Statistica Sinica **33** (2023), 1003-1024
doi:https://doi.org/10.5705/ss.202021.0070

SEMIPARAMETRIC ESTIMATION OF PROBABILISTIC INDEX MODELS: EFFICIENCY AND BIAS

Karel Vermeulen, Jan De Neve, Gustavo Amorim, Olivier Thas and Stijn Vansteelandt

Ghent University, Vanderbilt University, Hasselt University, London School of Hygiene and Tropical Medicine

Abstract: Many well-known rank tests can be viewed as score tests under probabilistic index models (PIMs), that is, regression models for the conditional probability that the outcome of one randomly chosen subject exceeds the outcome of another independently chosen subject. PIMs provide a natural regression framework for nonparametric rank tests. In addition, PIMs supplement these tests with effect sizes and ease the development of more flexible tests, such as tests that allow for covariate adjustment. Inferences for PIMs are currently based on an estimator, referred to as the standard estimator, that is derived heuristically. By appealing to semiparametric theory and a Hoeffding decomposition, we rigorously derive the class of all consistent and asymptotically normal estimators for the parameters indexing a PIM. We identify the (locally) semiparametric efficient estimator in this class, and derive a second estimator with a smaller second-order finite-sample bias. The properties of the estimators are evaluated theoretically and empirically. The heuristic standard estimator turns out to be the preferred estimator in practice, because it is computationally superior to both the efficient and the bias-reduced estimators, and only suffers from a minor loss in efficiency. We also propose a partition strategy to further improve the computational performance of the standard estimator.

Key words and phrases: Cross correlation, influence function, second-order bias, semiparametric estimation, U-process.

1. Introduction

Probabilistic index models (PIMs, Thas et al. (2012)) form a class of semiparametric models for the probability that the outcome of one randomly chosen subject exceeds the outcome of another independently chosen subject, as a function of covariates. Let $\{\mathbf{Z}_i^T = (Y_i, \mathbf{X}_i^T) : i = 1, \ldots, n\}$ denote a sample of n independent and identically distributed (i.i.d.) random vectors, where Y_i denotes

Corresponding author: Jan De Neve, Department of Data Analysis, Ghent University, Ghent, Belgium. E-mail: Jan.DeNeve@UGent.be.

the outcome of interest associated with the p-dimensional vector of covariates $\mathbf{X}_i$. A PIM is defined by the constraint

$$\mathrm{P}\left(Y_i \preccurlyeq Y_j \mid \mathbf{X}_i, \mathbf{X}_j\right) = m(\mathbf{X}_i, \mathbf{X}_j; \boldsymbol{\beta}_0), \tag{1.1}$$

where $\mathrm{P}\left(Y_i \preccurlyeq Y_j \mid \mathbf{X}_i, \mathbf{X}_j\right) := \mathrm{P}\left(Y_i < Y_j \mid \mathbf{X}_i, \mathbf{X}_j\right) + 0.5\mathrm{P}\left(Y_i = Y_j \mid \mathbf{X}_i, \mathbf{X}_j\right)$. This probability is referred to as the *probabilistic index* (PI). The function $m(\cdot)$ is a known function with range $[0, 1]$, smooth in the p-dimensional parameter vector $\boldsymbol{\beta}$, and satisfying the antisymmetry condition $m(\mathbf{X}_i, \mathbf{X}_j; \boldsymbol{\beta}) = 1 - m(\mathbf{X}_j, \mathbf{X}_i; \boldsymbol{\beta})$. The function $m(\cdot)$ typically takes the form $m(\mathbf{X}_i, \mathbf{X}_j; \boldsymbol{\beta}) = g^{-1}\{(\mathbf{X}_j - \mathbf{X}_i)^T\boldsymbol{\beta}\}$, with $g(\cdot)$ an appropriate link function, such as the probit or logit link. We let $\boldsymbol{\beta}_0$ denote the true, but unknown value of $\boldsymbol{\beta}$ that corresponds to the observed data law.

Thas et al. (2012) demonstrate that PIMs form a versatile class of models applicable to continuous and ordinal outcomes, and they establish connections with the Cox proportional hazards model and rank regression. De Neve and Thas (2015) illustrate how PIMs provide a unified regression framework for many rank tests, such as the Wilcoxon–Mann–Whitney (WMW), Kruskal–Wallis, and Friedman rank tests. An attractive feature of a PIM is that, in addition to hypothesis testing, it enables us to estimate the effect sizes with an informative interpretation; however, as discussed in Section 4, interpreting a PI requires caution. De Neve and Thas (2015) show how PIMs can be used to construct new rank tests for complicated designs. This is further convincingly demonstrated in Vermeulen, Thas and Vansteelandt (2015), who employ PIMs to increase the power of the WMW test by including auxiliary covariate information in randomized designs.

