org/10.1037/met0000495

Grønneberg, S., Moss, J., & Foldnes, N. (2020). Partial identification of latent correlations with binary data. *Psychometrika*, *85*(4), 1028–1051. https://doi.org/10.1007/s11336-020-09737-y

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1986). *Robust statistics: The approach based on influence functions*. Wiley.

Huang, J. L., Liu, M., & Bowling, N. A. (2015). Insufficient effort responding: Examining an insidious confound in survey data. *Journal of Applied Psychology*, *100*(3), 828–845. https://doi.org/10.1037/a0038510

https://doi.org/10.1017/psy.2026.10091 Published online by Cambridge University Press

Huber, P. J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics*, *35*(1), 73–101. https://doi.org/10.1214/aoms/1177703732

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In L. M. Le Cam & J. Neyman (Eds.), *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. University of California Press.

Huber, P. J., & Ronchetti, E. M. (2009). *Robust statistics*. (2nd ed.) Wiley. https://doi.org/10.1002/9780470434697

Itaya, Y., & Hayashi, K. (2025). *Robust estimation of item parameters via divergence measures in item response theory*. Preprint, arXiv:2502.10741. https://doi.org/10.48550/arXiv.2502.10741

Jaspen, N. (1946). Serial correlation. *Psychometrika*, *11*(1), 23–30. https://doi.org/10.1007/BF02288896

Jin, S., & Yang-Wallentin, F. (2017). Asymptotic robustness study of the polychoric correlation estimation. *Psychometrika*, *82*(1), 67–85. https://doi.org/10.1007/s11336-016-9512-2

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8: User's reference guide*. Scientific Software International.

Koopman, R. F. (1983). On the standard error of the modified biserial correlation. *Psychometrika*, *48*(4), 639–641. https://doi.org/10.1007/BF02293887

Kraemer, H. C. (1981). Modified biserial correlation coefficients. *Psychometrika*, *46*(3), 275–282. https://doi.org/10.1007/BF02293735

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, *22*(1), 79–86. https://doi.org/10.1214/aoms/1177729694

Lehmann, E. L., & Casella, G. (1998). *Theory of point estimation*. (2nd ed.) Springer. https://doi.org/10.1007/b98854

Li, C.-H. (2016). The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychological Methods*, *21*(3), 369–387. https://doi.org/10.1037/met0000093

Lord, F. M. (1963). Biserial estimates of correlation. *Psychometrika*, *28*(1), 81–85. https://doi.org/10.1007/BF02289550

Lyhagen, J., & Ornstein, P. (2023). Robust polychoric correlation. *Communications in Statistics—Theory and Methods*, *52*(10), 3241–3261. https://doi.org/10.1080/03610926.2021.1970770

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*(3), 437–455. https://doi.org/10.1037/a0028085

Monroe, S. (2018). Contributions to estimation of polychoric correlations. *Multivariate Behavioral Research*, *53*(2), 247–266. https://doi.org/10.1080/00273171.2017.1419851

Moss, J., & Grønneberg, S. (2023). Partial identification of latent correlations with ordinal data. *Psychometrika*, *88*(1), 241–252. https://doi.org/10.1007/s11336-022-09898-y

Muthén, L. K., & Muthén, B. O. (2026). Mplus user's guide, version 9. [Software and documentation for Mplus Version 9]. Software and documentation for Mplus Version 9. Muthén & Muthén. Los Angeles, CA, USA. https://www.statmodel.com/

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, *7*(4), 308–313. https://doi.org/10.1093/comjnl/7.4.308

Nocedal, J., & Wright, S. J. (2006). *Numerical optimization*. (2nd ed.) Springer. https://doi.org/10.1007/978-0-387-40065-5

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, *44*(4), 443–460. https://doi.org/10.1007/BF02296207

Olsson, U., Drasgow, F., & Dorans, N. J. (1982). The polyserial correlation coefficient. *Psychometrika*, *47*(3), 337–347. https://doi.org/10.1007/BF02294164

Pearson, K. (1909). On a new method of determining correlation between a measured character *A*, and a character *B*, of which only the percentage of cases wherein *B* exceeds (or falls short of) a given intensity is recorded for each grade of *A*. *Biometrika*, *7*(1–2), 96–105. https://doi.org/10.1093/biomet/7.1-2.96

Pearson, K. (1913). On the measurement of the influence of "broad categories" on correlation. *Biometrika*, *9*(1–2), 116–139. https://doi.org/10.1093/biomet/9.1-2.116

Pearson, K., & Pearson, E. S. (1922). On polychoric coefficients of correlation. *Biometrika*, *14*(1–2), 127–156. https://doi.org/10.1093/biomet/14.1-2.127

R Core Team. (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Danish Institute of Educational Research.

Revelle, W. (2025). psychTools: Tools to accompany the psych package for psychological research. [R package version 2.5.3]. R package version 2.5.3. Northwestern University. Evanston, Illinois. https://CRAN.R-project.org/package=psychTools

Roscino, A., & Pollice, A. (2006). A generalization of the polychoric correlation coefficient. In S. Zani, A. Cerioli, M. Riani, & M. Vichi (Eds.), *Data analysis, classification and the forward search*. https://doi.org/10.1007/3-540-35978-8˙16

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Saviola, F., Pappaianni, E., Monti, A., Grecucci, A., Jovicich, J., & De Pisapia, N. (2020). Trait and state anxiety are mapped differently in the human brain. *Scientific Reports*, *10*(1), 11112. https://doi.org/10.1038/s41598-020-68008-z

https://doi.org/10.1017/psy.2026.10091 Published online by Cambridge University Press

Tate, R. F. (1955a). Applications of correlation models for biserial data. *Journal of the American Statistical Association*, *50*(272), 1078–1095. https://doi.org/10.1080/01621459.1955.10501293

Tate, R. F. (1955b). The theory of correlation between two continuous variables when one is dichotomized. *Biometrika*, *42*(1–2), 205–216. https://doi.org/10.1093/biomet/42.1-2.205

Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, *48*, 465–471. https://doi.org/10.1007/BF02293687

Van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge University Press.

Ward, M., & Meade, A. W. (2023). Dealing with careless responding in survey data: Prevention, identification, and recommended best practices. *Annual Review of Psychology*, *74*(1), 577–596. https://doi.org/10.1146/annurev-psych-040422-045007

Welz, M. (2024). *Robust estimation and inference for categorical data*. Preprint, arXiv:2403.11954. https://doi.org/10.48550/arXiv.2403.11954

Welz, M., Alfons, A., & Mair, P. (2026a). robcat: Robust categorical data analysis. R package version 0.2.0. https://CRAN.R-project.org/package=robcat

Welz, M., Mair, P., & Alfons, A. (2026b). Robust estimation of polychoric correlation. *Psychometrika*. [forthcoming]. https://doi.org/10.1017/psy.2025.10066

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, *50*(1), 1–26. https://doi.org/10.2307/1912526

Yuan, K.-H., Bentler, P. M., & Chan, W. (2004). Structural equation modeling with heavy tailed distributions. *Psychometrika*, *69*(3), 421–436. https://doi.org/10.1007/BF02295644

Cite this article: Welz, M. (2026). Robust Estimation of Polyserial Correlation Coefficients: A Density Power Divergence Approach. *Psychometrika*, 445–472. https://doi.org/10.1017/psy.2026.10091

https://doi.org/10.1017/psy.2026.10091 Published online by Cambridge University Press