Deconvolution from two order statistics

JoonHwan Cho*[1], Yao Luo†[2], and Ruli Xiao‡[3]

[1]Binghamton University
[2]University of Toronto
[3]Indiana University

June 4, 2024

Abstract

Economic data are often contaminated by measurement errors and truncated by ranking. This paper shows that the classical measurement error model with independent and additive measurement errors is identified nonparametrically using only two order statistics of repeated measurements. The identification result confirms a hypothesis by Athey and Haile (2002) for a symmetric ascending auction model with unobserved heterogeneity. Extensions allow for heterogeneous measurement errors, broadening the applicability to additional empirical settings, including asymmetric auctions and wage offer models. We adapt an existing simulated sieve estimator and illustrate its performance in finite samples.

Keywords: Measurement error, Order statistics, Nonparametric identification, Spacing, Cross-sum

We thank the editor and anonymous referees, as well as Gaurab Aryal, Christian Gouriéroux, Philip Haile, Lixiong Li, José Luis Montiel Olea, Chen Qiu, Cailin Slattery, and conference participants for helpful comments. We are also grateful to Cristián Hernández, Daniel Quint, and Christopher Turansick for sharing their code. Yao Luo acknowledges support from SSHRC.

*Department of Economics, Binghamton University, 4400 Vestal Parkway E., PO Box 6000, Binghamton, NY 13902-6000. Email: jcho13@binghamton.edu.

†Department of Economics, University of Toronto, Max Gluskin House, 150 St. George St., Toronto, ON M5S 3G7. Email: yao.luo@utoronto.ca.

‡Department of Economics, Indiana University, 100 S. Woodlawn Ave., Bloomington, IN 47405. Email: rulixiao@iu.edu.

1 Introduction

We consider the classical measurement error model with repeated measurements:

$$X_j = \xi + \varepsilon_j, \qquad j \in \{1, \dots, n\},$$

where the latent variable of interest ξ is measured n times with i.i.d. measurement errors $(\varepsilon_j)_{j=1,\dots,n}$ that are independent of ξ. The identification result under such a model is well-established when at least *two* repeated measurements are observed, known as Kotlarski's lemma. The result has been widely applied in econometrics since its introduction by Li and Vuong (1998).[1] In practice, the researcher may observe only a subset of order statistics of the measurements, i.e., $(X_{(j)})_{j\in J}$, where $X_{(j)}$ is the j^{th} smallest order statistic from a sample of size n and $J \subset \{1, \dots, n\}$. This paper shows the underlying probability distributions of the latent variable and the measurement errors are identified from the joint distribution of the ordered measurements $(X_{(j)})_{j\in J}$.

The problem took its shape in Athey and Haile (2002) as a nonparametric identification problem in ascending auctions, which has particular relevance in auctions with electronic bidding. They conjecture that the model consists of enough structure to attain point identification using *two* order statistics. However, the question has been long-standing for two decades.[2] Importantly, the standard approach using Kotlarski's lemma fails because of dependence in the order statistics of measurement errors. Further, as Athey and Haile (2002) point out, the attempt to difference out the latent variable and exploit the spacing distribution is shown to be insufficient for point identification. This is evidenced by Rossberg (1972)'s counterexample.

While the literature has documented the increasing importance of unobserved heterogeneity in auctions,[3] the lack of identification results in the existing literature has hindered allowance for such heterogeneity in the classical fashion, unless relying on additional external variations in the data.[4] We fill this important gap by showing

[1] Examples include Bonhomme and Robin (2010), Krasnokutskaya (2011) and Grundl and Zhu (2019). An outline of uses of Kotlarski's lemma in econometrics is in Schennach (2016).

[2] Some positive findings are made recently in the finite mixture context; see Mbakop (2017) using five order statistics, Luo and Xiao (2023) using two and an instrument and Luo et al. (2021) using three. However, these papers do not tackle the original conjecture in the spirit of Kotlarski's lemma.

[3] For example, see Aradillas-López et al. (2013), Krasnokutskaya (2011), and Li et al. (2000).

[4] For example, Hernández et al. (2020) use variation in the number of bidders across auctions. Freyberger and Larsen (2022) circumvent the issue of dependence between order statistics using observed reserve prices, rendering the identification problem classical.

that the underlying distributions are identified with only data on *two* order statistics, which need not be consecutive or extreme, without relying on extra variations.

We propose a new identification strategy that exploits both features in Kotlarski (1967) and Rossberg (1972). In particular, we make use of *within* independence of the latent variable and the additively separable measurement errors as in Kotlarski (1967) to derive that the model imposes an additional restriction beyond the spacing of order statistics.[5] More precisely, we find that if two measurement error distributions both rationalize the observed distribution of the ordered measurements, then the joint distributions of *spacing* and *cross-sum* are identical.[6] Exploiting the structure of order statistics and commonly seen conditions (a support or tail restriction), we show that the spacing in conjunction with this additional restriction is indeed sufficient to point identify the underlying distribution using any two order statistics. The latent variable distribution is subsequently identified by a standard deconvolution argument.

We extend our main result to the setting where measurement errors are independent but nonidentically distributed (i.n.i.d.). We show that the underlying distributions are identified when two order statistics are observed, provided *some* measurement errors are known to be identically distributed and the data consist of group identities of high-order statistics. The result applies to asymmetric ascending auctions, where only high-order dropout bids are recorded, and asymmetric first-price auctions and wage offer settings, where consecutive and extreme order statistics are observed.

The outline of the paper is as follows. Section 2 introduces two motivating examples and Section 3 presents the identification results. To complement the identification result, in Section 4 we adapt Bierens and Song (2012)'s simulated sieve estimator to consistently estimate the unknown distributions. Section 5 concludes.

2 Motivating Examples

To motivate the identification problem, we introduce two examples: an ascending auction model with auction-specific unobserved heterogeneity and a model of wage offers where wage is determined by worker-specific labor productivity. Throughout the

[5]The term *spacing* usually refers to the difference between two *consecutive* order statistics. Here, we use it more broadly as the difference between any two order statistics.

[6]The term *cross-sum* refers to the sum of two order statistics of two independent random samples from distinct underlying parent distributions.

paper, we focus only on unobserved heterogeneity since adding exogenous covariates does not require novel considerations.

Example 2.1 (Ascending auction). *Consider an ascending auction with one indivisible good and n bidders. Each bidder's valuation for the item is determined by a set of item features—summarized and denoted by $\xi \in \mathbb{R}$—and a private value component ε_j that is independent of ξ and across bidders. The valuation of bidder j is given by $X_j = \xi + \varepsilon_j$. Note that the valuation of each bidder can be cast as a measurement of the value ξ with measurement error ε_j. See, e.g., Athey and Haile (2002).*

In an ascending auction, the price rises until only one bidder remains. Suppose the auctioneer records prices $P_1 \leq P_2 \leq \cdots$ at which bidders drop out. Assuming that the bidders play the dominant strategy of remaining in the auction until the price reaches their valuations, the dropout bids reveal the ordered valuations of the bidders, i.e., $P_j = X_{(j)}$. Importantly, the highest valuation $X_{(n)}$ is never observed since the auction ends when the price reaches $P_{n-1} = X_{(n-1)}$, and the bidder with the highest valuation wins. Therefore, in an ascending auction, we observe only an incomplete set of order statistics on bidders' valuations. For example, Kim and Lee (2014) observe at least the three highest dropout prices in ascending used-car auctions; see also Larsen (2021) for used-car auctions in the U.S. Data on timber auctions by the U.S. Forest Service—analyzed in a number of empirical papers, e.g., Haile and Tamer (2003), Athey et al. (2011), and Aradillas-López et al. (2013) to name a few—records at most top twelve bids regardless of the number of bidders.

Both the distributions of ξ and ε_j are of interest in an empirical analysis of auctions. For example, the counterfactual expected revenue to the auctioneer under a hypothetical auction rule requires the knowledge of both distributions.

Example 2.2 (Wage offer determination). *Consider a simple wage offer model $X_j = \xi + w_j + \varepsilon_j$ where X_j is the j^{th} (log-)wage offer an individual receives. X_j is assumed to be composed of the worker's productivity ξ, the wage rate w_j, and measurement error ε_j (on, e.g., labor productivity) associated with the offer. It is assumed that the period in consideration is relatively short such that the worker's productivity remains unchanged, but the worker may receive multiple offers; see Guo (2021).*

Data from the Survey of Consumer Expectations (SCE) Labor Market Survey records salary-related responses of up to three best job offers received by labor market participants within the last four months. In this setting, the researcher observes

$X_{(n)}, \ldots, X_{(n-k+1)}$ *where* $k = \min\{3, n\}$ *and* n *is the number of offers received. The identification results in the current paper show that the distributions* F_ξ *and* $F_{w_j+\varepsilon_j}$ *are nonparametrically identified.*

3 The Model and Main Results

In this section, we first formalize the i.i.d. framework considered throughout the paper and provide sufficient conditions to identify the latent variable and measurement error distributions. In Section 3.3, we consider i.n.i.d. measurement errors.

3.1 The Setup

We begin by stating the sampling process.

Assumption 3.1 (Sampling process)**.** *For each* $1 \leq j \leq n$, $X_j := \xi + \varepsilon_j$, *where the random variable* ξ *is independent of the random vector* $(\varepsilon_1, \ldots, \varepsilon_n)$.[7]

The researcher observes $(X_{(r)}, X_{(s)})$, the r^{th} and s^{th} order statistics from n observations, where r, s, and n are known and $1 \leq r < s \leq n$. That is, while the total number of measurements is known, one only observes two measurements of known ranks. In this section, we identify the unknown distributions of ξ and $(\varepsilon_1, \ldots, \varepsilon_n)$ using $F_{X_{(r,s)}}$, the joint distribution of $(X_{(r)}, X_{(s)})$, which is estimable from a random sample of the two order statistics.

For the benchmark case, we make the following distributional assumptions.

Assumption 3.2 (i.i.d. errors)**.**

(a) $\varepsilon_1, \ldots, \varepsilon_n$ *are i.i.d. with a common distribution* F_ε *on* $\mathbb{R}$.

(b) F_ε *is absolutely continuous with a probability density function* f_ε *that is light-tailed, i.e., for some* $C > 0$, $f_\varepsilon(\epsilon) = O(e^{-C|\epsilon|})$ *as* $|\epsilon| \to \infty$.

The tail condition in Assumption 3.2(b) is trivially satisfied when the support is bounded. If the support is unbounded on either side, the assumption restricts the tail of the density function to decay at an exponential rate.[8] The same assumption is found

[7] The identification results herein also applies to a setting with multiplicatively separable measurement error by taking logs when ξ and ε_j's are positive, as in Example 2.2.

[8] Throughout the paper, we use the term "light-tailed" to refer to densities with exponentially decaying tails as stated in the assumption.

in Evdokimov and White (2012). Alternatively, as in Kotlarski (1967) and Miller (1970), we may assume F_ε has a characteristic function (ch.f.) that is either (a.e.-)nonvanishing or analytic. Assumption 3.2(b) is a sufficient condition for the ch.f. of F_ε to be analytic. To our knowledge, there exists no result regarding its implication on the joint ch.f. of order statistics, the property we need for our identification results. Lemma A.1 in Appendix A shows that the joint ch.f. of the order statistics $(\varepsilon_{(r)}, \varepsilon_{(s)})$ (and thus its marginals) is also analytic under this assumption.

In addition to Assumption 3.2, we further restrict the support of the measurement errors as follows. For any distribution F on $\mathbb{R}$, let $S(F) \subseteq \mathbb{R}$ denote its support.[9]

Assumption 3.3 (Support condition)**.** *The measurement errors are bounded from below, which is normalized to zero, i.e.,* $\inf S(F_\varepsilon) = 0$.

Two aspects are worth mentioning regarding the above assumption. First, we require the measurement error to be bounded at one end; nevertheless, we allow the support to be possibly unbounded from above.[10] We do not assume the upper bound of the support is known nor assume any additional structure on the support, e.g., connectedness. Second, as is typical with measurement error models, the underlying distributions are identified only up to location. Although one may alternatively normalize the mean and have an unknown but finite lower bound, normalizing the lower bound turns out to be more convenient for our identification argument. Our identification strategy heavily utilizes the condition that one boundary of the support of F_ε is finite. On the other hand, the distribution of ξ is left completely unspecified.

The support restriction is nonrestrictive in many applications. The literature on games with incomplete information typically assumes bounded support of agent types (Athey (2001)), such as bidder valuation in empirical auctions (Guerre et al. (2000), Athey and Haile (2007)), private costs of exerting efforts in contest games (Huang and He (2021)), and private information about variable costs in Cournot games (Aryal and Zincenko (2021)). Observed wage offers are also bounded below by a positive constant in job search models (Burdett and Mortensen (1998), Guo (2021)), for example, when there is a reservation wage or a minimum wage that is nonbinding for the population of interest.

[9]The support of F is defined as the smallest closed set $K \subseteq \mathbb{R}$ such that $P_F(K) = 1$, where P_F is the Borel probability measure induced by F. Thus, the density $f(x)$ may be zero for some values $x \in K$, including points on the boundary of K.

[10]The results herein apply to the case when only the upper bound is finite, e.g., $S(F_\varepsilon) = (-\infty, 0]$, by considering $(X'_{(r')}, X'_{(s')}) = (-X_{(s)}, -X_{(r)})$ where $r' = n - s + 1$ and $s' = n - r + 1$.

3.2 Nonparametric Identification

Throughout the section, we treat the observable joint distribution $F_{X_{(r,s)}}$ as known (i.e., as a datum), and show that the distribution F_ξ of the latent variable ξ and F_ε are identified under the aforementioned assumptions. The bulk of our identification result is concerned with showing that there is a unique measurement error distribution that rationalizes the observed distribution $F_{X_{(r,s)}}$. Then the latent variable distribution is identified by a standard deconvolution argument. To focus on identifying the measurement error distribution, we first isolate the empirical content about the measurement errors.[11] In Lemma 3.1 below, we do so by exploiting the multiplicative structure of the ch.f. of a sum of independent random variables.

In the following lemmas, let $\eta_1, \ldots, \eta_n$ be n i.i.d. copies of $\eta \in \mathbb{R}$ and let $(\eta_{(r)}, \eta_{(s)})$ be order statistics. $\psi_{\eta_{(r,s)}}$ and $\psi_{\eta_{(r)}}$ denote the joint and marginal ch.f., respectively.

Lemma 3.1. *Suppose the sampling process is as described in Assumption 3.1 and the measurement errors satisfy Assumption 3.2. If F (on $\mathbb{R}$) is a data-consistent measurement error distribution,*[12] *then*

$$\frac{\psi_{X_{(r,s)}}(t_r, t_s)}{\psi_{X_{(j)}}(t_r + t_s)} = \frac{\psi_{\eta_{(r,s)}}(t_r, t_s)}{\psi_{\eta_{(j)}}(t_r + t_s)}, \quad \textit{for all } (t_r, t_s) \in B_0,\ j \in \{r, s\}, \tag{1}$$

where $\eta \sim F$ and B_0 is an open ball in $\mathbb{R}^2$ centered at zero. Further, F induces a unique data-consistent latent variable distribution.

Suppose F and G are two measurement error distributions that are data-consistent. Lemma 3.1 states that the order statistics $(\eta_{(r)}, \eta_{(s)})$ from the parent distribution F and $(\eta'_{(r)}, \eta'_{(s)})$ from the parent distribution G must have the same ratio of joint and marginal ch.f.s.

To obtain identification, it remains to show that such a measurement error distribution is unique, i.e., $F = G$. Lemma 3.1 implies that any two data-consistent

[11] In the existing literature that expands on Kotlarski (1967), the prevalent identification strategy is to begin with identifying the distribution of the latent variable. Nonetheless, alternative strategies that begin with identifying the measurement error distribution exist in the classical repeated measurement error setting; see, e.g., Hall and Yao (2003) and Evdokimov and White (2012).

[12] We say a pair of distribution functions (G_ξ, G_ε) rationalizes the data, or is data-consistent, if $\xi' \sim G_\xi$ and $(\varepsilon'_j)_{j=1,\ldots,n} \sim \times_{j=1}^n G_\varepsilon$, independent of ξ', implies $(\xi' + \varepsilon'_{(r)}, \xi' + \varepsilon'_{(s)}) =_d (X_{(r)}, X_{(s)})$. We say G_ξ (resp., G_ε) is data-consistent if (G_ξ, G_ε) is data-consistent for some G_ε (resp., G_ξ).

measurement error distributions must satisfy

$$\psi_{\eta_{(r,s)}}(t_r,t_s)\psi_{\eta'_{(j)}}(t_r+t_s)=\psi_{\eta'_{(r,s)}}(t_r,t_s)\psi_{\eta_{(j)}}(t_r+t_s),\quad j\in\{r,s\},\tag{2}$$

for all $(t_r,t_s)\in B_0$. In fact, the equality extends to all of $\mathbb{R}^2$ because all ch.f.s in (2) are analytic.[13] Finally, noticing that the ch.f. of the sum of two independent random vectors is multiplicatively separable, the following lemma establishes necessary conditions for any two data-consistent measurement error distributions.

Lemma 3.2. *Let $\eta_{(r)}$ and $\eta_{(s)}$ be two order statistics of a random sample of size n from F, and $\eta'_{(r)}$ and $\eta'_{(s)}$ be two order statistics of a random sample of size n from G, where $(\eta_{(r)},\eta_{(s)})$ and $(\eta'_{(r)},\eta'_{(s)})$ are independent. Under Assumptions 3.1, if F and G (on $\mathbb{R}$) are two data-consistent measurement error distributions that admit light-tailed densities,[14] then*

$$Z_{1j}:=\begin{pmatrix}\eta'_{(j)}+\eta_{(r)}\\ \eta'_{(j)}+\eta_{(s)}\end{pmatrix}\overset{d}{=}\begin{pmatrix}\eta_{(j)}+\eta'_{(r)}\\ \eta_{(j)}+\eta'_{(s)}\end{pmatrix}=:Z_{2j},\quad j\in\{r,s\}.\tag{3}$$

The lemma shows that if F and G are two data-consistent measurement error distributions, then the two surrogate measurement error models in (3) are observationally equivalent, where Z_{1j} is order statistics of measurements of $\eta'_{(j)}$ with errors η_i's from parent distribution F and Z_{2j} is order statistics of measurements of $\eta_{(j)}$ with errors η'_i's from parent distribution G. As stated below in Corollary 3.1, the lemma has an important implication: F and G not only have the same spacing distributions (i.e., $\eta_{(s)}-\eta_{(r)}=_d\eta'_{(s)}-\eta'_{(r)}$) but also have the same distributions for what we call cross-sums (i.e., $\eta'_{(s)}+\eta_{(r)}=_d\eta_{(s)}+\eta'_{(r)}$). The latter information is not exploited in the classical setting; however, it turns out to be relevant information for identification when only order statistics of measurements are observed.

Because both Z_{1r} and Z_{2r} involve three order statistics from two potentially different parent distributions F and G, it appears difficult to show directly from Lemma 3.2 that F and G are the same. A reasonable attempt to tackle the problem would be to

[13] Lemma A.1 shows that order statistics from light-tailed parent densities have analytic joint (and thus marginal) ch.f. Since the product of two analytic functions is also analytic, (2) shows that two analytic functions coincide on an open ball B_0 in $\mathbb{R}^2$, which implies they coincide everywhere. Thus Assumption 3.2(b) plays a crucial role in pinning down the entire distribution based only on the information of the ch.f. about the origin.