The parameter estimation and statistical inference proposed in Thas et al. (2012) rely on reformulating the PIM (1.1) as a semiparametric conditional moment model (Chamberlain (1987); Newey (1988)) for the *pseudo-observations* $\mathrm{I}\left(Y_i \preccurlyeq Y_j\right)$:

$$\begin{aligned} \mathrm{E}\{\mathrm{I}\left(Y_i \preccurlyeq Y_j\right) \mid \mathbf{X}_i, \mathbf{X}_j\} \;&=\; m(\mathbf{X}_i, \mathbf{X}_j; \boldsymbol{\beta}_0), \\ \text{with}\;\; \mathrm{I}\left(Y_i \preccurlyeq Y_j\right) &:= \mathrm{I}\left(Y_i < Y_j\right) + 0.5\mathrm{I}\left(Y_i = Y_j\right), \end{aligned} \tag{1.2}$$

where $\mathrm{I}\left(\cdot\right)$ denotes the ordinary indicator function, such that $\mathrm{I}\left(A\right) = 1$ if A is true, and zero otherwise. Thas et al. (2012) propose a semiparametric consistent estimator of $\boldsymbol{\beta}_0$ by mimicking quasi-likelihood estimating equations with an independence working correlation matrix. However, several authors have noted that this heuristic estimator is not necessarily semiparametrically efficient under

a correct specification of (1.1) (Van Keilegom (2012); Leng and Cheng (2012); Oja (2012)). This is because the pseudo observations are *cross-correlated*. For example, $\mathrm{I}(Y_i \preccurlyeq Y_j)$ and $\mathrm{I}(Y_i \preccurlyeq Y_k)$ are dependent, because they share the outcome Y_i. The potential inefficiency is a consequence of the limitations of the estimation theory of Thas et al. (2012), which only allows for an independence working correlation matrix ignoring the cross-correlation of the pseudo observations.

We therefore develop a more general estimation theory. Specifically, we derive the class of all consistent and asymptotically normal estimators of $\boldsymbol{\beta}$ in the semiparametric model induced by (1.1) by appealing to the theory of semiparametrics, and identify the efficient influence function of $\boldsymbol{\beta}$ using a Hoeffding decomposition (Newey (1990); Tsiatis (2006)). Next, we propose estimating equations based on the efficient influence function, the solution of which is equal to a locally efficient estimator of $\boldsymbol{\beta}$. A semiparametric locally efficient estimator of $\boldsymbol{\beta}$ is obtained by exploiting the relationship between PIMs and semiparametric transformation models (Cheng, Wei and Ying (1995)), allowing us to empirically evaluate the efficient estimator under a variety of scenarios. A second estimator is proposed that reduces the second-order bias of the estimator of Thas et al. (2012), where the latter is referred to as the standard estimator. Because the standard, efficient, and bias-reduced estimators have a computation complexity of at least $O(n^2)$, we propose computationally more convenient, but asymptotically equivalent variants of these estimators based on partitioning the data. In practice, this partition estimator is especially useful for the standard estimator.

The remainder of the paper is organized as follows. In Section 2, we present the main results of the estimation theory, identify the efficient influence function, and construct a locally efficient estimator. We also discuss the computational complexity of these estimators. In Section 3, we study the efficiency and bias properties of this estimator for a variety of well-chosen data-generating models, and in Section 4 we illustrate the methodology using a case study. In Section 5, we discuss our main results.

2. Estimation Theory

We denote the semiparametric model imposed by (1.1) by $\mathcal{M}_{\mathrm{PIM}}$, which is the set of all joint density functions $f_{\mathbf{Z}}(\mathbf{z};\boldsymbol{\beta},\boldsymbol{\eta}) = f_{Y|\mathbf{X}}(y \mid \mathbf{x};\boldsymbol{\beta},\boldsymbol{\eta}_1) f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\eta}_2)$ with $\mathbf{z}^T = (y, \mathbf{x}^T)$ obeying (1.1), with $\boldsymbol{\beta}$ the p-dimensional parameter of interest and $\boldsymbol{\eta} = (\boldsymbol{\eta}_1^T, \boldsymbol{\eta}_2^T)^T$ a (possibly) infinite-dimensional vector of variation-independent nuisance parameters. Let the data-generating law be $f_0(\mathbf{z}) = f_{\mathbf{Z}}(\mathbf{z};\boldsymbol{\beta}_0,\boldsymbol{\eta}_0) = f_{Y|\mathbf{X}}(y \mid \mathbf{x};\boldsymbol{\beta}_0,\boldsymbol{\eta}_{10}) f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\eta}_{20})$.

For example, a logistic PIM can be formulated as $\text{logit}\{\text{P}(Y_i \preccurlyeq Y_j \mid \mathbf{X}_i, \mathbf{X}_j)\} = (\mathbf{X}_j - \mathbf{X}_i)^T\boldsymbol{\beta}_0$, where $\text{logit}(x) = \log\{x/(1-x)\}$. The standard estimator of $\boldsymbol{\beta}_0$ in this model solves the estimating equation

$$\sum_{i=1}^{n}\sum_{j=1}^{n}(\mathbf{X}_j - \mathbf{X}_i)[\text{I}_{ij} - \text{expit}\{(\mathbf{X}_j - \mathbf{X}_i)^T\boldsymbol{\beta}\}] = \mathbf{0}, \tag{2.1}$$

where $\text{I}_{ij} = \text{I}(Y_i \preccurlyeq Y_j)$ and $\text{expit}(x) = 1/(1+e^{-x})$. Alternatively, a PIM with a probit link can be formulated as $\Phi^{-1}\{\text{P}(Y_i \preccurlyeq Y_j \mid \mathbf{X}_i, \mathbf{X}_j)\} = (\mathbf{X}_j - \mathbf{X}_i)^T\boldsymbol{\beta}_0$, where $\Phi^{-1}(\cdot)$ is the inverse cumulative standard normal distribution function. In this case, the standard estimator of $\boldsymbol{\beta}_0$ solves the estimating equation

$$\sum_{i=1}^{n}\sum_{j=1}^{n}(\mathbf{X}_j - \mathbf{X}_i)\frac{\phi\{(\mathbf{X}_j - \mathbf{X}_i)^T\boldsymbol{\beta}\}}{\Phi\{(\mathbf{X}_j - \mathbf{X}_i)^T\boldsymbol{\beta}\}[1 - \Phi\{(\mathbf{X}_j - \mathbf{X}_i)^T\boldsymbol{\beta}\}]}[\text{I}_{ij} - \Phi\{(\mathbf{X}_j - \mathbf{X}_i)^T\boldsymbol{\beta}\}] = \mathbf{0}, \tag{2.2}$$

and $\phi(\cdot)$ is the standard normal density function.

The solutions to (2.1) and (2.2) ignore the cross-correlation of the pseudo observations. Hence they may fail to exploit all available information in the data, and therefore fail to be efficient. To overcome this, we appeal to the theory of semiparametrics to derive the set of all unbiased estimating functions for $\boldsymbol{\beta}$ in model $\mathcal{M}_{\text{PIM}}$ (up to asymptotic equivalence) and to identify the estimating function that leads to the (locally) most efficient estimator.

The set of all unbiased estimating functions for the p-dimensional parameter $\boldsymbol{\beta}$ under model $\mathcal{M}_{\text{PIM}}$ is obtained by using the relationship between regular *asymptotically linear* (RAL) estimators and the geometry of *influence functions* (Newey (1990); Tsiatis (2006)). Specifically, an estimator $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ based on the i.i.d. data $\{\mathbf{Z}_i^T = (Y_i, \mathbf{X}_i^T) : i = 1, \ldots, n\}$ is said to be asymptotically linear under model $\mathcal{M}_{\text{PIM}}$ if it obeys the expansion