[14] Cf. Assumption 3.2(b) and footnote 8.

explore features of Z_{1r} and Z_{2r} that depend on the parent distributions in a simple manner. Under the support condition in Assumption 3.3, we show that the distributions of Z_{1r} and Z_{2r} depend on only one parent distribution upon conditioning on an extreme event.[15] Thus the joint distribution of the cross-sum and spacing together with Assumption 3.3 provide information that is sufficient to claim any two data-consistent measurement error distributions are the same, i.e., F_ε is identified.

Lemma 3.3. *Let $\eta_{(r)}$, $\eta_{(s)}$, $\eta'_{(r)}$ and $\eta'_{(s)}$ be as specified in Lemma 3.2. If measurement error distributions F and G admit light-tailed densities and* $\inf S(F) = \inf S(G) = 0$*, we have, for all $c \in \mathbb{R}$,*

$$\lim_{\delta \downarrow 0} \mathbb{P}(\eta'_{(r)} + \eta_{(s)} \leq c \mid \eta'_{(r)} + \eta_{(r)} \leq \delta) = F_{s-r:n-r}(c), \tag{4}$$

$$\lim_{\delta \downarrow 0} \mathbb{P}(\eta_{(r)} + \eta'_{(s)} \leq c \mid \eta_{(r)} + \eta'_{(r)} \leq \delta) = G_{s-r:n-r}(c), \tag{5}$$

where $F_{s-r:n-r}$ (resp., $G_{s-r:n-r}$) denotes the distribution of the $(s-r)^{th}$ order statistic of a random sample of size $n-r$ from F (resp., G).

For the sake of intuition, consider a simple variant of (4) that conditions on the event $\{\eta'_{(r)} + \eta_{(r)} = 0\}$, assuming the conditioning is well-defined.[16] This event is equivalent to that where both $\eta'_{(r)} = 0$ and $\eta_{(r)} = 0$, which simplifies the conditional distribution to $\mathbb{P}(\eta_{(s)} \leq c \mid \eta_{(r)} = 0)$. Standard arguments in order statistics (cf. Theorem 2.4.2 in Arnold et al. (2008)) suggest that this conditional distribution has a simple form:

$$\mathbb{P}(\eta_{(s)} \leq c \mid \eta_{(r)} = 0) = F_{s-r:n-r}(c).$$

In words, this equality says that the conditional distribution of the s^{th} order statistic when r observations take the lowest possible value is the same as the distribution of the $(s-r)^{\text{th}}$ order statistic obtained from a sample of size $n-r$.

Lemmas 3.1–3.3 show that if F and G are both data-consistent, then the condi-

[15] The identification argument here and below investigates the implications of the condition $Z_{1r} =_d Z_{2r}$ only. $Z_{1s} =_d Z_{2s}$ in (3) is the relevant condition to exploit under the assumption, in place of Assumption 3.3, that the measurement error is bounded from above (cf. footnote 10).

[16] The event $\{\eta'_{(r)} + \eta_{(r)} = 0\}$ may not be well-defined as it occurs with zero probability. In particular, the event $\{\eta_{(r)} = 0\}$ always has zero density unless $r = 1$, i.e., the minimum order statistic. Further, if $f(0) = 0$ where f is the density of η, even the minimum order statistic has zero density at $\eta_{(r)} = 0$. In Lemma 3.3, we make rigorous the intuitive claim provided here by considering the limiting argument as in (4) and (5), which is well-defined.

tional distributions (4) and (5) must be equal, i.e., for all $c \in \mathbb{R}$,

$$F_{s-r:n-r}(c) = G_{s-r:n-r}(c).$$

As the distribution of an order statistic uniquely identifies the parent distribution, we can conclude that $F = G$.[17] We formally state the main identification result for the i.i.d. case.

Theorem 3.1. *Under Assumption 3.1, if the measurement errors satisfy Assumptions 3.2 and 3.3, both F_ξ and F_ε are identified.*

A Discussion on Rossberg (1972)'s Counterexample

Athey and Haile (2002) point out that exploiting the distribution of spacing between two order statistics is insufficient to point identify their parent distribution. Their discussion relies on a counterexample by Rossberg (1972). A straightforward corollary to Lemma 3.2 highlights the model restrictions we exploit in addition to the spacing.

Corollary 3.1. *Under the same assumptions as in Lemma 3.2, if F and G are two data-consistent measurement error distributions that admit light-tailed densities, then*

$$\begin{pmatrix} \eta_{(s)} - \eta_{(r)} \\ \eta'_{(r)} + \eta_{(s)} \end{pmatrix} \overset{d}{=} \begin{pmatrix} \eta'_{(s)} - \eta'_{(r)} \\ \eta_{(r)} + \eta'_{(s)} \end{pmatrix} \quad \textit{and} \quad \begin{pmatrix} \eta_{(s)} - \eta_{(r)} \\ \eta_{(r)} + \eta'_{(s)} \end{pmatrix} \overset{d}{=} \begin{pmatrix} \eta'_{(s)} - \eta'_{(r)} \\ \eta'_{(r)} + \eta_{(s)} \end{pmatrix}.$$

If F and G are two data-consistent measurement error distributions, then they have the same spacing and cross-sum distributions. In the classical setting, the difference (i.e., spacing) between two independent measurement errors identifies the underlying measurement error distribution (cf. Hall and Yao (2003)). However, in the current context, the spacing between two order statistics of measurement errors is insufficient to identify the underlying (parent) distribution as evidenced by Rossberg (1972)'s counterexample. The additional information we incorporate in order to obtain identification lies in the cross-sums, a condition which emerges from exploiting the within independence between the latent variable and measurement errors when we take the ratio of ch.f.s. Combining the information contained in spacing as investigated by Rossberg (1972) and the information contained in within independence as

[17]The one-to-one mapping between the distribution of an order statistic and the parent distribution is a standard result, e.g., see David and Nagaraja (2003), p.10, (2.1.5). The mapping in (2.1.5) is an invertible map of the c.d.f. F because $p \mapsto I_p(a, b)$ is strictly monotone.

investigated by Kotlarski (1967) thus delivers the current identification result. The difference between Rossberg (1972)'s nonidentification result and our identification result is explained in more detail in Appendix B.

3.3 Extensions of the Identification Result

We extend our identification result to the case of independent but nonidentically distributed (i.n.i.d.) measurement errors. To motivate the extension to the problem, we expand on Example 2.1 by introducing bidder asymmetry and discuss additional features available in some auction data.

Example 3.1 (Asymmetric auction)**.** *In empirical auctions, bidders are said to be asymmetric if the bidders' private values $\varepsilon_1, \ldots, \varepsilon_n$ have different marginal (or parent, in the case of order statistics) distributions. Such asymmetry arises naturally in procurement auctions, where contractors differ in cost efficiency and productivity (see, e.g., Flambard and Perrigne (2006)), and in timber auctions, where mills have the manufacturing capacity and loggers do not (see, e.g., Athey et al. (2011)).*[18]

In order to identify bidder-specific valuation distributions, bidder identities must be observable. Fortunately, the auctioneer often publishes such identities despite some bids being missing; see, e.g., Athey et al. (2011). Another common practice in the asymmetric auction literature is grouping bidders by commonly known bidder types, such as mills and loggers in Athey et al. (2011) and strong and weak bidders in Luo et al. (2018), which leads to more tractable theory and empirics. Such grouping uses additional information about the bidders, typically available as a public record, such as bidders' manufacturing capacity in a timber auction and pre-qualified contractors' experience in Department of Transportation procurement auctions.

In this section, we first show that any two order statistics suffice to point identify the underlying distributions when there are a relatively small degree of heterogeneity in measurement errors and some membership information about them. In particular, as a cost of relaxing the homogeneity assumption, we require observing membership information about measurement errors that correspond to high-order statistics (for ranks above r). We then discuss the implication of the result for measurement errors that are completely heterogeneous.

[18] In a wage offer setting as in Example 2.2 employers may value the same productivity differently, which may result in heterogeneous measurement errors.

Assumption 3.4 (i.n.i.d. errors)**.**

(a) $\varepsilon_1, \ldots, \varepsilon_n$ *are independent with distributions* $F_{\varepsilon_1}, \ldots, F_{\varepsilon_n}$ *on* $\mathbb{R}$*, respectively.*

(b) For each $j \in \{1, \ldots, n\}$*,* F_{ε_j} *admits a probability density function* f_{ε_j} *that is light-tailed, i.e.,* $f_{\varepsilon_j}(\epsilon) = O(e^{-C_j|\epsilon|})$ *as* $|\epsilon| \to \infty$ *for some* $C_j > 0$*.*

(c) For each $j \in \{1, \ldots, n\}$*,* $\inf S(F_{\varepsilon_j}) = 0$*.*

Assumption 3.4(c) assumes both finite and common support lower bound across distinct measurement error distributions, the latter of which holds trivially in the i.i.d. case. Note that apart from the lower boundary, the supports may differ (cf. Corollary 3.2). The following assumption records any prior information on the different types of measurement errors, including an additional support condition.

Assumption 3.5 (Group structure)**.** *There exists a partition* $g_1, \ldots, g_p$ *of* $\{1, \ldots, n\}$ *such that* $\varepsilon_j =_d \varepsilon_k$ *if* $j, k \in g_q$ *for some* $q \in \{1, \ldots, p\}$*. In addition, the measurement errors* $\varepsilon_1, \ldots, \varepsilon_n$ *have common support.*

Assumption 3.5 posits there are at most p distinct measurement error distributions. The case $p = 1$ corresponds to the i.i.d. case in Section 3.2; and $p = n$ corresponds to the case with no prior information on homogeneity. We do not preclude the possibility that two groups g_q and $g_{q'}$ have the same distribution beyond the researcher's knowledge. For example, the extreme case $p = n$ subsumes the i.i.d. setting as a special case. Assumption 3.5 implicitly assumes that the group structure is held constant across observations. In an application to auctions, the assumption may be imposed by restricting to auctions with the same composition of bidder types when all bidder types of participants are available in the dataset. In such a case, the analysis should be interpreted conditionally on the bidder composition.

The common support assumption is a sufficient condition to guarantee that all measurement error distributions are identified on the entirety of their support. Otherwise, some distributions may be identified only on a strict subset of their support, as we highlight in a discussion below. Let $R_{(j)} = \{q : X_{(j)} = X_k \text{ for some } k \in g_q\}$ be the group identity of the j^{th} order statistic.[19]

Theorem 3.2. *Suppose Assumptions 3.1, 3.4, and 3.5 hold. Further, suppose two order statistics and the group identities of high-order statistics* $(X_{(r)}, X_{(s)}, R_{(r+1)}, \ldots, R_{(n)})$

[19] Under Assumption 3.4(b), $R_{(j)}$ is singleton with probability 1. Thus, we abuse notation and use $R_{(j)}$ to denote both the set and the a.s. unique element in $\{1, \ldots, p\}$.

are observed. If there exists a group g_q with at least $n-r$ members, both F_ξ and $(F_{\varepsilon_j} : j = 1, \ldots, n)$ are identified.

Note that the identification result does not require the researcher to observe the group identity $R_{(r)}$ of the r^{th} order statistic or those of lower order statistics. A heuristic explanation is provided in a discussion below. Theorem 3.2 has an important implication when the measurement errors are left completely heterogeneous, i.e., $p = n$. In particular, the theorem implies that the observed order statistics should be consecutive and extreme because there is no group of a larger size, i.e., $n - r = 1$.[20]

Corollary 3.2. *Suppose Assumptions 3.1 and 3.4 hold, and $(X_{(n-1)}, X_{(n)}, R_{(n)})$ are observed. Both F_ξ and $(F_{\varepsilon_j} : j = 1, \ldots, n)$ are identified.*

So as to appreciate the additional conditions assumed in Theorem 3.2, we first illustrate the identification strategy in the setting of Corollary 3.2, which delivers a simpler argument. Suppose results similar to Lemmas 3.1–3.3 hold in the i.n.i.d. case, in the sense that two sets of data-consistent measurement error distributions $(F_j : j = 1, \ldots, n)$ and $(G_j : j = 1, \ldots, n)$ must satisfy

$$\mathbb{P}(\eta_{(n)} \leq c \,|\, \eta_{(n-1)} = 0) = \mathbb{P}(\eta'_{(n)} \leq c \,|\, \eta'_{(n-1)} = 0), \tag{6}$$

for every constant c. Had the measurement errors been i.i.d., (6) corresponds the equivalence of two parent distributions. In the i.n.i.d. case, the top-order statistic may arise from any of the parent distributions. Intuition suggests that this should be a mixture of measurement errors from different parent distributions. Since the mixing weights depend on $(F_j : j = 1, \ldots, n)$, it appears formidable to show its equivalence with $(G_j : j = 1, \ldots, n)$ from (6) without additional assumptions. Alternatively, suppose—as we formally show in the proof of Theorem 3.2—data-consistency implies a condition similar to (6) conditional on the member index of the highest order statistic. Heuristically speaking, because the index is known, say j, the conditional distribution simplifies to the j^{th} marginal distribution (i.e., a mixture with trivial weights). Thus $F_j = G_j$ and the j^{th} measurement error distribution is identified. Under the assumption of a common support lower bound, each measurement error has a nonzero chance of being the largest, which implies all n measurement error distributions are identified.

[20]The support condition in Assumption 3.5 plays no role in this corollary because the observed order statistics are the largest among all measurement errors.

Corollary 3.2 does not grant identification of an ascending auction model when the data on dropout bids is incomplete; e.g., see Freyberger and Larsen (2022). Nonetheless, a group structure as in Theorem 3.2 is commonly present in the empirical auction literature. Since Theorem 3.2 does not rely on observing extreme or consecutive order statistics, the result may be applicable even if the bid data is incomplete.

To illustrate the identification strategy in the setting of Theorem 3.2, suppose out of $n = 4$ measurements, one observes $(X_{(2)}, X_{(3)}, R_{(3)}, R_{(4)})$ and it is known *a priori* that two of the measurement errors have a common distribution (call it group 1 and the other groups 2 and 3). By conditioning on the event that $\{R_{(3)} = R_{(4)} = 1\}$, we can homogenize the conditional distribution of the 3^{rd}-order statistic of measurement errors conditional on the 2^{nd}-order statistic taking the smallest possible value, which allows us to identify the measurement error distribution for group 1. This rests on (i) being able to observe the group identities and (ii) group 1 being large enough to condition on such an event. Provided the measurement errors have a common support, the remaining group distributions can be identified sequentially. For example, the group-2 distribution can be identified by conditioning on $\{R_{(3)} = 2, R_{(4)} = 1\}$.[21]

Remark 3.1. *Corollary 3.2 has natural applications in first-price auctions and wage offer settings, where consecutive and extreme order statistics are common and identities are observed. For instance, the Washington State Department of Transportation archives three apparent low bids and bidder identities of six months or older online. FDIC auction data contain the winning and second-highest bids and the associated identities (Allen et al. (2023)). U.S. Forest Service timber auctions only record up to top twelve bids and bidder identities regardless of the number of bidders. Lastly, data from the Survey of Consumer Expectations (SCE) Labor Market Survey records salary-related responses on the three best offers for those who received more than three offers within the last four months.*

Remark 3.2. *If* all *dropout bids are observed, Corollary 3.2 is applicable to ascending auctions. Specifically, if private values have a common upper bound, i.e.,* $\sup S(F_{\varepsilon_j}) = \overline{\varepsilon} < +\infty$*, the lowest order statistics* $(X_{(1)}, X_{(2)})$ *identify the underlying distributions.*

[21] If one measurement error has larger support than another, the distribution may not be fully identified. If group-1 measurement errors have support $[0, 1]$ but group 2 has support $[0, 2]$, regardless of whether one conditions on $\{R_{(3)} = 1, R_{(4)} = 2\}$ or $\{R_{(3)} = 2, R_{(4)} = 1\}$, the 3^{rd}-order statistic only has support $[0, 1]$. Thus group-2 measurement error distribution is not identified on $(1, 2]$.

4 Nonparametric Estimation

We propose a simulated sieve estimator under the i.i.d. measurement error framework, which can be easily modified for the i.n.i.d. case, albeit with the curse of dimensionality. The procedure, motivated by the sieve estimator in Bierens and Song (2012), estimates the distribution functions of the latent variable and of the measurement error simultaneously.[22] For a general survey of the method of sieves, see Chen (2007).

4.1 A Simulated Sieve Estimator

We follow closely the development in Bierens (2008) to specify the parameter space and its sieve space. It has the advantage that we can incorporate prior information on the support without having to choose different orthogonal bases. This is an attractive feature for our purpose because, unlike the latent variable ξ, the measurement errors are restricted to be nonnegative-valued. The sieve space is constructed using only Legendre polynomials instead of using, e.g., Hermite polynomials for the latent variable and Laguerre polynomials for the measurement errors.

The construction of the sieve space begins with the observation that any absolutely continuous c.d.f. F on $\mathbb{R}$ can be expressed as $F(\cdot) = (H \circ G)(\cdot)$ where H is an absolutely continuous c.d.f. on $[0,1]$ and G is an absolutely continuous c.d.f. that is strictly increasing on $S(F)$. Equivalently,

$$F(\cdot) = \int_0^{G(\cdot)} h(u)\,du = \int_0^{G(\cdot)} \pi^2(u)\,du, \tag{7}$$

where $h = \pi^2$ is the density of H and π is a Borel-measurable square-integrable function on $[0,1]$. Thus, e.g., with a fixed G with support $S(G) = [0,\infty)$, a large enough set $\mathcal{P}$ of Borel-measurable square-integrable functions maps via (7) to a large enough set of distribution functions with support contained in $[0,\infty)$.

The sieve space is constructed by using orthonormal polynomials to approximate π. Bierens (2008) considers the following compact set of square-integrable functions

$$\mathcal{P} := \left\{ \pi(\cdot) = \frac{1 + \sum_{\ell=1}^{\infty} \delta_\ell \rho_\ell(\cdot)}{\sqrt{1 + \sum_{\ell=1}^{\infty} \delta_\ell^2}} : |\delta_\ell| \le \frac{c}{1 + \sqrt{\ell}\ln \ell},\ \ell = 1, 2, \ldots \right\}, \tag{8}$$

[22] Two-sample approach in Carroll et al. (2010) does not apply to our setting in which one latent variable is measured with correlated measurement errors.

for some large constant $c > 0$, where ρ_k, defined on $[0,1]$, is a recentered and rescaled Legendre polynomial of order k: if ℓ_k is a Legendre polynomial of order k on $[-1,1]$, then $\rho_k(u) := \sqrt{2k+1}\ell_k(2u-1)$ for $u \in [0,1]$ (see Section 2.2 in Bierens (2008)). This implicitly defines a compact parameter space $\mathcal{F}$ for the unknown c.d.f. F by (7) for some fixed G and $\pi \in \mathcal{P}$. The sieve $\{\mathcal{F}_k\}_k$ is constructed by truncating the series in (8) at order k. Since we estimate two distribution functions F_ξ and F_ε, we consider a product sieve space $\{\mathcal{F}_k^\xi \times \mathcal{F}_k^\varepsilon\}_k$ for two choices of G, denoted G_ξ and G_ε.

As Bierens (2008) notes, the c.d.f. G not only restricts the support but also acts as an initial guess of the unknown c.d.f.[23] With prior information on the shape of the distribution functions, an educated initial guess helps reduce the approximation error resulting from low-order sieve spaces. Without any prior information, one clearly cannot expect any *a priori* advantage to choosing a particular distribution. In light of the often-used standard, Hermite, and Laguerre polynomial sieves, it may be reasonable to choose a uniform distribution if the support is known to be contained in a bounded interval, a normal distribution if one is agnostic about the support, and an exponential distribution if the support is known to be contained in $[0,\infty)$.

We consider a sieve extremum estimator that minimizes the average (squared) contrast between two empirical ch.f.s: one based on the factual sample and the other based on simulated draws. The population criterion function is devised as follows. For any candidate pair of distribution functions $F = (F_\xi, F_\varepsilon)$, let

$$X_{(r,s)}(F) = (X_{(r)}(F), X_{(s)}(F)) := (\xi(F_\xi) + \varepsilon_{(r)}(F_\varepsilon), \xi(F_\xi) + \varepsilon_{(s)}(F_\varepsilon)),$$

where $\xi(F_\xi)$ is a random draw from F_ξ and $\varepsilon_{(j)}(F_\varepsilon)$ is the j^{th} order statistic of n i.i.d. draws from F_ε. For any $t = (t_r, t_s) \in \mathbb{R}^2$, let $\varphi(t;F) := \mathbb{E}e^{it^\top X_{(r,s)}(F)}$ denote its ch.f. where the expectation is induced by the $(n+1)$ uniform draws in the simulation process described below. We consider the following population criterion function:

$$Q(F) := \frac{1}{4\kappa^2} \int 1\left\{t \in (-\kappa,\kappa)^2\right\} \left|\psi_{X_{(r,s)}}(t) - \varphi(t;F)\right|^2 dt, \tag{9}$$

[23]The flexibility of choosing a base distribution G in Bierens (2008) is more a norm than an exception in nonparametric methods using orthogonal bases. The standard polynomial sieve on the unit interval may be seen to have the uniform distribution as an initial guess. The Hermite and Laguerre polynomial sieves take normal and exponential distributions as initial guesses, respectively. While these "initial guesses" are determined naturally by the weight function associated with the orthogonal bases, the construction in Bierens (2008) allows for an explicit choice by the researcher.

where $\kappa > 0$ is a tuning parameter that determines the integration region. In place of the box weight $1\{(t_r, t_s) \in (-\kappa, \kappa)^2\}$, one may also consider a smooth weight. We choose the former because it has a closed-form expression for the empirical criterion function (see Appendix C).[24] We define a simulated sieve extremum estimator:

$$\widehat{F}_N \coloneqq (\widehat{F}_{\xi,N}, \widehat{F}_{\varepsilon,N}) \in \operatorname*{arg\,min}_{F \in \mathcal{F}^{\xi}_{k_N} \times \mathcal{F}^{\varepsilon}_{k_N}} \widehat{Q}_N(F), \tag{10}$$

where $\{k_N\}$ is an arbitrary sequence of positive integers such that $k_N \to \infty$ and $\widehat{Q}_N(F)$ is the empirical criterion function with the empirical ch.f. $\widehat{\psi}_N$ and the simulated ch.f. $\widehat{\varphi}_N(\cdot; F)$ in place of $\psi_{X_{(r,s:n)}}$ and $\varphi(\cdot; F)$, respectively. $\widehat{\varphi}_N(\cdot; F)$ is constructed from N repeated draws of $(n+1)$ uniform random variables $(V, U_1, \ldots, U_n)$ and computing the inverse transform $X_{(j)}(F_k) = F^{-1}_{\xi,k}(V) + F^{-1}_{\varepsilon,k}(U_{(j)})$ for $j \in \{r, s\}$.

We show that the estimator is consistent under the following set of assumptions.

Assumption 4.1 (Consistency)**.**

(a) $\{(X_{(r),i}, X_{(s),i})\}_{i=1,\ldots,N}$ *are* N *i.i.d. realizations of* $(X_{(r)}, X_{(s)})$*, where* $(X_{(r)}, X_{(s)})$ *has a bounded support.*

(b) G_ξ *and* G_ε *are known absolutely continuous c.d.f.s with support on* $\mathbb{R}$ *and* $[0, \infty)$*, respectively.*

(c) Given G_ξ *and* G_ε *in part (b), the pair of true underlying distributions* (F_ξ, F_ε) *is in the closure of* $\bigcup_k \mathcal{F}^{\xi}_k \times \mathcal{F}^{\varepsilon}_k$*.*

(d) $\{(V_i, U_{1,i}, \ldots, U_{n,i})\}_{i=1,\ldots,N}$ *are* N *i.i.d. draws from* $\mathcal{U}(0,1)^{n+1}$*, independent of the sampling process.*

The support restriction in Assumption 4.1(a) is a sufficient condition to ensure that the measurement errors have light tails. Despite being restrictive, this appears to be the most straightforward assumption to impose on the *observables* to guarantee identification.[25] Note that the researcher does not have to know *a priori* the true support of the observables, nor the implied bounded supports for F_ξ and F_ε.

[24]While a data-driven choice of κ (as well as the polynomial order k_N in Theorem 4.1 below) may be of interest, we do not explore its possibility here.

[25]Note that as remarked in Section 3, one may consider alternatives to Assumption 3.2(b), e.g., (a.e.-)nonvanishing or analytic ch.f.s, and still achieve the same identification result. Thus, one may also consider estimating the ch.f.s over a space of analytic functions and then estimate the densities via inverse Fourier transform. Clearly, various other estimation methods can be considered. For example, one may consider a sequential approach where one estimates the measurement error distribution via (1) in the first stage and then estimate the latent variable distribution using nonparametric deconvolution in the second stage. Alternatively, one may consider estimating

Assumption 4.1(b) restricts the support of the base distribution for F_ε to be in line with the normalization in Assumption 3.2(b). Assumption 4.1(c) is a standard assumption that the model is correctly specified. Assumption 4.1(d) specifies the simulation process. Note that the random draws $(V_i, U_{j,i})_{j,i}$ are obtained once and not repeatedly drawn across different candidate parameter values $F_{\xi,k}$ and $F_{\varepsilon,k}$ in order to ensure that the criterion function is continuous with respect to the parameter.

Theorem 4.1. *Let $\kappa > 0$ and let $\{k_N\}_N$ be any sequence of positive integers such that $k_N \to \infty$. Under Assumptions 3.1–3.3 and 4.1, the estimator in (10) is strongly uniformly consistent, i.e.,*

$$\max\left\{\|\widehat{F}_{\xi,N} - F_\xi\|_\infty, \|\widehat{F}_{\varepsilon,N} - F_\varepsilon\|_\infty\right\} \xrightarrow{a.s.} 0.$$

Steps for implementing the estimator is described in Appendix C and its finite sample performance is illustrated below. As is standard in nonparametric estimation, the choice of the sieve order k_N affects the approximation bias and variance of the estimator. An information-criterion-based approach analogous to that found in Bierens and Song (2012) may be used to choose the sieve order k_N, although the procedure may be computationally intensive. With larger κ, the estimator is expected to perform better as it accounts for more discrepancy between the two ch.f.s. There appears to be no theoretical reason to restrict κ except to ensure that the criterion function is well-defined, but limited simulation suggests the aggregation may come with larger variance. We do not have a useful criterion, but in light of the discussion, one may consider minimizing $\widehat{Q}_N$ over κ as well as F with a large upper bound on the parameter space for κ. From simulation studies, we find that the estimator is less sensitive to the choice of κ as long as κ is not too small. So we set $\kappa = 1$ as its baseline value in this paper, as is done in Bierens and Song (2012), and explore its properties in a separate paper. We also leave inference procedures for future research.[26]

the densities via a sieve maximum likelihood. We consider the proposed simulated sieve extremum estimator as it both allows to estimate the underlying distributions simultaneously and admits a closed-form objective function.

[26]Valid inference procedures exist in similar settings, for example, with independent measurement errors (Kato et al., 2021) or order statistics without unobserved heterogeneity (Menzel and Morganti, 2013). A common feature they tackle is a certain lack of continuity in the inverse problems. Similar irregularity concerns may have to be addressed here.

4.2 Monte Carlo Evidence

To illustrate the performance of the proposed estimator, we conduct a simple Monte Carlo experiment. We begin by describing the data generating process (DGP). For observation i, the variable of interest ξ_i is measured $n = 3$ times with i.i.d. measurement errors $\varepsilon_{1,i}, \ldots, \varepsilon_{3,i}$. The measurement is constructed as $X_{j,i} = \xi_i + \varepsilon_{j,i}$. We assume that only two order statistics of ranks $r = 1$ and $s = 2$ remain. That is, only $X_{(1),i}$ and $X_{(2),i}$ are recorded. We repeat the process to obtain N pairs of observations. The experiment is replicated $R = 500$ times to obtain 500 random samples of size N of the form $\{x^{(r)}_{(1),i}, x^{(r)}_{(2),i}\}_{i=1,\ldots,N}$.

For the distributions of ξ_i and $\varepsilon_{j,i}$, we set up a design that resembles Hernández et al. (2020)'s application on eBay Motors auctions. Specifically, we use their estimated distributions of unobserved heterogeneity (F_ξ in our set-up) and private values (F_ε in our set-up) to calibrate the DGP for our simulation exercise. To construct these two distributions, we approximate the estimates in Figure 4 of Hernández et al. (2020) with a sieve of order 6, which resulted in almost identical distributions to those in the original article.

We then simulate data from the DGP and investigate finite-sample performance of our estimator. We consider sample sizes $N = 1000$, 2000, and 4000 with $k = 4$, 5, and 6, respectively. Note that by construction, there is no sieve approximation error when $N = 4000$, i.e., any estimation error is associated only with sampling error. Furthermore, we set the bandwidth $\kappa = 1$, 3.14, and 5, and the base distribution $G_\xi = \mathcal{N}(0, 1/4)$ for ξ and $G_\varepsilon = \mathcal{N}(2, 1)_+$ for ε, where $\mathcal{N}(2, 1)_+$ denotes the truncated normal between 0 and ∞.

We present the estimation results for F_ξ and F_ε with $\kappa = 1$ in Figure 1. Simulation results with $\kappa = 3.14$ and 5 are similar and are presented in Figures 4 and 5 in Appendix C. The true distribution functions are shown in black. We also plot some randomly selected estimates along with some box plots that illustrate the pointwise sampling error of $\widehat{F}_\xi$ and $\widehat{F}_\varepsilon$ at various evaluation points. The figure suggests that the estimator performs reasonably well under all three sample sizes, and the performance improves with larger sample size. Although the choice of κ does not seem to affect the behavior of the estimator in any ill-behaved manner, there appear to be larger pointwise variance when $\kappa = 3.14$ and 5 relative to the case when $\kappa = 1$.

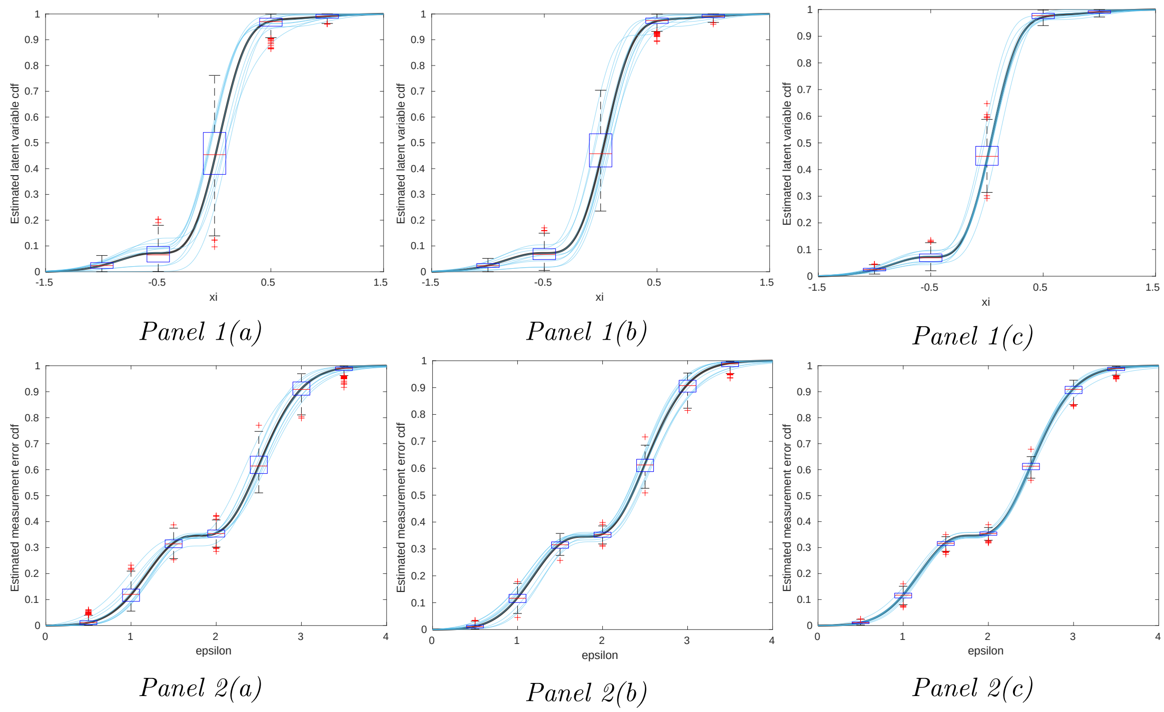


Figure 1: Monte Carlo simulation results ($\kappa = 1$). Panel 1 displays results for F_ξ and Panel 2 for F_ε. Subpanels *(a)*, *(b)*, and *(c)* correspond to sample sizes $N = 1000$, 2000, and 4000, respectively.

5 Conclusion

This paper shows that distributions of the latent variable and measurement errors are identified nonparametrically under mild assumptions when two or more order statistics are recorded from repeated measurements with independent errors, providing a positive answer to the hypothesis in Athey and Haile (2002) for an ascending auction with unobserved heterogeneity. Our results are also applicable to other applications with unobserved heterogeneity when order statistics are observed, survey data on wage offers being a notable example. More examples include repeated experiments with type II censoring, such as in reliability testing, where consecutive low-order failure times are recorded, and estimating the effects and damages of collusion in auctions, a setting in which Asker (2010) emphasizes the importance of accounting for unobserved heterogeneity. Relatedly, the identification result may be applied to extend the framework in Marmer et al. (2017) for testing collusion in ascending auctions.

References

Allen, J., Clark, R., Hickman, B., and Richert, E. (2023). Resolving failed banks: Uncertainty, multiple bidding & auction design. *Review of Economic Studies*, page rdad062.

Aradillas-López, A., Gandhi, A., and Quint, D. (2013). Identification and inference in ascending auctions with correlated private values. *Econometrica*, 81(2):489–534.

Arnold, B. C., Balakrishnan, N., and Nagaraja, H. N. (2008). *A First Course in Order Statistics*. SIAM.

Aryal, G. and Zincenko, F. (2021). Empirical framework for Cournot oligopoly with private information. *arXiv preprint arXiv:2106.15035*.

Asker, J. (2010). A study of the internal organization of a bidding cartel. *American Economic Review*, 100(3):724–62.

Athey, S. (2001). Single crossing properties and the existence of pure strategy equilibria in games of incomplete information. *Econometrica*, 69(4):861–889.

Athey, S. and Haile, P. A. (2002). Identification of standard auction models. *Econometrica*, 70(6):2107–2140.

Athey, S. and Haile, P. A. (2007). Nonparametric approaches to auctions. *Handbook of Econometrics*, 6:3847–3965.

Athey, S., Levin, J., and Seira, E. (2011). Comparing open and sealed bid auctions: Evidence from timber auctions. *The Quarterly Journal of Economics*, 126(1):207–257.

Bierens, H. J. (2008). Semi-nonparametric interval-censored mixed proportional hazard models: Identification and consistency results. *Econometric Theory*, 24(3):749–794.

Bierens, H. J. and Song, H. (2012). Semi-nonparametric estimation of independently and identically repeated first-price auctions via an integrated simulated moments method. *Journal of Econometrics*, 168(1):108–119.

Bonhomme, S. and Robin, J.-M. (2010). Generalized non-parametric deconvolution with an application to earnings dynamics. *The Review of Economic Studies*, 77(2):491–533.

Burdett, K. and Mortensen, D. T. (1998). Wage differentials, employer size, and unemployment. *International Economic Review*, pages 257–273.

Carroll, R. J., Chen, X., and Hu, Y. (2010). Identification and estimation of nonlinear models using two samples with nonclassical measurement errors. *Journal of Nonparametric Statistics*, 22(4):379–399.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. *Handbook of Econometrics*, 6:5549–5632.

David, H. A. and Nagaraja, H. N. (2003). *Order Statistics*. John Wiley & Sons.

Evdokimov, K. and White, H. (2012). Some extensions of a lemma of Kotlarski. *Econometric Theory*, 28(4):925–932.

Flambard, V. and Perrigne, I. (2006). Asymmetry in procurement auctions: Evidence from snow removal contracts. *The Economic Journal*, 116(514):1014–1036.

Freyberger, J. and Larsen, B. J. (2022). Identification in ascending auctions, with an application to digital rights management. *Quantitative Economics*, 13(2):505–543.

Gallant, A. R. (1987). Identification and consistency in semi-nonparametric regression. *Advances in Econometrics: Fifth World Congress*, 1:145–170.

Gallant, A. R. and Nychka, D. W. (1987). Semi-nonparametric maximum likelihood estimation. *Econometrica*, 55(2):363–390.

Grundl, S. and Zhu, Y. (2019). Identification and estimation of risk aversion in first-price auctions with unobserved auction heterogeneity. *Journal of Econometrics*, 210(2):363–378.

Guerre, E., Perrigne, I., and Vuong, Q. (2000). Optimal nonparametric estimation of first-price auctions. *Econometrica*, 68(3):525–574.

Guo, J. (2021). Identification of the wage offer distribution using order statistics. Working Paper 3910119, SSRN.

Haile, P. A. and Tamer, E. (2003). Inference with an incomplete model of english auctions. *Journal of Political Economy*, 111(1):1–51.

Hall, P. and Yao, Q. (2003). Inference in components of variance models with low replication. *The Annals of Statistics*, 31(2):414–441.

Hernández, C., Quint, D., and Turansick, C. (2020). Estimation in English auctions with unobserved heterogeneity. *The RAND Journal of Economics*, 51(3):868–904.

Hörmander, L. (1973). *An Introduction to Complex Analysis in Several Variables*. Elsevier.

Huang, Y. and He, M. (2021). Structural analysis of Tullock contests with an application to us house of representatives elections. *International Economic Review*, 62(3):1011–1054.

Kato, K., Sasaki, Y., and Ura, T. (2021). Robust inference in deconvolution. *Quantitative Economics*, 12(1):109–142.

Kim, K. i. and Lee, J. (2014). Nonparametric estimation and testing of the symmetric IPV framework with unknown number of bidders. Working paper.

Kotlarski, I. (1967). On characterizing the gamma and the normal distribution. *Pacific Journal of Mathematics*, 20(1):69–76.

Krasnokutskaya, E. (2011). Identification and estimation of auction models with unobserved heterogeneity. *The Review of Economic Studies*, 78(1):293–327.

Kruse, R. and Deely, J. (1969). Joint continuity of monotonic functions. *The American Mathematical Monthly*, 76(1):74–76.

Larsen, B. J. (2021). The efficiency of real-world bargaining: Evidence from wholesale used-auto auctions. *The Review of Economic Studies*, 88(2):851–882.

Li, T., Perrigne, I., and Vuong, Q. (2000). Conditionally independent private information in OCS wildcat auctions. *Journal of Econometrics*, 98(1):129–161.

Li, T. and Vuong, Q. (1998). Nonparametric estimation of the measurement error model using multiple indicators. *Journal of Multivariate Analysis*, 65(2):139–165.

Luo, Y., Perrigne, I., and Vuong, Q. (2018). Auctions with ex post uncertainty. *The RAND Journal of Economics*, 49(3):574–593.

Luo, Y., Sang, P., and Xiao, R. (2021). Order statistics approaches to unobserved heterogeneity in auctions. Working Paper 3704644, SSRN.