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\varphi}(Y_i, \mathbf{X}_i; \boldsymbol{\beta}_0, \boldsymbol{\eta}_0) + o_p(1) \tag{2.3}$$

for a p-dimensional function $\boldsymbol{\varphi}(\cdot; \boldsymbol{\beta}_0, \boldsymbol{\eta}_0)$ of the observed data satisfying the following moment conditions: (*i*) $\text{E}\{\boldsymbol{\varphi}(Y_i, \mathbf{X}_i; \boldsymbol{\beta}_0, \boldsymbol{\eta}_0)\} = \mathbf{0}$; (*ii*) $\text{E}\{\boldsymbol{\varphi}^T(Y_i, \mathbf{X}_i; \boldsymbol{\beta}_0, \boldsymbol{\eta}_0)\boldsymbol{\varphi}(Y_i, \mathbf{X}_i; \boldsymbol{\beta}_0, \boldsymbol{\eta}_0)\} < \infty$; and (*iii*) $\text{E}\{\boldsymbol{\varphi}(Y_i, \mathbf{X}_i; \boldsymbol{\beta}_0, \boldsymbol{\eta}_0)\boldsymbol{\varphi}^T(Y_i, \mathbf{X}_i; \boldsymbol{\beta}_0, \boldsymbol{\eta}_0)\}$ is nonsingular and $o_p(1)$ denotes a term that converges to zero in probability under the true data-generating law $f_0(\mathbf{z})$. This p-dimensional function $\boldsymbol{\varphi}(Y_i, \mathbf{X}_i; \boldsymbol{\beta}_0, \boldsymbol{\eta}_0)$ is referred to as the ith influence function ($i = 1, \ldots, n$) of the RAL estimator

$\widehat{\boldsymbol{\beta}}$, which is consistent and asymptotically normal with an asymptotic variance of $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ given by $\mathrm{E}\{\boldsymbol{\varphi}(Y, \mathbf{X}; \boldsymbol{\beta}_0, \boldsymbol{\eta}_0)\boldsymbol{\varphi}^T(Y, \mathbf{X}; \boldsymbol{\beta}_0, \boldsymbol{\eta}_0)\}$, the variance of the influence function. We restrict our development to regular estimators, thereby excluding estimators that have undesirable local properties (Newey (1990)), such as super-efficiency. Because the influence function of a RAL estimator $\widehat{\boldsymbol{\beta}}$ is asymptotically uniquely determined, it fully describes the first-order asymptotic behavior of the estimator $\widehat{\boldsymbol{\beta}}$. We therefore focus on identifying the set of all such influence functions from which we can subsequently construct unbiased estimating functions for $\boldsymbol{\beta}_0$. Next, we identify the influence function with the *smallest* variance, from which a *(locally) efficient* RAL estimator can be constructed.

Theorem 1 gives the set of all unbiased estimating functions for $\boldsymbol{\beta}_0$ in model $\mathcal{M}_{\mathrm{PIM}}$, delivering the class of all RAL estimators of $\boldsymbol{\beta}_0$ in model $\mathcal{M}_{\mathrm{PIM}}$ (up to asymptotic equivalence), together with their corresponding influence function and asymptotic distribution. The proof of Theorem 1 is given in Section 1 and Section 2 of the Supplementary Material.

Theorem 1. *If $\widehat{\boldsymbol{\beta}}$ is a RAL estimator of $\boldsymbol{\beta}_0$ in model $\mathcal{M}_{\mathrm{PIM}}$, then there exists a p-dimensional function $\mathbf{B}_{ij}(\boldsymbol{\beta}) = \mathbf{b}(\mathbf{X}_i, \mathbf{X}_j; \boldsymbol{\beta})$ in the set $\mathcal{B}$ of antisymmetric p-dimensional functions of $\mathbf{X}_i$ and $\mathbf{X}_j$ (satisfying $\mathbf{b}(\mathbf{X}_i, \mathbf{X}_j; \boldsymbol{\beta}) + \mathbf{b}(\mathbf{X}_j, \mathbf{X}_i; \boldsymbol{\beta}) = \mathbf{0}$) such that $\widehat{\boldsymbol{\beta}}$ is asymptotically equivalent to the solution of the estimating equation*

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{U}_{ij}(\boldsymbol{\beta}) = \mathbf{0}, \quad \textit{with} \quad \mathbf{U}_{ij}(\boldsymbol{\beta}) = \mathbf{B}_{ij}(\boldsymbol{\beta})\{\mathrm{I}_{ij} - M_{ij}(\boldsymbol{\beta})\}, \tag{2.4}$$