Luo, Y. and Xiao, R. (2023). Identification of auction models using order statistics. *Journal of Econometrics*, 236(1):105457.

Marmer, V., Shneyerov, A. A., and Kaplan, U. (2017). Identifying collusion in English auctions. Working Paper 2738789, SSRN.

Mbakop, E. (2017). Identification of auctions with incomplete bid data in the presence of unobserved heterogeneity. Working paper, Northwestern University.

Menzel, K. and Morganti, P. (2013). Large sample properties for estimators based on the order statistics approach in auctions. *Quantitative Economics*, 4(2):329–375.

Miller, P. (1970). Characterizing the distributions of three independent n-dimensional random variables, X_1, X_2, X_3, having analytic characteristic functions by the joint distribution of (X_1+X_3, X_2+X_3). *Pacific Journal of Mathematics*, 35(2):487–491.

Rossberg, H. J. (1972). Characterization of the exponential and the Pareto distributions by means of some properties of the distributions which the differences and quotients of order statistics are subject to. *Statistics: A Journal of Theoretical and Applied Statistics*, 3(3):207–216.

Schennach, S. M. (2016). Recent advances in the measurement error literature. *Annual Review of Economics*, 8:341–377.

Stein, E. M. and Shakarchi, R. (2010). *Complex Analysis*, volume 2. Princeton University Press.

Appendix A Proofs

A.1 Supplementary Results

Lemma A.1. *Let $(\varepsilon_{(r)}, \varepsilon_{(s)})$ be r^{th} and s^{th} order statistics from $\varepsilon_1, \ldots, \varepsilon_n$ which are independent with distributions $F_{\varepsilon_1}, \ldots, F_{\varepsilon_n}$. If every F_{ε_j} has a density function f_{ε_j} that is light-tailed (i.e., for some $C_j > 0$, $f_{\varepsilon_j}(\epsilon) = O(e^{-C_j|\epsilon|})$ as $|\epsilon| \to \infty$), then the ch.f. of order statistics $\psi_{\varepsilon_{(r,s)}} : \mathbb{R}^2 \to \mathbb{C}$ is (jointly) analytic.*

Proof of Lemma A.1. Let $C_0 < \min_{1\leq j\leq n} C_j$ be a positive constant and define $\Omega = \{z = (z_r, z_s) \in \mathbb{C}^2 : |\Im z_j| < C_0,\ j \in \{r, s\}\}$, an open set in $\mathbb{C}^2$, where $\Im z_j$ denotes the imaginary part of z_j. Consider

$$\psi^*_{\varepsilon_{(r,s)}} : z \in \Omega \mapsto \int_{\mathbb{R}^2} e^{iz^\top \epsilon}\, dF_{\varepsilon_{(r,s)}}(\epsilon).$$

To show that $\psi_{\varepsilon_{(r,s)}}$ is analytic on $\mathbb{R}^2$, it suffices to show that $\psi^*_{\varepsilon_{(r,s)}}$ is analytic on the open set Ω (since $\mathbb{R}^2 \subset \Omega$). By Hartogs' theorem on separate analyticity (cf. Hörmander (1973), Theorem 2.2.8), it suffices to show that $\psi^*_{\varepsilon_{(r,s)}}$ is separately analytic on a strip $\{z_r \in \mathbb{C} : |\Im z_r| < C_0\}$ for any fixed value of z_s, and vice versa on a strip for z_s for any fixed value of z_r. The remainder of the proof shows that $\psi^*_{\varepsilon_{(r,s)}}$ is (1) indeed well-defined on Ω and (2) separately analytic with respect to each variable.

For any complex vector $z \in \mathbb{C}^d$, let $\Re z \in \mathbb{R}^d$ denote the real part of the vector and $\Im z \in \mathbb{R}^d$ the imaginary part. To show that $\psi^*_{\varepsilon_{(r,s)}}$ is well-defined on Ω, it suffices to show that $\int_{\mathbb{R}^2} e^{-\Im z^\top \epsilon} f_{\varepsilon_{(r,s)}}(\epsilon)\, d\epsilon < \infty$ for any $z \in \Omega$ since

$$\psi^*_{\varepsilon_{(r,s)}}(z) = \int_{\mathbb{R}^2} e^{-\Im z^\top \epsilon} e^{i\Re z^\top \epsilon} f_{\varepsilon_{(r,s)}}(\epsilon)\, d\epsilon.$$

From the definition of the joint density $f_{\varepsilon_{(r,s)}}$ (see, e.g., (5.2.8) in David and Nagaraja (2003)), there exists a positive constant $K_{n,r,s}$ such that for every $\epsilon = (\epsilon_r, \epsilon_s) \in \mathbb{R}^2$,

$$f_{\varepsilon_{(r,s)}}(\epsilon) \leq K_{n,r,s} \sum_{1\leq k,\ell\leq n} f_{\varepsilon_k}(\epsilon_r) f_{\varepsilon_\ell}(\epsilon_s).$$

where by assumption, $f_{\varepsilon_k}(\epsilon) = O(e^{-C_k|\epsilon|})$ as $|\epsilon| \to \infty$ for every k. Hence, it follows from the fact $|\Im z_r| < C_0$ and $|\Im z_s| < C_0$ in Ω that

$$\int_{\mathbb{R}^2} e^{-\Im z^\top \epsilon} f_{\varepsilon_{(r,s)}}(\epsilon)\, d\epsilon \le K_{n,r,s} \sum_{1 \le k,\ell \le n} \int_{\mathbb{R}^2} e^{-\Im z_r \epsilon_r} f_{\varepsilon_k}(\epsilon_r) e^{-\Im z_s \epsilon_s} f_{\varepsilon_\ell}(\epsilon_s)\, d\epsilon < \infty,$$

which concludes that $\psi^*_{\varepsilon_{(r,s)}}$ is well-defined on Ω.

Now we show that $\psi^*_{\varepsilon_{(r,s)}}(\cdot, z_s)$ is analytic on $\Omega_r = \{z_r \in \mathbb{C} : |\Im z_r| < C_0\}$ for every z_s in the domain. By Theorem 5.2 in Stein and Shakarchi (2010), it suffices to construct a sequence of analytic functions $\{\psi^*_m\}_m$ that converges uniformly to $\psi^*_{\varepsilon_{(r,s)}}(\cdot, z_s)$ in every compact subset of Ω_r. We show that the sequence

$$\psi^*_m(z_r) = \int_{[-m,m]^2} e^{i(z_r \epsilon_r + z_s \epsilon_s)}\, dF_{\varepsilon_{(r,s)}}(\epsilon)$$

satisfies the criteria above. Since the region of integration is bounded, it follows from the dominated convergence theorem that for every m,

$$\psi^*_m(z_r) = \int_{[-m,m]^2} \sum_{\ell=0}^{\infty} \frac{(iz^\top \epsilon)^\ell}{\ell!}\, dF_{\varepsilon_{(r,s)}}(\epsilon) = \sum_{\ell=0}^{\infty} \int_{[-m,m]^2} \frac{(iz^\top \epsilon)^\ell}{\ell!}\, dF_{\varepsilon_{(r,s)}}(\epsilon)$$

is entire on the complex plane and thus analytic on $\Omega_r \subset \mathbb{C}$. Further,

$$\begin{aligned}
\left|\psi^*_{\varepsilon_{(r,s)}}(z_r, z_s) - \psi^*_m(z_r)\right| &\le \int_{\mathbb{R}^2 \setminus [-m,m]^2} \left|e^{iz^\top \epsilon}\right| dF_{\varepsilon_{(r,s)}}(\epsilon) \\
&= \int_{\mathbb{R}^2 \setminus [-m,m]^2} e^{-\Im z^\top \epsilon}\, dF_{\varepsilon_{(r,s)}}(\epsilon) \\
&\le K_{n,r,s} \sum_{1 \le k,\ell \le n} \int_{\mathbb{R}^2 \setminus [-m,m]^2} e^{-\Im z^\top \epsilon} f_{\varepsilon_k}(\epsilon_r) f_{\varepsilon_\ell}(\epsilon_s)\, d\epsilon \\
&\le K_{n,r,s} \sum_{1 \le k,\ell \le n} \int_{\mathbb{R} \times (\mathbb{R} \setminus [-m,m])} e^{-\Im z^\top \epsilon} f_{\varepsilon_k}(\epsilon_r) f_{\varepsilon_\ell}(\epsilon_s)\, d\epsilon \\
&\quad + K_{n,r,s} \sum_{1 \le k,\ell \le n} \int_{(\mathbb{R} \setminus [-m,m]) \times \mathbb{R}} e^{-\Im z^\top \epsilon} f_{\varepsilon_k}(\epsilon_r) f_{\varepsilon_\ell}(\epsilon_s)\, d\epsilon.
\end{aligned}$$

Since $f_{\varepsilon_k}(\epsilon_r) = O(e^{-C_k|\epsilon_r|})$ for every k by assumption and $|\Im z_r| < C_0$, for any compact subset $D \subset \Omega_r$, we have

$$\sup_{z_r \in D} K_k(z_r) \coloneqq \sup_{z_r \in D} \int_{\mathbb{R}} e^{-\Im z_r \epsilon_r} f_{\varepsilon_k}(\epsilon_r)\, d\epsilon_r \le \int_{\mathbb{R}} e^{\sup_{z_r \in D} |\Im z_r| \epsilon_r} f_{\varepsilon_k}(\epsilon_r)\, d\epsilon_r < \infty.$$

In addition, for m sufficiently large (independent of z_r), there exists a constant $K_{\ell,0}$ such that

$$\begin{aligned}
&\int_{\mathbb{R}\times(\mathbb{R}\setminus[-m,m])} e^{-\Im z^\top \epsilon} f_{\varepsilon_k}(\epsilon_r) f_{\varepsilon_\ell}(\epsilon_s)\, d\epsilon \\
&\quad \le K_k(z_r) K_0 \int_{\mathbb{R}\setminus[-m,m]} e^{-C_0|\epsilon_s| - \Im z_s \epsilon_s}\, d\epsilon_s. \\
&\quad = K_k(z_r) K_0 \left(\int_{-\infty}^{-m} e^{-(C_0 - \Im z_s)|\epsilon_s|}\, d\epsilon_s + \int_m^\infty e^{-(C_0+\Im z_s)|\epsilon_s|}\, d\epsilon_s \right) \longrightarrow 0,
\end{aligned}$$

as $m \to \infty$. Similarly, we have

$$\int_{(\mathbb{R}\setminus[-m,m])\times\mathbb{R}} e^{-\Im z^\top \epsilon} f_{\varepsilon_k}(\epsilon_r) f_{\varepsilon_\ell}(\epsilon_s)\, d\epsilon \le K_\ell(z_r) K_{k,0} \int_{\mathbb{R}\setminus[-m,m]} e^{-C_0|\epsilon_r| - \Im z_r \epsilon_r}\, d\epsilon_r,$$

where, as $m \to \infty$,

$$\sup_{z_r \in D} \int_{\mathbb{R}\setminus[-m,m]} e^{-C_0|\epsilon_r| - \Im z_r \epsilon_r}\, d\epsilon_r \longrightarrow 0.$$

Therefore, we conclude that $\{\psi_m^*\}_m$ converges uniformly to $\psi^*_{\varepsilon_{(r,s)}}(\cdot, z_s)$ on any compact subset of Ω_r for any fixed z_s in the domain. Therefore, $\psi^*_{\varepsilon_{(r,s)}}(\cdot, z_s)$ is analytic on Ω_r. An analogous proof shows that $\psi^*_{\varepsilon_{(r,s)}}(z_r, \cdot)$ is analytic on $\Omega_s = \{z_s \in \mathbb{C} : |\Im z_s| < C_0\}$ for any fixed z_r in the domain.

This concludes the proof that $\psi_{\varepsilon_{(r,s)}}$ is (jointly) analytic on $\Omega = \{z = (z_r, z_s) \in \mathbb{C}^2 : z_r \in \Omega_r, z_s \in \Omega_s\}$. ☐

Lemma A.2. *Suppose the measurement errors satisfy Assumption 3.2(a) and 3.3. Define, for every $\epsilon_s \in \mathbb{R}$ and $\epsilon_r > 0$,*

$$F_{\varepsilon_{(s|r)}}(\epsilon_s; \epsilon_r) = \frac{F_{\varepsilon_{(r,s)}}(\epsilon_r, \epsilon_s)}{F_{\varepsilon_{(r)}}(\epsilon_r)}.$$

Then $\lim_{\epsilon_r \downarrow 0} F_{\varepsilon_{(s|r)}}(\cdot; \epsilon_r) = F_{\varepsilon_{s-r:n-r}}(\cdot)$, where the latter denotes the distribution of the $(s-r)^{th}$ order statistic of a random sample of size $n-r$ from F_ε.

Proof. When $\epsilon_s \le 0$, clearly $\lim_{\epsilon_r \downarrow 0} F_{\varepsilon_{(s|r)}}(\epsilon_s; \epsilon_r) = F_{\varepsilon_{s-r:n-r}}(\epsilon_s) = 0$ by Assumption 3.3. Thus, fix any $\epsilon_s > 0$ and consider small enough ϵ_r such that $\epsilon_s > \epsilon_r \downarrow 0$.

It follows from standard results (e.g., see (2.1.3) and (2.2.4) in David and Nagaraja

(2003)) that the joint and marginal distribution functions may be expressed as

$$F_{\varepsilon_{(s|r)}}(\epsilon_s;\epsilon_r)=\frac{\sum_{k=s}^{n}\sum_{j=r}^{k}C^n_{j,k-j,n-k}F_\varepsilon(\epsilon_r)^j(F_\varepsilon(\epsilon_s)-F_\varepsilon(\epsilon_r))^{k-j}(1-F_\varepsilon(\epsilon_s))^{n-k}}{\sum_{\ell=r}^{n}C^n_\ell F_\varepsilon(\epsilon_r)^\ell(1-F_\varepsilon(\epsilon_r))^{n-\ell}},$$

where $C^n_{j,k-j,n-k}$ and C^n_ℓ are multinomial and binomial coefficients, respectively. Observe that

$$\begin{aligned}
F_{\varepsilon_{(s|r)}}(\epsilon_s;\epsilon_r)&=\frac{\sum_{j=r}^{n}\sum_{k=\max(s,j)}^{n}C^n_{j,k-j,n-k}F_\varepsilon(\epsilon_r)^j(F_\varepsilon(\epsilon_s)-F_\varepsilon(\epsilon_r))^{k-j}(1-F_\varepsilon(\epsilon_s))^{n-k}}{\sum_{\ell=r}^{n}C^n_\ell F_\varepsilon(\epsilon_r)^\ell(1-F_\varepsilon(\epsilon_r))^{n-\ell}}\\
&=\sum_{j=r}^{n}\frac{C^n_j F_\varepsilon(\epsilon_r)^j(1-F_\varepsilon(\epsilon_r))^{n-j}}{\sum_{\ell=r}^{n}C^n_\ell F_\varepsilon(\epsilon_r)^\ell(1-F_\varepsilon(\epsilon_r))^{n-\ell}}\times\\
&\qquad\sum_{k=\max(s,j)}^{n}C^{n-j}_{k-j}\left(\frac{F_\varepsilon(\epsilon_s)-F_\varepsilon(\epsilon_r)}{1-F_\varepsilon(\epsilon_r)}\right)^{k-j}\left(\frac{1-F_\varepsilon(\epsilon_s)}{1-F_\varepsilon(\epsilon_r)}\right)^{n-k}\\
&=\sum_{j=r}^{n}\frac{C^n_j F_\varepsilon(\epsilon_r)^j(1-F_\varepsilon(\epsilon_r))^{n-j}}{\sum_{\ell=r}^{n}C^n_\ell F_\varepsilon(\epsilon_r)^\ell(1-F_\varepsilon(\epsilon_r))^{n-\ell}}\times\\
&\qquad\sum_{k=\max(s-j,0)}^{n-j}C^{n-j}_{k}\left(\frac{F_\varepsilon(\epsilon_s)-F_\varepsilon(\epsilon_r)}{1-F_\varepsilon(\epsilon_r)}\right)^{k}\left(\frac{1-F_\varepsilon(\epsilon_s)}{1-F_\varepsilon(\epsilon_r)}\right)^{n-j-k}.
\end{aligned}$$

As $\epsilon_r\downarrow 0$, the weight component vanishes for all but $j=r$, i.e.,

$$\begin{aligned}
\frac{C^n_j F_\varepsilon(\epsilon_r)^j(1-F_\varepsilon(\epsilon_r))^{n-j}}{\sum_{\ell=r}^{n}C^n_\ell F_\varepsilon(\epsilon_r)^\ell(1-F_\varepsilon(\epsilon_r))^{n-\ell}}&=\left(\sum_{\ell=r}^{n}\frac{C^n_\ell}{C^n_j}\left(\frac{F_\varepsilon(\epsilon_r)}{1-F_\varepsilon(\epsilon_r)}\right)^{\ell-j}\right)^{-1}\\
&=\left(\sum_{\ell=r-j}^{n-j}\frac{C^n_{\ell+j}}{C^n_j}\left(\frac{F_\varepsilon(\epsilon_r)}{1-F_\varepsilon(\epsilon_r)}\right)^{\ell}\right)^{-1}\longrightarrow\begin{cases}1 & \text{if } j=r,\\ 0 & \text{if } j\neq r.\end{cases}
\end{aligned}$$

This implies that for any $\epsilon_s>0$,

$$\lim_{\epsilon_r\downarrow 0}F_{\varepsilon_{(s|r)}}(\epsilon_s;\epsilon_r)=\sum_{k=s-r}^{n-r}C^{n-r}_k F_\varepsilon(\epsilon_s)^k(1-F_\varepsilon(\epsilon_s))^{n-r-k}=F_{\varepsilon_{s-r:n-r}}(\epsilon_s).$$

Therefore, we conclude that $\lim_{\epsilon_r\downarrow 0}F_{\varepsilon_{(s|r)}}(\cdot;\epsilon_r)=F_{\varepsilon_{s-r:n-r}}(\cdot)$ on $\mathbb{R}$. □

Remark A.1. *The conditional distribution $F_{\varepsilon_{(s|r)}}(\epsilon_s;\epsilon_r)$ is well-defined for any $\epsilon_r>0$ since $F_{\varepsilon_{(r)}}(\epsilon_r)>0$ under the support normalization in Assumption 3.3, but $F_{\varepsilon_{(r)}}(0)=0$. Lemma A.2 allows one to define $F_{\varepsilon_{(s|r)}}(\cdot;0)$ by a continuous extension from above.*

Suppose there exists a known group structure for the measurement errors as in Assumption 3.5 and let n_q denote the size of group q. Without loss of generality, let $(\varepsilon_1, \ldots, \varepsilon_n)$ be ordered such that the first n_1 random variables are from group 1, next n_2 random variables from group 2, etc. Further without loss of generality, let g_1 be a group with at least $(n-r)$ members, i.e., $n_1 \geq n-r$. Let $R_{(j)} = \{q : X_{(j)} = X_k \text{ for some } k \in g_q\}$ be the group identity of the j^{th} order statistic. We abuse notation and use $R_{(j)}$ to denote both the set and the a.s. unique element in $\{1, \ldots, p\}$. Further, let E_q denote the event where the top $(n-r)$ order statistics are all from group 1 except for the s^{th} order statistic, which belongs to group q (where it may be that $q = 1$), i.e.,

$$E_q = \{R_{(r+1)} = 1, \ldots, R_{(s-1)} = 1, R_{(s)} = q, R_{(s+1)} = 1, \ldots, R_{(n)} = 1\}.$$

In the following two lemmas, we derive the distribution of order statistics conditional on the event E_1 and E_q ($q \neq 1$), respectively.

Lemma A.3. *Let $(\varepsilon_{(r)}, \varepsilon_{(s)})$ be order statistics from an independent but nonidentically distributed sample $(\varepsilon_1, \ldots, \varepsilon_n) \sim \times_{j=1}^n F_{\varepsilon_j}$. Let ζ be the maximum of $\{\varepsilon_{n-r+1}, \ldots, \varepsilon_n\}$. Then*

$$\begin{aligned}
&\mathbb{P}(\varepsilon_{(r)} \leq \epsilon_r, \varepsilon_{(s)} \leq \epsilon_s \,|\, E_1) \\
&\quad \propto \sum_{k=s}^{n} \sum_{j=r}^{k} C^{n_1}_{j-r,k-j,n-k} \mathbb{E}_\zeta \Big(1\{\zeta \leq \epsilon_r\} (F_{\varepsilon_1}(\epsilon_r) - F_{\varepsilon_1}(\zeta))^{j-r} \Big) \times \\
&\qquad\qquad (F_{\varepsilon_1}(\epsilon_s) - F_{\varepsilon_1}(\epsilon_r))^{k-j} (1 - F_{\varepsilon_1}(\epsilon_s))^{n-k},
\end{aligned}$$

and

$$\mathbb{P}(\varepsilon_{(r)} \leq \epsilon_r \,|\, E_1) \propto \sum_{j=r}^{n} C^{n_1}_{j-r,n-j} \mathbb{E}_\zeta \Big(1\{\zeta \leq \epsilon_r\} (F_{\varepsilon_1}(\epsilon_r) - F_{\varepsilon_1}(\zeta))^{j-r} \Big) (1 - F_{\varepsilon_1}(\epsilon_r))^{n-j}.$$

Proof. Note that

$$\{\varepsilon_{(r)} \leq \epsilon_r, \varepsilon_{(s)} \leq \epsilon_s, E_1\} = \bigcup_{k=s}^{n} \bigcup_{j=r}^{k} \{\varepsilon_{(j)} \leq \epsilon_r, \varepsilon_{(j+1)} > \epsilon_r, \varepsilon_{(k)} \leq \epsilon_s, \varepsilon_{(k+1)} > \epsilon_s, E_1\}.$$

Let $\mathbb{S}_{j,k}$ be a collection of reordered vectors of $(1, \ldots, n)$ where $\sigma = (\sigma_1, \ldots, \sigma_n) \in \mathbb{S}_{j,k}$

uniquely partitions $\{1,\ldots,n\}$ in the sense that

$$\{1,\ldots,n\}=\{\sigma_1,\ldots,\sigma_r\}\cup\{\sigma_{r+1},\ldots,\sigma_j\}\cup\{\sigma_{j+1},\ldots,\sigma_k\}\cup\{\sigma_{k+1},\ldots,\sigma_n\},$$

and $\sigma_\ell\in g_1$ for all $\ell\geq r+1$. In words, σ divides group 1 into four, where three of them consist solely of group 1, and the first r coordinates of σ collect remaining members of group 1, if any, along with all other groups. We will use this partition to assign group-1 measurement errors in a way that (1) all top $(n-r)$ order statistics belong to group 1, and (2) the three sets of these order statistics divide the group-1 measurement errors into the ranges $(-\infty,\epsilon_r]$, $(\epsilon_r,\epsilon_s]$, and (ϵ_s,∞), respectively. The remaining members of group 1 and all other groups are then associated with the lowest r order statistics. More precisely, we have

$$\begin{aligned}&\{\varepsilon_{(r)}\leq\epsilon_r,\varepsilon_{(s)}\leq\epsilon_s,E_1\}\\&\quad=\bigcup_{k=s}^{n}\bigcup_{j=r}^{k}\bigcup_{\sigma\in\mathbb{S}_{j,k}}\{\varepsilon_{\sigma_{(r)}}\leq\epsilon_r,\varepsilon_{\sigma_{(r)}}<\varepsilon_{r+1}^{j}\leq\epsilon_r,\epsilon_r<\varepsilon_{j+1}^{k}\leq\epsilon_s,\epsilon_s<\varepsilon_{k+1}^{n}\},\end{aligned}$$

where $\varepsilon_{\sigma_{(r)}}=\max_{1\leq j\leq r}\varepsilon_{\sigma_j}$ and ε_ℓ^m is a shorthand for $\varepsilon_{\sigma_\ell},\ldots,\varepsilon_{\sigma_m}$. Denote by $\mathbb{E}_{\sigma_{(r)}}$ the expectation with respect to $\varepsilon_{\sigma_{(r)}}$. It follows that

$$\begin{aligned}&\mathbb{P}(\varepsilon_{(r)}\leq\epsilon_r,\varepsilon_{(s)}\leq\epsilon_s,E_1)\\&\quad=\sum_{k=s}^{n}\sum_{j=r}^{k}\sum_{\sigma\in\mathbb{S}_{j,k}}\mathbb{E}_{\sigma_{(r)}}\mathbb{P}(\varepsilon_{\sigma_{(r)}}\leq\epsilon_r,\varepsilon_{\sigma_{(r)}}<\varepsilon_{r+1}^{j}\leq\epsilon_r,\epsilon_r<\varepsilon_{j+1}^{k}\leq\epsilon_s,\epsilon_s<\varepsilon_{k+1}^{n}\,|\,\varepsilon_{\sigma_{(r)}})\\&\quad=\sum_{k=s}^{n}\sum_{j=r}^{k}\sum_{\sigma\in\mathbb{S}_{j,k}}\mathbb{E}_{\sigma_{(r)}}\Big(1\{\varepsilon_{\sigma_{(r)}}\leq\epsilon_r\}(F_{\varepsilon_1}(\epsilon_r)-F_{\varepsilon_1}(\varepsilon_{\sigma_{(r)}}))^{j-r}\times\\&\qquad\qquad(F_{\varepsilon_1}(\epsilon_s)-F_{\varepsilon_1}(\epsilon_r))^{k-j}(1-F_{\varepsilon_1}(\epsilon_s))^{n-k}\Big)\\&\quad=\sum_{k=s}^{n}\sum_{j=r}^{k}C^{n_1}_{j-r,k-j,n-k}\mathbb{E}_{\sigma_{(r)}}\Big(1\{\varepsilon_{\sigma_{(r)}}\leq\epsilon_r\}(F_{\varepsilon_1}(\epsilon_r)-F_{\varepsilon_1}(\varepsilon_{\sigma_{(r)}}))^{j-r}\Big)\times\\&\qquad\qquad(F_{\varepsilon_1}(\epsilon_s)-F_{\varepsilon_1}(\epsilon_r))^{k-j}(1-F_{\varepsilon_1}(\epsilon_s))^{n-k},\end{aligned}$$

where $C^{n_1}_{j-r,k-j,n-k}=n_1!/[(n_1-(n-r))!(j-r)!(k-j)!(n-k)!]$ is the multinomial coefficient equal to the size of $\mathbb{S}_{j,k}$ (the number of ways to classify members of group 1 to four sets that partition $\{1,\ldots,n\}$). The last equality holds because the distribution of $\varepsilon_{\sigma_{(r)}}$ is invariant with respect to σ. This proves the original statement for the joint distribution in the lemma since $\zeta =_d \varepsilon_{\sigma_{(r)}}$. This is because $\{\varepsilon_{n-r+1}, \ldots, \varepsilon_n\}$ consists of all measurement errors except $(n-r)$ members from group 1, as is the definition of $\varepsilon_{\sigma_{(r)}}$ for any $\sigma \in \mathbb{S}_{j,k}$.

An analogous derivation shows that

$$\mathbb{P}(\varepsilon_{(r)} \leq \epsilon_r, E_1) = \sum_{j=r}^{n} C^{m_1}_{j-r,n-j} \mathbb{E}_{\sigma_{(r)}} \Big(1\{\varepsilon_{\sigma_{(r)}} \leq \epsilon_r\} (F_{\varepsilon_1}(\epsilon_r) - F_{\varepsilon_1}(\varepsilon_{\sigma_{(r)}}))^{j-r} \Big) \times (1 - F_{\varepsilon_1}(\epsilon_r))^{n-j}.$$

□

Lemma A.4. *Let $(\varepsilon_{(r)}, \varepsilon_{(s)})$ be order statistics from an independent but nonidentically distributed sample $(\varepsilon_1, \ldots, \varepsilon_n) \sim \times_{j=1}^n F_{\varepsilon_j}$. If $q \neq 1$, for $m \in g_q$ and $\zeta = (\zeta_r, \zeta_s)$ where ζ_r is the maximum of $\{\varepsilon_{n-r}, \ldots, \varepsilon_n\} \backslash \{\varepsilon_m\}$ and $\zeta_s = \varepsilon_m$,*

$$\begin{aligned}
&\mathbb{P}(\varepsilon_{(r)} \leq \epsilon_r, \varepsilon_{(s)} \leq \epsilon_s \,|\, E_q) \\
&\quad \propto \sum_{k=s}^{n} \sum_{j=r}^{s-1} n_q C^{m_1}_{j-r,s-j-1,k-s,n-k} \mathbb{E}_\zeta \Big(1\{\zeta_r \leq \epsilon_r < \zeta_s \leq \epsilon_s\} (F_{\varepsilon_1}(\epsilon_r) - F_{\varepsilon_1}(\zeta_r))^{j-r} \times \\
&\qquad\qquad (F_{\varepsilon_1}(\zeta_s) - F_{\varepsilon_1}(\epsilon_r))^{s-j-1} (F_{\varepsilon_1}(\epsilon_s) - F_{\varepsilon_1}(\zeta_s))^{k-s} \Big) (1 - F_{\varepsilon_1}(\epsilon_s))^{n-k} \\
&\qquad + \sum_{k=s}^{n} \sum_{j=s}^{k} n_q C^{m_1}_{s-r-1,j-s,k-j,n-k} \mathbb{E}_\zeta \Big(1\{\zeta_r \leq \zeta_s \leq \epsilon_r\} (F_{\varepsilon_1}(\zeta_s) - F_{\varepsilon_1}(\zeta_r))^{s-r-1} \times \\
&\qquad\qquad (F_{\varepsilon_1}(\epsilon_r) - F_{\varepsilon_1}(\zeta_s))^{j-s} \Big) (F_{\varepsilon_1}(\epsilon_s) - F_{\varepsilon_1}(\epsilon_r))^{k-j} (1 - F_{\varepsilon_1}(\epsilon_s))^{n-k},
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{P}(\varepsilon_{(r)} \leq \epsilon_r \,|\, E_q) \\
&\quad \propto \sum_{j=r}^{s-1} n_q C^{m_1}_{j-r,s-j-1,n-s} \mathbb{E}_\zeta \Big(1\{\zeta_r \leq \epsilon_r < \zeta_s\} (F_{\varepsilon_1}(\epsilon_r) - F_{\varepsilon_1}(\zeta_r))^{j-r} \\
&\qquad\qquad (F_{\varepsilon_1}(\zeta_s) - F_{\varepsilon_1}(\epsilon_r))^{s-j-1} (1 - F_{\varepsilon_1}(\zeta_s))^{n-s} \Big) \\
&\qquad + \sum_{j=s}^{n} n_q C^{m_1}_{s-r-1,j-s,n-j} \mathbb{E}_\zeta \Big(1\{\zeta_r \leq \zeta_s \leq \epsilon_r\} (F_{\varepsilon_1}(\zeta_s) - F_{\varepsilon_1}(\zeta_r))^{s-r-1} \\
&\qquad\qquad (F_{\varepsilon_1}(\epsilon_r) - F_{\varepsilon_1}(\zeta_s))^{j-s} \Big) (1 - F_{\varepsilon_1}(\epsilon_r))^{n-j}.
\end{aligned}$$

Proof. Note that

$$
\begin{aligned}
\{\varepsilon_{(r)} \leq \epsilon_r, \varepsilon_{(s)} \leq \epsilon_s, E_q\} &= \bigcup_{k=s}^{n} \bigcup_{j=r}^{k} \{\varepsilon_{(j)} \leq \epsilon_r, \varepsilon_{(j+1)} > \epsilon_r, \varepsilon_{(k)} \leq \epsilon_s, \varepsilon_{(k+1)} > \epsilon_s, E_q\} \\
&= \left(\bigcup_{k=s}^{n} \bigcup_{j=r}^{s-1} \{\varepsilon_{(j)} \leq \epsilon_r, \varepsilon_{(j+1)} > \epsilon_r, \varepsilon_{(k)} \leq \epsilon_s, \varepsilon_{(k+1)} > \epsilon_s, E_q\} \right) \\
&\quad \bigcup \left(\bigcup_{k=s}^{n} \bigcup_{j=s}^{k} \{\varepsilon_{(j)} \leq \epsilon_r, \varepsilon_{(j+1)} > \epsilon_r, \varepsilon_{(k)} \leq \epsilon_s, \varepsilon_{(k+1)} > \epsilon_s, E_q\} \right).
\end{aligned}
$$

The set is partitioned into two parts where $\varepsilon_{(s)} > \epsilon_r$ (in the first part of the partition) and where $\varepsilon_{(s)} \leq \epsilon_r$ (in the second part of the partition). We partition the event into two different events because, as will be evident in the derivations below, the functional form for the probability of these events differs depending on the location of $\varepsilon_{(s)}$ relative to ϵ_r.

For $j < s$, let $\mathbb{S}^1_{j,k}$ be a collection of reordered vectors of $(1, \ldots, n)$ where each vector $\sigma \in \mathbb{S}^1_{j,k}$ uniquely partition $\{1, \ldots, n\}$ in the sense that

$$
\{1, \ldots, n\} = \{\sigma_\ell\}_{\ell=1}^{r} \cup \{\sigma_\ell\}_{\ell=r+1}^{j} \cup \{\sigma_\ell\}_{\ell=j+1}^{s-1} \cup \{\sigma_s\} \cup \{\sigma_\ell\}_{\ell=s+1}^{k} \cup \{\sigma_\ell\}_{\ell=k+1}^{n},
$$

where $\sigma_s \in g_q$ and $\sigma_\ell \in g_1$ for all $\ell \geq r+1$ such that $\ell \neq s$. We use $\sigma \in \mathbb{S}^1_{j,k}$ to divide group 1 such that all top $(n-r)$ order statistics except for the s^{th} belong to group 1 and are in the ranges $(-\infty, \epsilon_r]$, $(\epsilon_r, \varepsilon_{\sigma_s}]$ $(\varepsilon_{\sigma_s}, \epsilon_s]$, and (ϵ_s, ∞), respectively; the s^{th} order statistic ε_{σ_s} is in $(\epsilon_r, \epsilon_s]$ (because $j < s \leq k$); and the remaining members are associated with the lowest r order statistics.

Similarly, for $j \geq s$, let $\mathbb{S}^2_{j,k}$ be a collection of reordered vectors of $(1, \ldots, n)$ where each vector $\sigma \in \mathbb{S}^2_{j,k}$ uniquely partition $\{1, \ldots, n\}$ in the sense that

$$
\{1, \ldots, n\} = \{\sigma_\ell\}_{\ell=1}^{r} \cup \{\sigma_\ell\}_{\ell=r+1}^{s-1} \cup \{\sigma_s\} \cup \{\sigma_\ell\}_{\ell=s+1}^{j} \cup \{\sigma_\ell\}_{\ell=j+1}^{k} \cup \{\sigma_\ell\}_{\ell=k+1}^{n},
$$

where $\sigma_s \in g_q$ and $\sigma_\ell \in g_1$ for all $\ell \geq r+1$ such that $\ell \neq s$. We use $\sigma \in \mathbb{S}^2_{j,k}$ to divide group 1 such that all top $(n-r)$ order statistics except for the s^{th} belong to group 1 and are in the ranges $(-\infty, \varepsilon_{\sigma_s}]$, $(\varepsilon_{\sigma_s}, \epsilon_r]$ $(\epsilon_r, \epsilon_s]$, and (ϵ_s, ∞), respectively; the s^{th} order statistic ε_{σ_s} is in $(-\infty, \epsilon_r]$ (because $s \leq j$); and the remaining members are associated with the lowest r order statistics.

It follows that

$$
\begin{aligned}
&\{\varepsilon_{(r)} \le \epsilon_r, \varepsilon_{(s)} \le \epsilon_s, E_q\} \\
&\quad = \Bigg(\bigcup_{k=s}^{n} \bigcup_{j=r}^{s-1} \bigcup_{\sigma \in \mathbb{S}^1_{j,k}} \{\varepsilon_{\sigma_{(r)}} \le \epsilon_r, \varepsilon_{\sigma_{(r)}} < \varepsilon^{j}_{r+1} \le \epsilon_r, \epsilon_r < \varepsilon^{s-1}_{j+1} \le \varepsilon_{\sigma_s}, \\
&\qquad\qquad \epsilon_r < \varepsilon_{\sigma_s} \le \epsilon_s, \varepsilon_{\sigma_s} \le \varepsilon^{k}_{s+1} \le \epsilon_s, \epsilon_s < \varepsilon^{n}_{k+1}\} \Bigg) \\
&\qquad \bigcup \Bigg(\bigcup_{k=s}^{n} \bigcup_{j=s}^{k} \bigcup_{\sigma \in \mathbb{S}^2_{j,k}} \{\varepsilon_{\sigma_{(r)}} \le \varepsilon_{\sigma_s}, \varepsilon_{\sigma_{(r)}} < \varepsilon^{s-1}_{r+1} \le \varepsilon_{\sigma_s}, \varepsilon_{\sigma_s} \le \epsilon_r, \\
&\qquad\qquad \varepsilon_{\sigma_s} < \varepsilon^{j}_{s+1} \le \epsilon_r, \epsilon_r < \varepsilon^{k}_{j+1} \le \epsilon_s, \epsilon_s < \varepsilon^{n}_{k+1}\} \Bigg),
\end{aligned}
$$

where $\varepsilon_{\sigma_{(r)}} = \max_{1 \le j \le r} \varepsilon_{\sigma_j}$ and ε^m_ℓ is a shorthand for $\varepsilon_{\sigma_\ell}, \ldots, \varepsilon_{\sigma_m}$. Denote by $\mathbb{E}_{\sigma(r),s}$ the expectation with respect to $(\varepsilon_{\sigma_{(r)}}, \varepsilon_{\sigma_s})$. Then, we have

$$
\begin{aligned}
&\mathbb{P}(\varepsilon_{(r)} \le \epsilon_r, \varepsilon_{(s)} \le \epsilon_s, E_q) \\
&\quad = \sum_{k=s}^{n} \sum_{j=r}^{s-1} \sum_{\sigma \in \mathbb{S}^1_{j,k}} \mathbb{E}_{\sigma(r),s} \Big(1\{\varepsilon_{\sigma_{(r)}} \le \epsilon_r < \varepsilon_{\sigma_s} \le \epsilon_s\} (F_{\varepsilon_1}(\epsilon_r) - F_{\varepsilon_1}(\varepsilon_{\sigma_{(r)}}))^{j-r} \times \\
&\qquad (F_{\varepsilon_1}(\varepsilon_{\sigma_s}) - F_{\varepsilon_1}(\epsilon_r))^{s-j-1} (F_{\varepsilon_1}(\epsilon_s) - F_{\varepsilon_1}(\varepsilon_{\sigma_s}))^{k-s} (1 - F_{\varepsilon_1}(\epsilon_s))^{n-k} \Big) \\
&\qquad + \sum_{k=s}^{n} \sum_{j=s}^{k} \sum_{\sigma \in \mathbb{S}^2_{j,k}} \mathbb{E}_{\sigma(r),s} \Big(1\{\varepsilon_{\sigma_{(r)}} \le \varepsilon_{\sigma_s} \le \epsilon_r\} (F_{\varepsilon_1}(\varepsilon_{\sigma_s}) - F_{\varepsilon_1}(\varepsilon_{\sigma_{(r)}}))^{s-r-1} \times \\
&\qquad (F_{\varepsilon_1}(\epsilon_r) - F_{\varepsilon_1}(\varepsilon_{\sigma_s}))^{j-s} (F_{\varepsilon_1}(\epsilon_s) - F_{\varepsilon_1}(\epsilon_r))^{k-j} (1 - F_{\varepsilon_1}(\epsilon_s))^{n-k} \Big) \\
&\quad = \sum_{k=s}^{n} \sum_{j=r}^{s-1} n_q C^{n_1}_{j-r, s-j-1, k-s, n-k} \mathbb{E}_{\sigma(r),s} \Big(1\{\varepsilon_{\sigma_{(r)}} \le \epsilon_r < \varepsilon_{\sigma_s} \le \epsilon_s\} (F_{\varepsilon_1}(\epsilon_r) - F_{\varepsilon_1}(\varepsilon_{\sigma_{(r)}}))^{j-r} \times \\
&\qquad (F_{\varepsilon_1}(\varepsilon_{\sigma_s}) - F_{\varepsilon_1}(\epsilon_r))^{s-j-1} (F_{\varepsilon_1}(\epsilon_s) - F_{\varepsilon_1}(\varepsilon_{\sigma_s}))^{k-s} \Big) (1 - F_{\varepsilon_1}(\epsilon_s))^{n-k} \\
&\qquad + \sum_{k=s}^{n} \sum_{j=s}^{k} n_q C^{n_1}_{s-r-1, j-s, k-j, n-k} \mathbb{E}_{\sigma(r),s} \Big(1\{\varepsilon_{\sigma_{(r)}} \le \varepsilon_{\sigma_s} \le \epsilon_r\} (F_{\varepsilon_1}(\varepsilon_{\sigma_s}) - F_{\varepsilon_1}(\varepsilon_{\sigma_{(r)}}))^{s-r-1} \times \\
&\qquad (F_{\varepsilon_1}(\epsilon_r) - F_{\varepsilon_1}(\varepsilon_{\sigma_s}))^{j-s} \Big) (F_{\varepsilon_1}(\epsilon_s) - F_{\varepsilon_1}(\epsilon_r))^{k-j} (1 - F_{\varepsilon_1}(\epsilon_s))^{n-k}.
\end{aligned}
$$