$\mathrm{I}_{ij} = \mathrm{I}(Y_i \preccurlyeq Y_j)$, and $M_{ij}(\boldsymbol{\beta}) = m(\mathbf{X}_i, \mathbf{X}_j; \boldsymbol{\beta})$. Under suitable smoothness and regularity conditions (listed in Section 1 and Section 2 of the Supplementary Material), the estimator $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}_0$ obeys expansion (2.3) with the influence function $\boldsymbol{\varphi}(Y_i, \mathbf{X}_i; \boldsymbol{\beta}_0, \boldsymbol{\eta}_0) = \boldsymbol{C}_0\mathrm{E}\{\mathbf{U}_{ij}(\boldsymbol{\beta}_0) \,|\, Y_i, \mathbf{X}_i\}$ and normalization constant $\boldsymbol{C}_0 = -2\mathbf{J}(\boldsymbol{\beta}_0)^{-1}$, with $\mathbf{J}(\boldsymbol{\beta}_0) = \mathrm{E}\{\partial\mathbf{U}_{ij}(\boldsymbol{\beta})/\partial\boldsymbol{\beta}^T|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\}$. It follows that $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_0)$ with the variance-covariance matrix $\boldsymbol{\Sigma}_0 = 4\mathbf{J}(\boldsymbol{\beta}_0)^{-1}\mathrm{cov}[\mathrm{E}\{\mathbf{U}_{ij}(\boldsymbol{\beta}_0) \mid Y_i, \mathbf{X}_i\}]\mathbf{J}(\boldsymbol{\beta}_0)^{-T}$.

A consistent estimator for the asymptotic variance $\boldsymbol{\Sigma}_0$ can be obtained using the sandwich formula: $\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\beta}}) = 4\widehat{\mathbf{J}}(\widehat{\boldsymbol{\beta}})^{-1}\widehat{\mathbf{K}}(\widehat{\boldsymbol{\beta}})\widehat{\mathbf{J}}(\widehat{\boldsymbol{\beta}})^{-T}$, with

$$\widehat{\mathbf{J}}(\widehat{\boldsymbol{\beta}}) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial\mathbf{U}_{ij}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^T}\Big|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}}$$

$$\text{and} \quad \widehat{\mathbf{K}}(\widehat{\boldsymbol{\beta}}) = \frac{1}{n}\sum_{i=1}^{n}\bar{\mathbf{U}}_i(\widehat{\boldsymbol{\beta}})\bar{\mathbf{U}}_i^T(\widehat{\boldsymbol{\beta}}), \quad \bar{\mathbf{U}}_i(\widehat{\boldsymbol{\beta}}) = \frac{1}{n}\sum_{j=1}^{n}\mathbf{U}_{ij}(\widehat{\boldsymbol{\beta}}).$$

2.1. The standard estimator of Thas et al. (2012)

We denote the estimator of Thas et al. (2012) as $\widehat{\boldsymbol{\beta}}^{\mathrm{ST}}$, and refer to it as the *standard estimator*. Before we identify the element of $\mathcal{B}$ that corresponds to the most efficient estimator, we show how the estimating equations proposed in Thas et al. (2012) are a special case of (2.4). In particular, their equations (2.1) and (2.2) are obtained by choosing $\mathbf{B}_{ij}(\boldsymbol{\beta}) = \mathbf{B}_{ij}^{\mathrm{ST}}(\boldsymbol{\beta}) = \mathbf{b}^{\mathrm{ST}}(\mathbf{X}_i, \mathbf{X}_j; \boldsymbol{\beta}) = \{\partial M_{ij}(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\}/[M_{ij}(\boldsymbol{\beta})\{1 - M_{ij}(\boldsymbol{\beta})\}]$, where the denominator corresponds to the conditional variance of the pseudo observations I_{ij}, given the covariates $\mathbf{X}_i$ and $\mathbf{X}_j$, mimicking quasi-likelihood estimating equations with an independence working correlation matrix. For the logistic PIM, (2.1) corresponds to $\mathbf{B}_{ij}^{\mathrm{ST}}(\boldsymbol{\beta}) = (\mathbf{X}_j - \mathbf{X}_i)$, and for the probit PIM, (2.2) corresponds to the choice $\mathbf{B}_{ij}^{\mathrm{ST}}(\boldsymbol{\beta}) = (\mathbf{X}_j - \mathbf{X}_i)\phi\{(\mathbf{X}_j - \mathbf{X}_i)^T\boldsymbol{\beta}\}(\Phi\{(\mathbf{X}_j - \mathbf{X}_i)^T\boldsymbol{\beta}\}[1 - \Phi\{(\mathbf{X}_j - \mathbf{X}_i)^T\boldsymbol{\beta}\}])^{-1}$. Thus $\mathbf{B}_{ij}^{\mathrm{ST}}(\boldsymbol{\beta}) + \mathbf{B}_{ji}^{\mathrm{ST}}(\boldsymbol{\beta}) = \mathbf{0}$. Because these estimating equations ignore the cross-correlation structure of the transformed outcomes $\mathrm{I}(Y_i \preccurlyeq Y_j)$, they may not deliver an efficient estimator.