The last equality holds because the distribution of $(\varepsilon_{\sigma_{(r)}}, \varepsilon_{\sigma_s})$ is invariant with respect to σ. This proves the original statement for the joint distribution in the lemma since $\zeta_r =_d \varepsilon_{\sigma_{(r)}}$ because $\{\varepsilon_{n-r}, \ldots, \varepsilon_n\} \backslash \{\varepsilon_m\}$ consists of all measurement errors except r members from group 1 and 1 member from group q—as is the definition of $\varepsilon_{\sigma_{(r)}}$—and $\zeta_s =_d \varepsilon_m \sim F_q$.

An analogous derivation shows that

$$
\begin{aligned}
&\mathbb{P}(\varepsilon_{(r)} \leq \epsilon_r, E_q) \\
&\quad = \sum_{j=r}^{s-1} n_q C^{n_1}_{j-r,s-j-1,n-s} \mathbb{E}_{\sigma_{(r),s}} \Big(1\{\varepsilon_{\sigma_{(r)}} \leq \epsilon_r < \varepsilon_{\sigma_s}\} (F_{\varepsilon_1}(\epsilon_r) - F_{\varepsilon_1}(\varepsilon_{\sigma_{(r)}}))^{j-r} \times \\
&\qquad\qquad (F_{\varepsilon_1}(\varepsilon_{\sigma_s}) - F_{\varepsilon_1}(\epsilon_r))^{s-j-1} (1 - F_{\varepsilon_1}(\varepsilon_{\sigma_s}))^{n-s} \Big) \\
&\quad\quad + \sum_{j=s}^{n} n_q C^{n_1}_{s-r-1,j-s,n-j} \mathbb{E}_{\sigma_{(r),s}} \Big(1\{\varepsilon_{\sigma_{(r)}} \leq \varepsilon_{\sigma_s} \leq \epsilon_r\} (F_{\varepsilon_1}(\varepsilon_{\sigma_s}) - F_{\varepsilon_1}(\varepsilon_{\sigma_{(r)}}))^{s-r-1} \times \\
&\qquad\qquad (F_{\varepsilon_1}(\epsilon_r) - F_{\varepsilon_1}(\varepsilon_{\sigma_s}))^{j-s} \Big) (1 - F_{\varepsilon_1}(\epsilon_r))^{n-j}.
\end{aligned}
$$

□

A.2 Proofs of Main Results

Proof of Lemma 3.1. Let F be a data-consistent measurement error distribution satisfying Assumption 3.2 and $\eta \sim F$. We first prove the second part of the lemma. Note that the distribution of j^{th} order statistic $\eta_{(j)}$ is uniquely determined by F and, by independence,

$$\psi_\xi(t; F) = \psi_{X_{(j)}}(t) / \psi_{\eta_{(j)}}(t), \quad \text{for all } t \in \mathbb{R},$$

for any $j \in \{r, s\}$, where $\psi_\xi(t; F)$ is the induced latent variable distribution implied by F. Since $\psi_{\eta_{(j)}}$ is analytic (Lemma A.1), it has isolated real zeros. Thus, by the continuity of $\psi_\xi(\cdot; F)$, the equality is defined by the continuous extension at t_0 whenever $\psi_{\eta_{(j)}}(t_0) = 0$.

Now consider the first part of the lemma and note that because ξ is independent of the measurement errors, we have

$$\psi_{X_{(r,s)}}(t_r, t_s) = \psi_\xi(t_r + t_s; F) \psi_{\eta_{(r,s)}}(t_r, t_s),$$

and likewise for the marginal ch.f. Since $\psi_\xi(\cdot;F)$ is a ch.f., there exists $t_\xi > 0$ such that $\psi_\xi(t;F) \neq 0$ for all $t \in (-t_\xi, t_\xi)$. Similarly, there exists $t_\eta > 0$ such that $\psi_{\eta_{(j)}}(t) \neq 0$ for all $t \in (-t_\eta, t_\eta)$. Pick any positive $t_0 \leq \min(t_\xi, t_\eta)$ and let B_0 be an open ball around zero contained in $\{(t_r, t_s) \in \mathbb{R}^2 : |t_r + t_s| < t_0\}$. Thus, $\psi_\xi(t_r + t_s; F) \neq 0$ and $\psi_{\eta_{(j)}}(t_r + t_s) \neq 0$ for all $(t_r, t_s) \in B_0$. Then, on B_0, we have

$$\frac{\psi_{X_{(r,s)}}(t_r, t_s)}{\psi_{X_{(j)}}(t_r + t_s)} = \frac{\psi_\xi(t_r + t_s; F)\psi_{\eta_{(r,s)}}(t_r, t_s)}{\psi_\xi(t_r + t_s; F)\psi_{\eta_{(j)}}(t_r + t_s)} = \frac{\psi_{\eta_{(r,s)}}(t_r, t_s)}{\psi_{\eta_{(j)}}(t_r + t_s)}.$$

This concludes the proof. □

Proof of Lemma 3.2. By Lemma 3.1, if the distribution functions F and G are data-consistent, then

$$\psi_{\eta_{(r,s)}}(t_r, t_s)\psi_{\eta'_{(j)}}(t_r + t_s) = \psi_{\eta'_{(r,s)}}(t_r, t_s)\psi_{\eta_{(j)}}(t_r + t_s), \tag{A1}$$

for $j \in \{r, s\}$ and for all $(t_r, t_s) \in B_0$. Observe that the two products in (A1) are ch.f.s of Z_{1j} and Z_{2j}, respectively. By Lemma A.1, all ch.f.s in (A1) are analytic. Since the product of two analytic functions is also analytic, the equality of ch.f.s on B_0 implies their equality on all of $\mathbb{R}^2$. Hence, if F and G are both data-consistent, then the condition in (3) holds for $j \in \{r, s\}$. □

Proof of Lemma 3.3. We only prove the equality in (4) as the proof for (5) is analogous. Let $F_{(j)}$ and $f_{(j)}$ (resp., $G_{(j)}$ and $g_{(j)}$) denote the marginal distribution and density function of the j^{th} order statistic of a random sample of size n from F (resp. G), respectively. Also, we denote the joint distribution and density functions of order statistics from F by $F_{(r,s)}$ and $f_{(r,s)}$. Further, for any $y_r \geq 0$, define $F_{(s|r)}(\cdot; y_r)$ to be the distribution of the s^{th} order statistic conditional on the cumulative event that the r^{th} order statistic $\eta_{(r)} \leq y_r$ as in Lemma A.2.

For $c \leq 0$, the equality in (4) is an obvious consequence of the normalization in Assumption 3.3. Fix any $c > 0$ and consider a small enough δ such that $c > \delta \downarrow 0$. The conditional probability in (4) is a weighted average of the conditional distribution $F_{(s|r)}$:

$$\mathbb{P}(\eta'_{(r)} + \eta_{(s)} \leq c \mid \eta_{(r)} + \eta'_{(r)} \leq \delta) = \frac{\mathbb{P}(\eta_{(r)} + \eta'_{(r)} \leq \delta\ ,\ \eta_{(s)} + \eta'_{(r)} \leq c)}{\mathbb{P}(\eta_{(r)} + \eta'_{(r)} \leq \delta)}$$

$$
\begin{aligned}
&= \frac{\int_0^\delta F_{(r,s)}(\delta - x, c - x) g_{(r)}(x)\, dx}{\int_0^\delta F_{(r)}(\delta - x) g_{(r)}(x)\, dx} \\
&= \int_0^\delta \frac{F_{(r)}(\delta - x) g_{(r)}(x)}{\int_0^\delta F_{(r)}(\delta - x) g_{(r)}(x)\, dx} F_{(s|r)}(c - x; \delta - x)\, dx.
\end{aligned}
$$

Note that since F is absolutely continuous by Assumption 3.2, $F_{(s|r)}(y_s; \cdot)$ is continuous on $(0, \infty)$ for every $y_s \in \mathbb{R}$. Further, by the definition of $F_{(s|r)}(y_s; 0)$ as in Lemma A.2, we conclude that $F_{(s|r)}(y_s; \cdot)$ is continuous on $[0, \infty)$ for every $y_s \in \mathbb{R}$. Likewise, for every $y_r \in [0, \infty)$, $F_{(s|r)}(\cdot; y_r)$ is continuous. Finally, it follows from the monotonicity of $F_{(s|r)}(\cdot; y_r)$ for every $y_r \in [0, \infty)$ that the function $F_{(s|r)}(\cdot; \cdot)$ is (jointly) continuous on $\mathbb{R} \times [0, \infty)$ (see, e.g., Kruse and Deely (1969)).

Therefore, it follows that as $\delta \downarrow 0$,

$$
\begin{aligned}
&\int_0^\delta \frac{F_{(r)}(\delta - x) g_{(r)}(x)}{\int_0^\delta F_{(r)}(\delta - x) g_{(r)}(x)\, dx} F_{(s|r)}(c - x; \delta - x)\, dx \\
&\quad \leq F_{(s|r)}(c; \delta) + \int_0^\delta \frac{F_{(r)}(\delta - x) g_{(r)}(x)}{\int_0^\delta F_{(r)}(\delta - x) g_{(r)}(x)\, dx} \left| F_{(s|r)}(c - x; \delta - x) - F_{(s|r)}(c; \delta) \right| \, dx \\
&\quad \leq F_{(s|r)}(c; \delta) + \int_0^\delta \frac{F_{(r)}(\delta - x) g_{(r)}(x)}{\int_0^\delta F_{(r)}(\delta - x) g_{(r)}(x)\, dx} \sup_{0 \leq x \leq \delta} \left| F_{(s|r)}(c - x; \delta - x) - F_{(s|r)}(c; \delta) \right| \, dx \\
&\quad = F_{(s|r)}(c; \delta) + \sup_{0 \leq x \leq \delta} \left| F_{(s|r)}(c - x; \delta - x) - F_{(s|r)}(c; \delta) \right| \\
&\quad \longrightarrow F_{(s|r)}(c; 0),
\end{aligned}
$$

The convergence of the upper bound is due to the (joint) continuity of $F_{(s|r)}(\cdot; \cdot)$.

A similar derivation for the lower bound shows that the lower bound approaches the same limit. By the squeeze theorem and Lemma A.2, we conclude that for any $c > 0$,

$$
\mathbb{P}(\eta'_{(r)} + \eta_{(s)} \leq c \mid \eta_{(r)} + \eta'_{(r)} \leq \delta) \longrightarrow F_{(s|r)}(c; 0) = F_{s-r:n-r}(c).
$$

Thus, we conclude that the statement of the lemma holds for all $c \in \mathbb{R}$. $\square$

Proof of Theorem 3.1. Lemmas 3.2 and 3.3 imply that any two data-consistent measurement errors satisfying Assumptions 3.2 and 3.3 must have the same distribution of order statistics:

$$
F_{s-r:n-r} = G_{s-r:n-r}.
$$

It then follows from the one-to-one mapping between the distribution of an order statistic and the parent distribution (see, e.g., David and Nagaraja (2003), p.10, (2.1.5)) that $F = G$, i.e., the measurement error distribution $F_\varepsilon = F = G$ is identified. Therefore, by Lemma 3.1, the latent variable distribution F_ξ is also identified. $\square$

Proof of Corollary 3.1. The result follows directly from (3) in Lemma 3.2. Applying linear transformations $T_r : (z_1, z_2) \mapsto (z_2 - z_1, z_2)$ and $T_s : (z_1, z_2) \mapsto (z_2 - z_1, z_1)$ to (3) for $j \in \{r, s\}$, respectively, shows that

$$\begin{pmatrix} \eta_{(s)} - \eta_{(r)} \\ \eta'_{(r)} + \eta_{(s)} \end{pmatrix} \overset{d}{=} \begin{pmatrix} \eta'_{(s)} - \eta'_{(r)} \\ \eta_{(r)} + \eta'_{(s)} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \eta_{(s)} - \eta_{(r)} \\ \eta_{(r)} + \eta'_{(s)} \end{pmatrix} \overset{d}{=} \begin{pmatrix} \eta'_{(s)} - \eta'_{(r)} \\ \eta'_{(r)} + \eta_{(s)} \end{pmatrix}.$$

$\square$

Proof of Theorem 3.2. Without loss of generality, let g_1 be a group with at least $n-r$ members. For any $q \in \{1, \ldots, p\}$ (see Assumption 3.5), let E_q denote the event where the top $n - r$ order statistics are all from group 1 except for the s^{th} order statistic, which belong to group q (where it may be that $q = 1$), i.e.,

$$E_q = \{R_{(r+1)} = 1, \ldots, R_{(s-1)} = 1, R_{(s)} = q, R_{(s+1)} = 1, \ldots, R_{(n)} = 1\}. \tag{A2}$$

The identification proof proceeds in three steps. First, we claim that the distribution F_{ϵ_j} for $j \in g_1$—the distribution of a group with many members—is identified. Then we show that F_{ϵ_j} for $j \in g_q$ is identified for each $q \neq 1$. Finally, F_ξ is identified by a standard deconvolution argument.

Step 1. For notational simplicity, suppose $1 \in g_1$. A close inspection of the proof of Lemmas 3.1 and 3.2 reveals that the necessary condition (3) does not rely on the hypothesis that the measurement errors are identically distributed. Therefore, we can make use of a similar argument as in the proof of Lemmas 3.1 and 3.2, provided the ch.f. of $(\varepsilon_{(r)}, \varepsilon_{(s)})$ is analytic in the i.n.i.d. setting. Indeed, Lemma A.1 confirms $f_{(r,s)}(\cdot, \cdot; E_1)$ is analytic. Therefore, analogous to the proof of Lemmas 3.1 and 3.2 but conditional on the event E_1, two data-consistent measurement errors $\eta = (\eta_1, \ldots, \eta_n) \sim \times_{j=1}^n F_j$ and $\eta' = (\eta'_1, \ldots, \eta'_n) \sim \times_{j=1}^n G_j$ satisfying Assumption 3.4

must have the same joint distribution of sums:

$$\begin{pmatrix} \eta'_{(j)} + \eta_{(r)} \\ \eta'_{(j)} + \eta_{(s)} \end{pmatrix} \stackrel{d}{=} \begin{pmatrix} \eta_{(j)} + \eta'_{(r)} \\ \eta_{(j)} + \eta'_{(s)} \end{pmatrix} \Bigm| E_1^{\eta} \cap E_1^{\eta'}, \quad j \in \{r, s\},$$

where E_1^{η}, as in (A2), denotes the event that identifies the group association of order statistics that originate from the random vector η; and likewise for $E_1^{\eta'}$. Consider the left-hand side with $j = r$. Following a similar derivation as in the proof of Lemma 3.3, for any $c \in \mathbb{R}$, we have

$$\begin{aligned} &\mathbb{P}(\eta'_{(r)} + \eta_{(s)} \leq c \mid \eta_{(r)} + \eta'_{(r)} \leq \delta, E_1^{\eta} \cap E_1^{\eta'}) \\ &\qquad \leq F_{s-r:n-r}(c) + \sup_{0 \leq x \leq \delta} \left| F_{(s|r)}(c - x; \delta - x, E_1^{\eta}) - F_{s-r:n-r}(c) \right|, \end{aligned}$$

where $F_{s-r:n-r}$ is the distribution of the $(s-r)^{\text{th}}$ order statistic of a random sample of size $n - r$ from the parent distribution F_1. To show that the limit of the left-hand side as $\delta \downarrow 0$ is indeed $F_{s-r:n-r}(c)$, it suffices to show, in combination with a similar lower bound, that $F_{(s|r)}(\cdot;\cdot, E_1^{\eta})$ is continuous and $\lim_{\delta \downarrow 0} F_{(s|r)}(c; \delta, E_1^{\eta}) = F_{s-r:n-r}(c)$. Following the same proof as in Lemma A.2 using the expressions in Lemma A.3, we show that

$$\lim_{\delta \downarrow 0} F_{(s|r)}(c; \delta, E_1^{\eta}) = \sum_{k=s}^{n} C_{k-r}^{n-r} F_1(c)^{k-r} (1 - F_1(c))^{n-k},$$

for every $c \in \mathbb{R}$. The right-hand side equals $F_{s-r:n-r}(c)$. As in the proof of Lemma 3.3, the continuity of $F_{(s|r)}(\cdot;\cdot, E_1^{\eta})$ on $\mathbb{R} \times [0, \infty)$ follows by elementwise continuity of $F_{(s|r)}(\cdot;\cdot, E_1^{\eta})$ and monotonicity of $F_{(s|r)}(\cdot; \epsilon_r, E_1^{\eta})$ for every $\epsilon_r \in [0, \infty)$ (cf. Kruse and Deely (1969)). Therefore, we conclude that for any $c \in \mathbb{R}$,

$$\lim_{\delta \downarrow 0} \mathbb{P}(\eta'_{(r)} + \eta_{(s)} \leq c \mid \eta_{(r)} + \eta'_{(r)} \leq \delta, E_1^{\eta} \cap E_1^{\eta'}) = F_{s-r:n-r}(c).$$

A similar derivation for the right-hand side shows that

$$\begin{aligned} \lim_{\delta \downarrow 0} \mathbb{P}(\eta_{(r)} + \eta'_{(s)} \leq c \mid \eta_{(r)} + \eta'_{(r)} \leq \delta, E_1^{\eta} \cap E_1^{\eta'}) &= \sum_{k=s}^{n} C_{k-r}^{n-r} G_1(c)^{k-r} (1 - G_1(c))^{n-k} \\ &=: G_{s-r:n-r}(c), \end{aligned}$$

for every c. Therefore, we conclude that $F_{s-r:n-r} = G_{s-r:n-r}$ and thus $F_1 = G_1$, i.e., the distribution for group 1 (a group with large size) is identified.

Step 2. For notational simplicity, suppose $q \in g_q$. A sequence of arguments analogous to Step 1 remains to hold conditional on the event E_q for any $q \neq 1$. Therefore, it suffices to show that the equality

$$\lim_{\delta \downarrow 0} F_{(s|r)}(c; \delta, E_q^{\eta}) = \lim_{\delta \downarrow 0} G_{(s|r)}(c; \delta, E_q^{\eta'}), \qquad \text{for all } c \in \mathbb{R}, \tag{A3}$$

implies that $F_q = G_q$, where

$$F_{(s|r)}(c; \delta, E_q^{\eta}) = \mathbb{P}(\eta_{(s)} \leq c \,|\, \eta_{(r)} \leq \delta, E_q^{\eta}),$$

and $G_{(s|r)}(\cdot; \cdot, E_q^{\eta'})$ is defined similarly. Following the same proof as in Lemma A.2 using the expression in Lemma A.4, one can show that

$$\begin{aligned} F_{(s|r)}(c; 0, E_q^{\eta}) &:= \lim_{\delta \downarrow 0} F_{(s|r)}(c; \delta, E_q^{\eta}) \\ &= \sum_{k=s}^{n} C_{k-s}^{n-s} \frac{\mathbb{E}_{\zeta_s} 1\{\zeta_s \leq c\} F_1(\zeta_s)^{s-r-1}(F_1(c) - F_1(\zeta_s))^{k-s}(1 - F_1(c))^{n-k}}{\mathbb{E}_{\zeta_s} F_1(\zeta_s)^{s-r-1}(1 - F_1(\zeta_s))^{n-s}}, \end{aligned} \tag{A4}$$

where $\zeta_s \sim F_q$, the group-q distribution. Differentiating (A4) with respect to c, the density function is given by

$$\begin{aligned} f_{(s|r)}(c; 0, E_q^{\eta}) &= \frac{n-s}{\mathbb{E}_{\zeta_s} F_1(\zeta_s)^{s-r-1}(1 - F_1(\zeta_s))^{n-s}} F_1(c)^{s-r-1}(1 - F_1(c))^{n-s} f_q(c) \\ &\propto F_1(c)^{s-r-1}(1 - F_1(c))^{n-s} f_q(c), \end{aligned}$$

for almost all $c \in \mathbb{R}$. Similarly, the density $g_{(s|r)}(\cdot; 0, E_q^{\eta'})$ is given by

$$\begin{aligned} g_{(s|r)}(c; 0, E_q^{\eta}) &= \frac{n-s}{\mathbb{E}_{\zeta'_s} F_1(\zeta'_s)^{s-r-1}(1 - F_1(\zeta'_s))^{n-s}} F_1(c)^{s-r-1}(1 - F_1(c))^{n-s} g_q(c) \\ &\propto F_1(c)^{s-r-1}(1 - F_1(c))^{n-s} g_q(c), \end{aligned}$$

for almost all $c \in \mathbb{R}$, where $\zeta'_s \sim G_q$. Since (A3) implies $f_{(s|r)}(\cdot; 0, E_q^{\eta}) = g_{(s|r)}(\cdot; 0, E_q^{\eta})$ a.e. and $F_1(c)^{s-r-1}(1 - F_1(c))^{n-s} > 0$ for all c in the interior of the support of F_q (common support assumption in Assumption 3.5), we conclude from the above expression of the densities that $f_q = g_q$ a.e., and thus $F_q = G_q$ for all $q \neq 1$. Therefore, the distributions for all measurement error groups are identified.

Step 3. Since $F_{\varepsilon_1}, \ldots, F_{\varepsilon_n}$ are identified, so is the ch.f. $\psi_{\varepsilon_{(j)}}$. It follows that the ch.f. $\psi_\xi = \psi_{X_{(j)}}/\psi_{\varepsilon_{(j)}}$ is identified, where $\psi_\xi(t)$ is defined by the continuous extension whenever $\psi_{\varepsilon_{(j)}}(t) = 0$. It is well-defined since $\psi_{\varepsilon_{(j)}}$ is analytic and hence has isolated zeros. Thus, F_ξ is identified. □

Proof of Corollary 3.2. When $s = n$ and $r = n-1$, the proof of Theorem 3.2 reveals that

$$f_{(s|r)}(c; 0, E_q^\eta) \propto f_q(c), \qquad g_{(s|r)}(c; 0, E_q^\eta) \propto g_q(c).$$

Above implies $f_q = g_q$ a.e. in the absence of a common support assumption. □

Proof of Theorem 4.1. It follows from Theorem 3 in Bierens (2008) that the sieve $\{\mathcal{H}_k\}_k$ is dense in $\mathcal{H} = \{h \in \pi^2 : \pi \in \mathcal{P}\}$. This implies that $\{\mathcal{F}_k\}_k := \{\mathcal{F}_k^\xi \times \mathcal{F}_k^\varepsilon\}_k$ is dense in $\mathcal{F} := \mathcal{F}^\xi \times \mathcal{F}^\varepsilon$ where

$$\begin{aligned}
\mathcal{F}^\xi &= \{F(\cdot) = \int_0^{G_\xi(\cdot)} h(u)\,du \,:\, h \in \mathcal{H}\}, \quad \mathcal{F}^\varepsilon = \{F(\cdot) = \int_0^{G_\varepsilon(\cdot)} h(u)\,du \,:\, h \in \mathcal{H}\},\\
\mathcal{F}_k^\xi &= \{F(\cdot) = \int_0^{G_\xi(\cdot)} h(u)\,du \,:\, h \in \mathcal{H}_k\}, \quad \mathcal{F}_k^\varepsilon = \{F(\cdot) = \int_0^{G_\varepsilon(\cdot)} h(u)\,du \,:\, h \in \mathcal{H}_k\}.
\end{aligned}$$

We endow $\mathcal{F}$ with the supremum norm $\|\cdot\|_\infty$. Further, because $\mathcal{F}$ is compact, it suffices to show that (1) $Q(\cdot)$ is continuous on $\mathcal{F}$, (2) $Q(\cdot)$ has a unique minimizer on $\mathcal{F}$, and (3) the following uniform a.s. convergence holds:

$$\sup_{F \in \mathcal{F}} \left| \widehat{Q}_N(F) - Q(F) \right| \xrightarrow{\text{a.s.}} 0.$$

See Gallant (1987) or Gallant and Nychka (1987). A proof of each of (1)–(3) follows.

(1) Given a complex number $z \in \mathbb{C}$, let $\Re(z)$ and $\Im(z)$ denote the real and imaginary part, respectively. Let $F_k = (F_{\xi,k}, F_{\varepsilon,k}) \in \mathcal{F}$ such that $F_k \to F \in \mathcal{F}$. Consider:

$$\begin{aligned}
&|Q(F) - Q(F_k)| \\
&\quad = \frac{1}{4\kappa^2} \left| \int_{(-\kappa,\kappa)^2} \Re\{\psi_{X_{(r,s)}}(t) - \varphi(t;F)\}^2 + \Im\{\psi_{X_{(r,s)}}(t) - \varphi(t;F)\}^2\, dt \right.
\end{aligned}$$

$$
\begin{aligned}
&\qquad - \int_{(-\kappa,\kappa)^2} \Re\{\psi_{X_{(r,s)}}(t) - \varphi(t;F_k)\}^2 + \Im\{\psi_{X_{(r,s)}}(t) - \varphi(t;F_k)\}^2 \, dt \Bigg| \\
&= \frac{1}{4\kappa^2} \Bigg| \int_{(-\kappa,\kappa)^2} \Re\{2\psi_{X_{(r,s)}}(t) - \varphi(t;F) - \varphi(t;F_k)\} \Re\{\varphi(t;F) - \varphi(t;F_k)\} \\
&\qquad + \Im\{2\psi_{X_{(r,s)}}(t) - \varphi(t;F) - \varphi(t;F_k)\} \Im\{\varphi(t;F) - \varphi(t;F_k)\} \, dt \Bigg| \\
&\le \frac{1}{4\kappa^2} \int_{(-\kappa,\kappa)^2} \Big| \Re\{2\psi_{X_{(r,s)}}(t) - \varphi(t;F) - \varphi(t;F_k)\} \Big| \Big| \Re\{\varphi(t;F) - \varphi(t;F_k)\} \Big| \\
&\qquad + \Big| \Im\{2\psi_{X_{(r,s)}}(t) - \varphi(t;F) - \varphi(t;F_k)\} \Big| \Big| \Im\{\varphi(t;F) - \varphi(t;F_k)\} \Big| \, dt \\
&\le \frac{1}{\kappa^2} \int_{(-\kappa,\kappa)^2} \Big| \Re\{\varphi(t;F) - \varphi(t;F_k)\} \Big| + \Big| \Im\{\varphi(t;F) - \varphi(t;F_k)\} \Big| \, dt \\
&\le \frac{\sqrt{2}}{\kappa^2} \int_{(-\kappa,\kappa)^2} \Big| \varphi(t;F) - \varphi(t;F_k) \Big| \, dt \\
&\le 4\sqrt{2} \sup_{t \in (-\kappa,\kappa)^2} \Big| \varphi(t;F) - \varphi(t;F_k) \Big| .
\end{aligned}
$$

Therefore, to show $|Q(F) - Q(F_k)| \to 0$, it suffices to show that

$$
(\xi(F_{\xi,k}), \varepsilon_1(F_{\varepsilon,k}), \ldots, \varepsilon_n(F_{\varepsilon,k})) \xrightarrow{\text{a.s.}} (\xi(F_\xi), \varepsilon_1(F_\varepsilon), \ldots, \varepsilon_n(F_\varepsilon)),
$$

as this implies a.s. convergence of $(\xi(F_{\xi,k}), \varepsilon_{(r)}(F_{\varepsilon,k}), \varepsilon_{(s)}(F_{\varepsilon,k}))$ to $(\xi(F_\xi), \varepsilon_{(r)}(F_\varepsilon), \varepsilon_{(s)}(F_\varepsilon))$ and thus uniform convergence of $\varphi(t;F_k)$ to $\varphi(t;F)$ on any compact set. For each k, let H_k denote the c.d.f. such that $F_{\xi,k}(\cdot) = H_k(G_\xi(\cdot))$. $F_{\xi,k} \to F_\xi$ implies $H_k \to H$ where $F_\xi(\cdot) = H(G_\xi(\cdot))$. Therefore, since G_ξ^{-1} and H^{-1} are monotone and hence have at most countable discontinuity points, we conclude that V–a.s.,

$$
\lim_{k\to\infty} \xi(F_{\xi,k}) = \lim_{k\to\infty} F_{\xi,k}^{-1}(V) = \lim_{k\to\infty} G_\xi^{-1}(H_k^{-1}(V)) = G_\xi^{-1}(H^{-1}(V)) = F_\xi^{-1}(V) = \xi(F_\xi).
$$

Similarly, we have $\lim_{k\to\infty} \varepsilon_j(F_{\varepsilon,k}) = \varepsilon_j(F_\varepsilon)$, U_j–a.s. We thus conclude that $Q(\cdot)$ is continuous on $\mathcal{F}$.

(2) Existence of $F = (F_\xi, F_\varepsilon) \in \mathcal{F}$ satisfying $Q(F) = 0$ is obvious from Assumption 3.4(b). Uniqueness of the minimizer follows from Theorem 3.1.

(3) Note that we have

$$
\sup_{F\in\mathcal{F}} \Big| \widehat{Q}_N(F) - Q(F) \Big|
$$

$$
\begin{aligned}
&= \sup_{F\in\mathcal{F}} \frac{1}{4\kappa^2}\Bigg|\int_{(-\kappa,\kappa)^2} \Re\{\widehat{\psi}_N(t)-\widehat{\varphi}_N(t;F)\}^2 + \Im\{\widehat{\psi}_N(t)-\widehat{\varphi}_N(t;F)\}^2\,dt \\
&\qquad\qquad - \int_{(-\kappa,\kappa)^2} \Re\{\psi_{X_{(r,s)}}(t)-\varphi(t;F)\}^2 + \Im\{\psi_{X_{(r,s)}}(t)-\varphi(t;F)\}^2\,dt\Bigg| \\
&= \sup_{F\in\mathcal{F}} \frac{1}{4\kappa^2}\Bigg|\int_{(-\kappa,\kappa)^2} \Re\{\widehat{\psi}_N(t)-\widehat{\varphi}_N(t;F)+\psi_{X_{(r,s)}}(t)-\varphi(t;F)\} \\
&\qquad\qquad\qquad \Re\{(\widehat{\psi}_N(t)-\psi_{X_{(r,s)}}(t))-(\widehat{\varphi}_N(t;F)-\varphi(t;F))\} \\
&\qquad\qquad + \Im\{\widehat{\psi}_N(t)-\widehat{\varphi}_N(t;F)+\psi_{X_{(r,s)}}(t)-\varphi(t;F)\} \\
&\qquad\qquad\qquad \Im\{(\widehat{\psi}_N(t)-\psi_{X_{(r,s)}}(t))-(\widehat{\varphi}_N(t;F)-\varphi(t;F))\}\,dt\Bigg| \\
&\le \sup_{F\in\mathcal{F}} \frac{1}{\kappa^2}\int_{(-\kappa,\kappa)^2} \Big|\Re\{(\widehat{\psi}_N(t)-\psi_{X_{(r,s)}}(t))-\Re(\widehat{\varphi}_N(t;F)-\varphi(t;F))\}\Big| \\
&\qquad\qquad + \Big|\Im\{(\widehat{\psi}_N(t)-\psi_{X_{(r,s)}}(t))-\Im(\widehat{\varphi}_N(t;F)-\varphi(t;F))\}\Big|\,dt \\
&\le \sup_{F\in\mathcal{F}} \frac{2}{\kappa^2}\int_{(-\kappa,\kappa)^2} \Big|\widehat{\psi}_N(t)-\psi_{X_{(r,s)}}(t)\Big| + \Big|\widehat{\varphi}_N(t;F)-\varphi(t;F)\Big|\,dt \\
&\le \frac{2}{\kappa^2}\int_{(-\kappa,\kappa)^2} \Big|\widehat{\psi}_N(t)-\psi_{X_{(r,s)}}(t)\Big|\,dt + \frac{2}{\kappa^2}\int_{(-\kappa,\kappa)^2}\sup_{F\in\mathcal{F}}\Big|\widehat{\varphi}_N(t;F)-\varphi(t;F)\Big|\,dt.
\end{aligned}
$$

Recall ch.f.s are bounded uniformly by 1. Since $\widehat{\psi}_N \to_{\text{a.s.}} \psi_{X_{(r,s)}}$ pointwise on $\mathbb{R}^2$, the first integral on the right-hand side a.s. vanishes asymptotically. It remains to be shown that $\widehat{\varphi}_N(t;\cdot) \to_{\text{a.s.}} \varphi(t;\cdot)$ uniformly in F. the supremum in the last integral vanishes pointwise on $(-\kappa,\kappa)^2$. Since, by definition

$$
\begin{aligned}
&\Big|\widehat{\varphi}_N(t;F)-\varphi(t;F)\Big| \\
&\quad = \Big|\mathbb{E}_N e^{it^\top X(F)} - \mathbb{E} e^{it^\top X(F)}\Big| \\
&\quad \le \Big|\mathbb{E}_N \cos(t^\top X(F)) - \mathbb{E}\cos(t^\top X(F))\Big| + \Big|\mathbb{E}_N \sin(t^\top X(F)) - \mathbb{E}\sin(t^\top X(F))\Big|,
\end{aligned}
$$

where both terms are bounded, we conclude from the Borel–Cantelli lemma and Hoeffding's inequality that

$$
\Big|\widehat{\varphi}_N(t;F)-\varphi(t;F)\Big| \xrightarrow{\text{a.s.}} 0, \quad \text{for all } F\in\mathcal{F}. \tag{A5}
$$

We complete the proof by establishing its strong stochastic equicontinuity. Note that a.s. continuity of $(\xi(F), \varepsilon_1(F), \ldots, \varepsilon_n(F))$ in part (1) of the proof implies a.s. continuity of $X(F)$. Thus, for any $\eta > 0$, there exists $\delta > 0$ such that

$$d(F, F') = \max\big\{\|F_\xi - F'_\xi\|_\infty, \|F_\varepsilon - F'_\varepsilon\|_\infty\big\} < \delta$$

implies, a.s.,

$$\left|\cos(t^\top X(F)) - \cos(t^\top X(F'))\right| < \eta/4, \quad \left|\sin(t^\top X(F)) - \sin(t^\top X(F'))\right| < \eta/4.$$

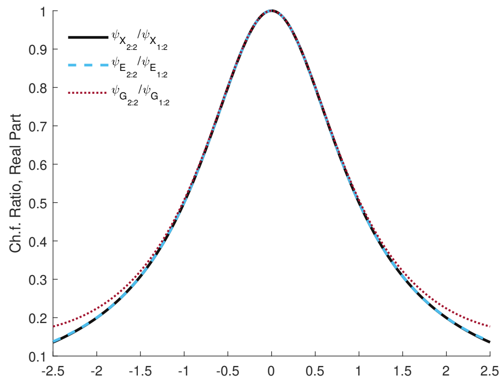
Therefore,

$$\begin{aligned}
&\sup_{d(F,F')<\delta}\left|\Big(\widehat{\varphi}_N(t;F) - \varphi(t;F)\Big) - \Big(\widehat{\varphi}_N(t;F') - \varphi(t;F')\Big)\right| \\
&\quad = \sup_{d(F,F')<\delta}\left|\Big(\mathbb{E}_N e^{it^\top X(F)} - \mathbb{E} e^{it^\top X(F)}\Big) - \Big(\mathbb{E}_N e^{it^\top X(F')} - \mathbb{E} e^{it^\top X(F')}\Big)\right| \\
&\quad \le \sup_{d(F,F')<\delta}\left|\Big(\mathbb{E}_N e^{it^\top X(F)} - \mathbb{E}_N e^{it^\top X(F')}\Big) - \Big(\mathbb{E} e^{it^\top X(F)} - \mathbb{E} e^{it^\top X(F')}\Big)\right| \\
&\quad \le \mathbb{E}_N \sup_{d(F,F')<\delta}\Big(\left|\cos(t^\top X(F)) - \cos(t^\top X(F'))\right| + \left|\sin(t^\top X(F)) - \sin(t^\top X(F'))\right|\Big) \\
&\qquad + \mathbb{E} \sup_{d(F,F')<\delta}\Big(\left|\cos(t^\top X(F)) - \cos(t^\top X(F'))\right| + \left|\sin(t^\top X(F)) - \sin(t^\top X(F'))\right|\Big) \\
&\quad < \eta.
\end{aligned}$$

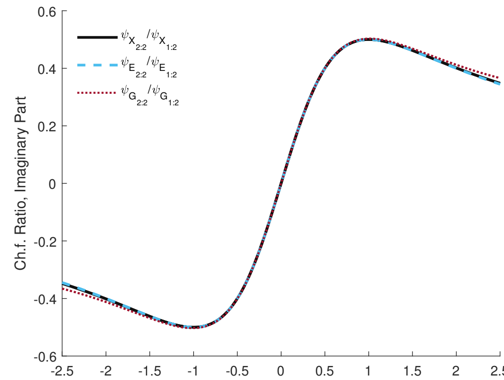
Since the inequality holds for all N, we conclude by strong stochastic equicontinuity that the a.s. convergence in (A5) holds uniformly on $\mathcal{F}$. □

Appendix B Rossberg's Counterexample

One natural approach towards investigating the identification problem of the measurement error distribution is to exploit the spacing between two order statistics: $X_{(s)} - X_{(r)} = \varepsilon_{(s)} - \varepsilon_{(r)}$. However, without additional restrictions, knowledge of the spacing distribution is not sufficient to identify F_ε. A constructive example is provided by Rossberg (1972). Specifically, suppose ε_1 and ε_2 are i.i.d. standard exponential random variables. Then the spacing $\varepsilon_{(2)} - \varepsilon_{(1)}$ is a standard exponential random variable. Rossberg (1972) shows there are infinitely many parent distributions that



(a) Real part of the ratios of ch.f.s.