2.2. The locally efficient estimator

Different choices of the function $\mathbf{B}_{ij}(\boldsymbol{\beta})$ result in RAL estimators with different asymptotic variances. Theorem 2 identifies the choice $\mathbf{B}_{ij}^{\mathrm{EFF}}(\boldsymbol{\beta}) = \mathbf{b}^{\mathrm{EFF}}(\mathbf{X}_i, \mathbf{X}_j; \boldsymbol{\beta})$ that results in an efficient RAL estimator $\widehat{\boldsymbol{\beta}}^{\mathrm{EFF}}$ under model $\mathcal{M}_{\mathrm{PIM}}$. The proof of Theorem 2 is outlined below; while a detailed proof is given in Section 3 of the Supplementary Material.

Theorem 2. *The efficient estimator* $\widehat{\boldsymbol{\beta}}^{\mathrm{EFF}}$ *is obtained by choosing* $\mathbf{B}_{ij}(\boldsymbol{\beta}) = \mathbf{B}_{ij}^{\mathrm{EFF}}(\boldsymbol{\beta})$ *in* (2.4)*, where* $\mathbf{B}_{ij}^{\mathrm{EFF}}(\boldsymbol{\beta})$ *is the solution to the integral equation*

$$\mathbf{D}_{ij}(\boldsymbol{\beta}_0) = \mathrm{E}\left\{\mathbf{B}_{ik}^{\mathrm{EFF}}(\boldsymbol{\beta}_0)V_{ijik}(\boldsymbol{\beta}_0) + \mathbf{B}_{jk}^{\mathrm{EFF}}(\boldsymbol{\beta}_0)V_{ijjk}(\boldsymbol{\beta}_0)\Big|\mathbf{X}_i, \mathbf{X}_j\right\}, \quad i \neq k \text{ and } j \neq k, \tag{2.5}$$

with $\mathbf{D}_{ij}(\boldsymbol{\beta}_0) = \partial M_{ij}(\boldsymbol{\beta})/\partial\boldsymbol{\beta}|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$ *and with conditional covariance*

$$V_{ijkl}(\boldsymbol{\beta}_0) = V(\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k, \mathbf{X}_l; \boldsymbol{\beta}_0) = \mathrm{cov}(\mathrm{I}_{ij}, \mathrm{I}_{kl}|\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k, \mathbf{X}_l).$$

Proof. The semiparametric nuisance tangent space Λ of model $\mathcal{M}_{\mathrm{PIM}}$ is equal to

$$\Lambda = \{\mathbf{s}(Y, \mathbf{X}) \in \mathcal{H} \,|\, \mathrm{E}[\{\mathbf{s}(Y, \mathbf{X}) + \mathbf{s}(Y^*, \mathbf{X}^*)\}\{\mathrm{I}(Y \preccurlyeq Y^*) - m(\mathbf{X}, \mathbf{X}^*; \boldsymbol{\beta}_0)\}|\mathbf{X}, \mathbf{X}^*] = \mathbf{0}\}, \tag{2.6}$$

for $(Y, \mathbf{X}^T) \perp\!\!\!\perp (Y^*, \mathbf{X}^{*,T})$, and $\mathcal{H}$ is the Hilbert space equipped with the covariance inner product of p-dimensional mean-zero and square-integrable measurable random functions $\mathbf{h}(Y, \mathbf{X})$. The orthogonal complement of the semiparametric nuisance tangent space $\Lambda^\perp$ is equal to

$$\Lambda^\perp = \{\mathbf{s}^\perp(Y, \mathbf{X}) \in \mathcal{H} \mid \mathbf{s}^\perp(Y, \mathbf{X}) = \mathrm{E}[\mathbf{b}(\mathbf{X}, \mathbf{X}^*)\{\mathrm{I}(Y \preccurlyeq Y^*) - m(\mathbf{X}, \mathbf{X}^*; \boldsymbol{\beta}_0)\} \mid Y, \mathbf{X}], \mathbf{b}(\mathbf{X}, \mathbf{X}^*) \in \mathcal{B}\}, \tag{2.7}$$

for $(Y, \mathbf{X}^T) \perp\!\!\!\perp (Y^*, \mathbf{X}^{*,T})$, where $\mathcal{B} = \{\mathbf{b}(\mathbf{X}, \mathbf{X}^*) \mid \mathbf{b}(\mathbf{X}, \mathbf{X}^*)$ is square integrable and $\mathbf{b}(\mathbf{X}, \mathbf{X}^*) + \mathbf{b}(\mathbf{X}^*, \mathbf{X}) = \mathbf{0}\}$. The score function for $\boldsymbol{\beta}$ is equal to $\mathbf{s}_{\boldsymbol{\beta}}(Y, \mathbf{X}; \boldsymbol{\beta}_0, \boldsymbol{\eta}_0) = \partial \log f_{Y\mathbf{X}}(y, \mathbf{x}; \boldsymbol{\beta}, \boldsymbol{\eta}_0)/\partial\boldsymbol{\beta}|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$, and the efficient influence function is given by

$$\boldsymbol{\varphi}^{\mathrm{EFF}}(Y, \mathbf{X}; \boldsymbol{\beta}_0, \boldsymbol{\eta}) = \mathrm{E}\{\mathbf{s}^{\mathrm{EFF}}(Y, \mathbf{X}; \boldsymbol{\beta}_0, \boldsymbol{\eta}_0)\mathbf{s}^{\mathrm{EFF},T}(Y, \mathbf{X}; \boldsymbol{\beta}_0, \boldsymbol{\eta}_0)\}^{-1}\mathbf{s}^{\mathrm{EFF}}(Y, \mathbf{X}; \boldsymbol{\beta}_0, \boldsymbol{\eta}_0),$$

with an efficient score $\mathbf{s}^{\mathrm{EFF}}(Y, \mathbf{X}; \boldsymbol{\beta}_0, \boldsymbol{\eta}_0)$ equal to the orthogonal projection of the score function onto the complement of the nuisance tangent space. In order to find the efficient score, we thus need to find the function $\mathbf{b}^{\mathrm{EFF}}(\mathbf{X}, \mathbf{X}^*) \in \mathcal{B}$ such that $\mathbf{s}^{\mathrm{EFF}}(Y, \mathbf{X}; \boldsymbol{\beta}_0, \boldsymbol{\eta}_0) = \mathrm{E}[\mathbf{b}^{\mathrm{EFF}}(\mathbf{X}, \mathbf{X}^*)\{\mathrm{I}(Y \preccurlyeq Y^*) - m(\mathbf{X}, \mathbf{X}^*; \boldsymbol{\beta}_0)\} \mid Y, \mathbf{X}]$. This means solving the integral equation

$$\begin{aligned}
&\mathrm{E}[\{\mathbf{s}_{\boldsymbol{\beta}}(Y, \mathbf{X}; \boldsymbol{\beta}_0, \boldsymbol{\eta}_0) + \mathbf{s}_{\boldsymbol{\beta}}(Y^*, \mathbf{X}^*; \boldsymbol{\beta}_0, \boldsymbol{\eta}_0)\}\{\mathrm{I}(Y \preccurlyeq Y^*) - m(\mathbf{X}, \mathbf{X}^*; \boldsymbol{\beta}_0)\}|\mathbf{X}, \mathbf{X}^*] \\
&= \mathrm{E}([\mathbf{b}^{\mathrm{EFF}}(\mathbf{X}, \mathbf{X}^\dagger)\{\mathrm{I}(Y \preccurlyeq Y^\dagger) - m(\mathbf{X}, \mathbf{X}^\dagger; \boldsymbol{\beta}_0)\} + \mathbf{b}^{\mathrm{EFF}}(\mathbf{X}^*, \mathbf{X}^\dagger)\{\mathrm{I}(Y^* \preccurlyeq Y^\dagger) \\
&\quad - m(\mathbf{X}^*, \mathbf{X}^\dagger; \boldsymbol{\beta}_0)\}] \times \{\mathrm{I}(Y \preccurlyeq Y^*) - m(\mathbf{X}, \mathbf{X}^*; \boldsymbol{\beta}_0)\}|\mathbf{X}, \mathbf{X}^*),
\end{aligned}$$