(b) Imaginary part of the ratios of ch.f.s.

Figure 2: An illustration of the departure of the ratio of ch.f.s of second and first Rossberg order statistics (*red dotted line*) from that of the measurement (observed) order statistics (*black line*). The ratio of ch.f.s of true measurement error (exponential) order statistics (*blue dashed line*) aligns with that of the measurement order statistics.

deliver a standard exponential spacing distribution, one of them being

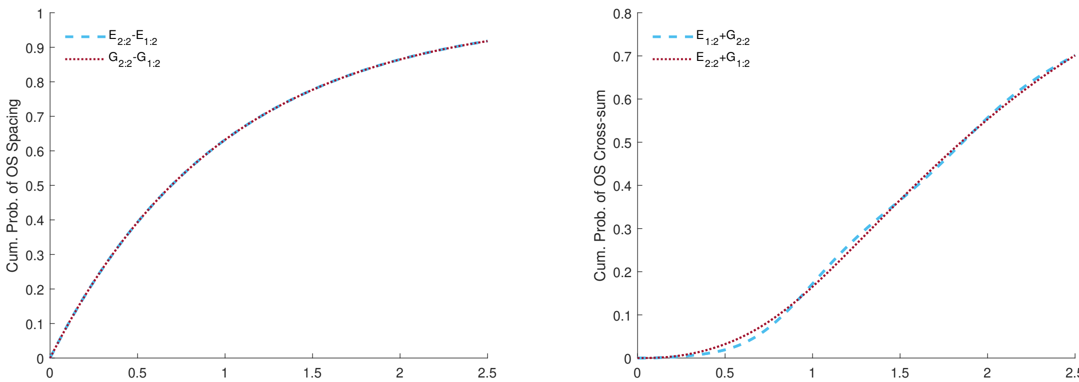
$$G(x) = 1 - e^{-x}\big[1 + \pi^{-2}(1 - \cos 2\pi x)\big],$$

with support $S(G) = [0, \infty)$.

However, our Corollary 3.1 requires that, for Rossberg's distribution G to be observationally equivalent to the true exponential distribution, the distributions of cross-sums of order statistics must be the same in addition to their spacing distributions. This additional constraint on the distribution of cross-sum that is absent in Rossberg (1972) arises from our model setup, in particular, from the within independence between the latent variable of interest, ξ, and the measurement errors. Whereas the empirical content comes from the random vector $(X_{(1)}, X_{(2)})$ in our context, the spacing is the only—and complete—empirical content available in the setting of Rossberg (1972). As discussed in Section 3.2, the independence assumption allows us to exploit the ratio of ch.f.s in a tractable manner despite the dependence between measurement errors of the observed order statistics.

Figure 2 compares the ratios of ch.f.s of order statistics.[27] In order to make

[27]Due to the lack of analytical tractability of the ratio for Rossberg's distribution, all functions are approximated by Monte Carlo integration with $N = 10^8$ pairs of random draws. To pin down the specification of $X_j = \xi + \varepsilon_j$, random quantity ξ is drawn from the standard normal distribution. The same draws are used to plot the distribution functions in Figure 3.



(a) Cumulative distributions of the spacings. (*blue dashed line*: exponential spacing; *red dotted line*: Rossberg spacing)

(b) Cumulative distributions of the cross-sums. (*blue dashed line*: exponential first order statistic; *red dotted line*: Rossberg first order statistic)

Figure 3: An illustration of the dissimilarity between the probability distributions of cross-sums of two exponential and Rossberg order statistics (Figure 3b). Consistent with Rossberg (1972), the probability distributions of exponential and Rossberg spacings are aligned (Figure 3a).

the minor departures clear, we forfeit the three-dimensional display of the ratios of ch.f.s in (1). The ratios are evaluated at $\psi_{X_{(1,2)}}(0,t)/\psi_{X_{(1)}}(t) = \psi_{X_{(2)}}(t)/\psi_{X_{(1)}}(t)$, and the real and imaginary parts are displayed separately. The ratios for the observations $(X_{(1)}, X_{(2)})$ and the true measurement errors $(\varepsilon_{(1)}, \varepsilon_{(2)})$ from the standard exponential distribution coincide. However, the ratio for G begins to depart substantially when $|t| > 1$, indicating that G cannot rationalize the observed data. Consequently, in Figure 3, while the spacing distributions for the exponential and Rossberg's random variables overlap, the cross-sum distributions differ over nontrivial regions. This connection between ratios of ch.f.s and probability distributions of spacings and cross-sums is established in Corollary 3.1.

Appendix C Implementation of the Estimator

This section provides a closed form for the empirical criterion function and describes how the estimation procedure is implemented in practice to calculate the estimator proposed in Section 4. Additional simulation results are also reported. While it is left implicit in the main text whether the vectorized form of an element is a column or row vector, it is made clear here. For instance, $x = (x_1, \ldots, x_n)$ is a row vector of length n and $x^\top$ a column vector.

C.1 Closed form for the Criterion Function

The simulated sieve estimator minimizes the empirical analogue of (9). By simulating the model-implied ch.f. φ, we avoid having to numerically integrate the criterion function over $(-\kappa, \kappa)^2$. Precisely, the empirical criterion function $\widehat{Q}_N(\cdot)$ has the following closed-form:

$$\widehat{Q}_N(F_{k_N};\kappa) = \frac{2}{N} + \frac{2}{N^2}\sum_{i>j}^{N} q(X_i - X_j) + \frac{2}{N^2}\sum_{i>j}^{N} q(X_i(F_{k_N}) - X_j(F_{k_N})) - \frac{2}{N^2}\sum_{i,j}^{N} q(X_i - X_j(F_{k_N})), \tag{C1}$$

where $X_i = (X_{(r),i}, X_{(s),i})^\top$ is the i^{th} observation, $X_i(F_{k_N})$ is the i^{th} simulated column vector given c.d.f.s F_{ξ,k_N} and F_{ε,k_N}, to be made precise below (see also Assumption 3.4(d)), and $q(x,y) = \sin(\kappa x)\sin(\kappa y)/(\kappa^2 xy)$, defined everywhere by the continuous extension. The above closed form is straightforward to derive using the identity $e^{ix} = \cos(x) + i\sin(x)$ and two trigonometric identities, $\cos(x-y) = \cos(x)\cos(y) + \sin(x)\sin(y)$ and $2\sin(x)\sin(y) = \cos(x-y) - \cos(x+y)$.

C.2 Estimation Procedure

Note that the sieve space $\{\mathcal{F}_k\}_k := \{\mathcal{F}_k^\xi \times \mathcal{F}_k^\varepsilon\}_k$ introduced in Section 4 is constructed by $\{\mathcal{P}_k\}_k$, a sequence of spaces of truncated series based on the series representation $\mathcal{P}$ in (8). That is, the infinite-dimensional spaces $\{\mathcal{P}_k\}_k$ are determined by a sequence of finite-dimensional sieves $\{\mathcal{D}_k\}_k$, where

$$\mathcal{D}_k := \left\{\delta \in \mathbb{R}^k : |\delta_\ell| \le \frac{c}{1+\sqrt{\ell}\ln\ell},\ \ell = 1,\dots,k\right\}.$$

Thus, so as to minimize $\widehat{Q}_N$ in (C1) with respect to $\mathcal{F}_{k_N}$, one can minimize

$$\widehat{Q}_N(F_{k_N}) = \widehat{Q}_N\big(H_k(G_\xi(\cdot);\delta_\xi), H_k(G_\varepsilon(\cdot);\delta_\varepsilon)\big)$$

with respect to $(\delta_\xi, \delta_\varepsilon) \in \mathcal{D}_{k_N}^2$, where

$$H_k(v;\delta) = \frac{\int_0^v \big(1 + \sum_{\ell=1}^k \delta_\ell \rho_\ell(v)\big)^2\, dv}{1 + \sum_{\ell=1}^k \delta_\ell^2}.$$

To alleviate computational concerns associated with solving the above finite-dimensional optimization problem with the parameterization in (8), Bierens (2008) suggests a reparametrization under which the c.d.f. $H_k(\cdot;\delta)$ on $[0,1]$ can be expressed as (see Sections 3 and 7 in Bierens (2008)):

$$H_k(v;\theta) = \frac{(1-\pi_k^\top\theta, \theta^\top)\Pi_{k+1}(v)(1-\pi_k^\top\theta, \theta^\top)^\top}{(1-\pi_k^\top\theta, \theta^\top)\Pi_{k+1}(1)(1-\pi_k^\top\theta, \theta^\top)^\top}, \quad v \in [0,1],$$

where $\Pi_{k+1}(v)$ is a $(k+1)$-square matrix and π_k a k-dimensional vector defined as

$$\Pi_{k+1}(v) = \left(\frac{v^{i+j+1}}{i+j+1}\right)_{i,j=0,1,\ldots,k} \qquad \text{and} \qquad \pi_k = \left(\frac{1}{i+1}\right)_{i=1,\ldots,k}.$$

The compactifying restrictions on $\delta \in \mathcal{D}_k$ in the definition of $\mathcal{D}_k$ translates to the following constraints on the space Θ_k for θ:

$$\Theta_k = \left\{\theta \in \mathbb{R}^k : \left|\sum_{m=0}^{k-\ell} \theta_{\ell+m}\mu_\ell(m)\right| \leq \frac{c}{1+\sqrt{\ell}\ln\ell}, \ \ell = 1,\ldots,k\right\},$$

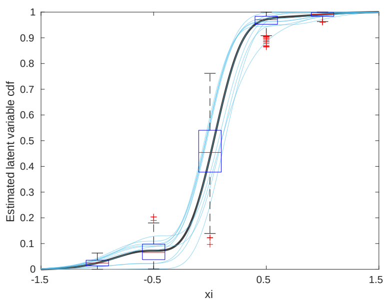
where $\mu_\ell(m) := \int_0^1 u^{\ell+m}\rho_\ell(u)\,du$.

We use the finite-dimensional sieves $\{\Theta_k^2\}_k$ for $(\theta_\xi, \theta_\varepsilon)$ to estimate the distribution functions in the Monte Carlo study in Section 4.2. For further details regarding the sieve space, we refer the readers to the original article by Bierens (2008).

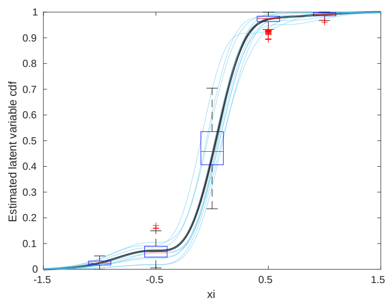
C.3 Additional Simulation Results

In this section, we report additional results from the simulation exercise described in Section 4.2. Figures 4 and 5 display pointwise box plots that summarizes the simulated coverage of the estimator for F_ξ and F_ε. Sample sizes $N = 1000$, 2000, and 4000 with $k = 4$, 5, and 6, respectively, are considered with $\kappa = 1$, 3.14, and 5 for each sample size.[28] For each κ, the result shows that the estimator improves with larger N. The pointwise variance, as indicated by more values outside the interquartile range, appears to be higher when κ is set to a larger value.

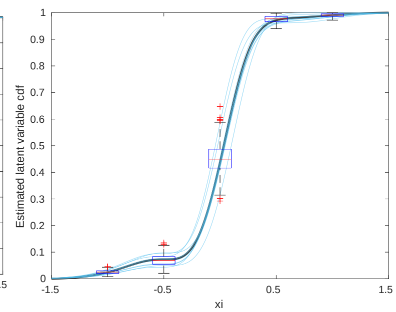
[28] Panel 1 of Figures 4 and 5 are reproduced in Figure 1.



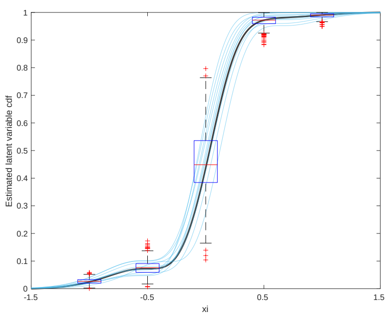
Panel 1(a)



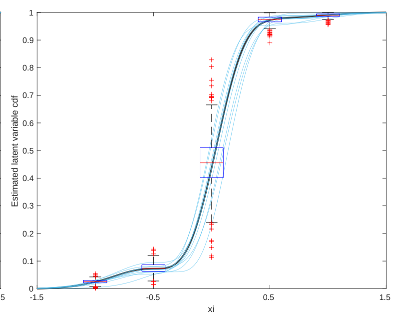
Panel 1(b)



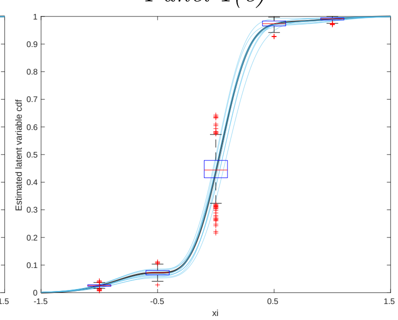
Panel 1(c)



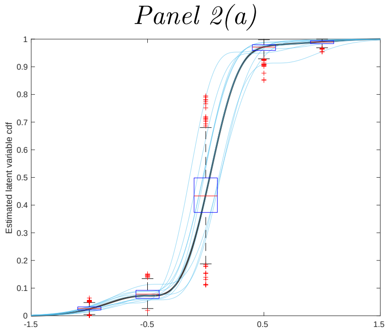
Panel 2(a)



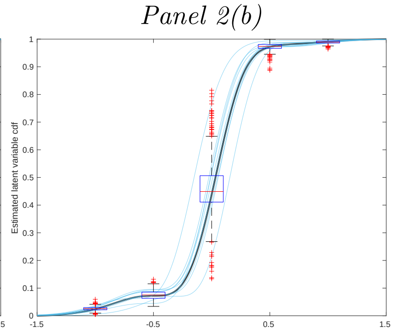
Panel 2(b)



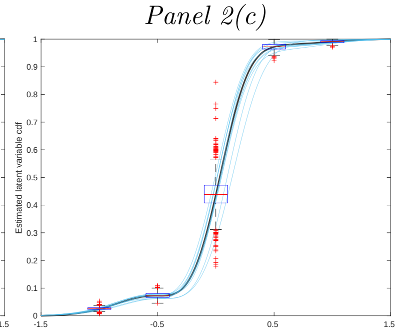
Panel 2(c)



Panel 3(a)



Panel 3(b)



Panel 3(c)

Figure 4: Monte Carlo simulation results for F_ξ. Panels 1, 2, and 3 correspond to the tuning parameter $\kappa = 1$, 3.14, and 5, respectively; Subpanels *(a)*, *(b)*, and *(c)* correspond to sample sizes $N = 1000$, 2000, and 4000, respectively.

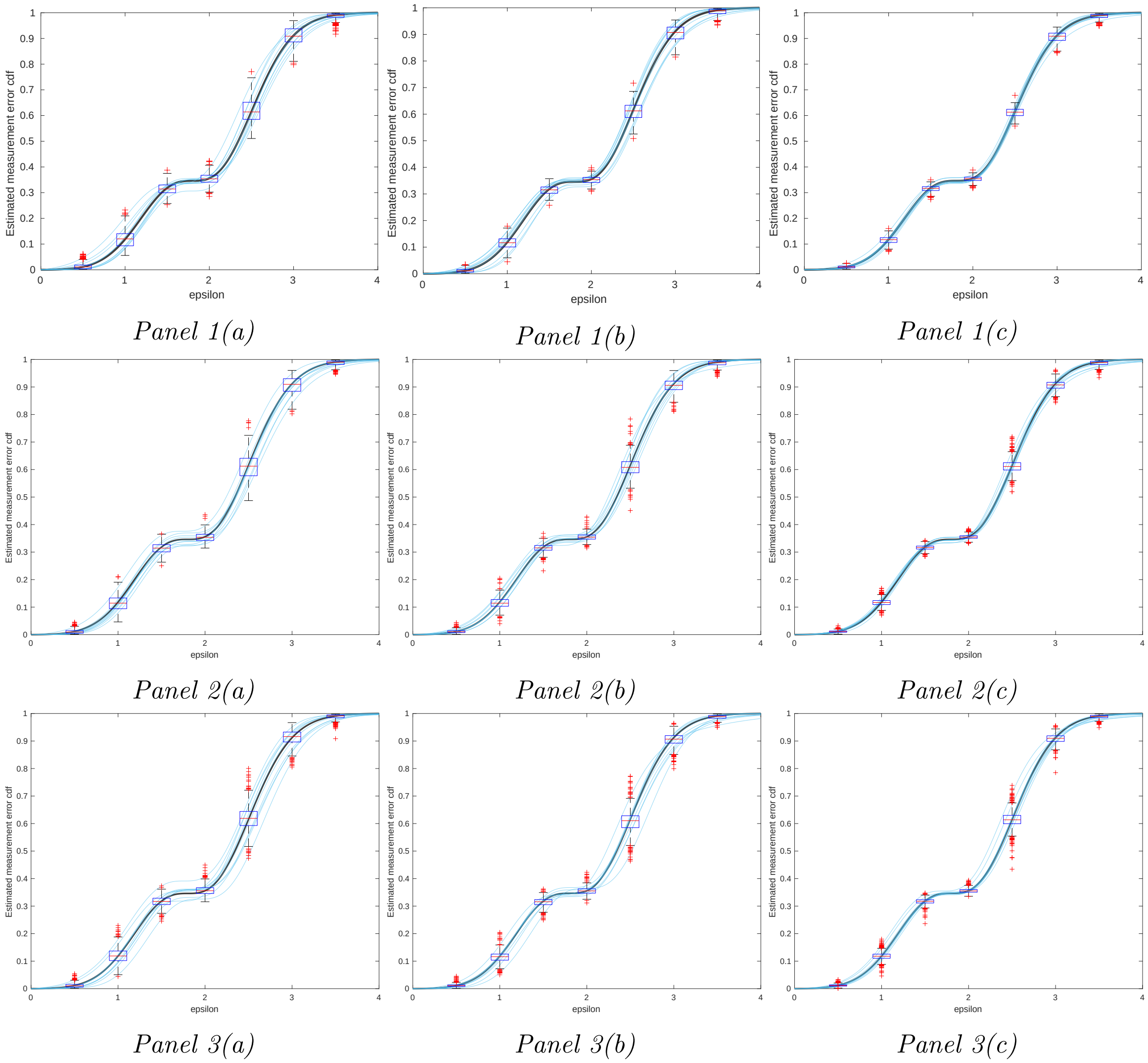


Figure 5: Monte Carlo simulation results for F_ε. Panels 1, 2, and 3 correspond to the tuning parameter $\kappa = 1$, 3.14, and 5, respectively; Subpanels *(a)*, *(b)*, and *(c)* correspond to sample sizes $N = 1000$, 2000, and 4000, respectively.