from which expression (2.5) follows, because $\mathrm{E}[\{\mathbf{s}^T_{\boldsymbol{\beta}}(Y, \mathbf{X}; \boldsymbol{\beta}_0, \boldsymbol{\eta}_0) + \mathbf{s}^T_{\boldsymbol{\beta}}(Y^*, \mathbf{X}^*; \boldsymbol{\beta}_0, \boldsymbol{\eta}_0)\}\{\mathrm{I}(Y \preccurlyeq Y^*) - m(\mathbf{X}, \mathbf{X}^*; \boldsymbol{\beta}_0)\} \mid \mathbf{X}, \mathbf{X}^*] = \partial m(\mathbf{X}, \mathbf{X}^*; \boldsymbol{\beta})/\partial\boldsymbol{\beta}^T|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$ for an $\mathcal{M}_{\mathrm{PIM}}$.

Unfortunately, the integral equation (2.5) does not admit a closed-form solution for the function $\mathbf{B}^{\mathrm{EFF}}_{ij}(\boldsymbol{\beta})$, especially when the conditional covariances $V_{ijkl}(\boldsymbol{\beta})$ depend on $\boldsymbol{\beta}$. It therefore needs to be solved numerically using computationally demanding iterative procedures. In doing so, we replace the expectation in (2.5) with its empirical counterpart, where the average is taken over $k = 1, \ldots, n$. Specifically, for a fixed $\boldsymbol{\beta}$, we approximate $\mathbf{B}^{\mathrm{EFF}}_{ij}(\boldsymbol{\beta})$ by $\widehat{\mathbf{B}}^{\mathrm{EFF}}_{ij}(\boldsymbol{\beta})$,

where $\widehat{\mathbf{B}}_{ij}^{\mathrm{EFF}}(\boldsymbol{\beta})$ solves the linear system of equations

$$\mathbf{D}_{ij}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{k=1}^{n}\left\{\widehat{\mathbf{B}}_{ik}^{\mathrm{EFF}}(\boldsymbol{\beta})V_{ijik}(\boldsymbol{\beta}) + \widehat{\mathbf{B}}_{jk}^{\mathrm{EFF}}(\boldsymbol{\beta})V_{ijjk}(\boldsymbol{\beta})\right\}, \quad i,j = 1,\ldots,n. \quad (2.8)$$

From their defining properties, it follows that for arbitrary $k,\ell \in \{1,\ldots,n\}$, we have that $\mathbf{D}_{k\ell}(\boldsymbol{\beta}) = -\mathbf{D}_{\ell k}(\boldsymbol{\beta})$, $\widehat{\mathbf{B}}_{k\ell}^{\mathrm{EFF}}(\boldsymbol{\beta}) = -\widehat{\mathbf{B}}_{\ell k}^{\mathrm{EFF}}(\boldsymbol{\beta})$, and $V_{ijk\ell}(\boldsymbol{\beta}) = -V_{ij\ell k}(\boldsymbol{\beta}) = V_{k\ell ij}(\boldsymbol{\beta})$. In particular, this implies that for all $i \in \{1,\ldots,n\}$, $\mathbf{D}_{ii}(\boldsymbol{\beta}) = \mathbf{0}$, $\mathbf{B}_{ii}^{\mathrm{EFF}}(\boldsymbol{\beta}) = \widehat{\mathbf{B}}_{ii}^{\mathrm{EFF}}(\boldsymbol{\beta}) = \mathbf{0}$ and $V_{iik\ell}(\boldsymbol{\beta}) = 0$. We conclude that equations for which $i = j$ do not contribute to the system. The antisymmetry conditions additionally guarantee that the equation $\mathbf{D}_{ij}(\boldsymbol{\beta}) = n^{-1}\sum_{k=1}^{n}\{\widehat{\mathbf{B}}_{ik}^{\mathrm{EFF}}(\boldsymbol{\beta})V_{ijik}(\boldsymbol{\beta}) + \widehat{\mathbf{B}}_{jk}^{\mathrm{EFF}}(\boldsymbol{\beta})V_{ijjk}(\boldsymbol{\beta})\}$ is equivalent to the equation

$$\mathbf{D}_{ji}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{k=1}^{n}\left\{\widehat{\mathbf{B}}_{jk}^{\mathrm{EFF}}(\boldsymbol{\beta})V_{jijk}(\boldsymbol{\beta}) + \widehat{\mathbf{B}}_{ik}^{\mathrm{EFF}}(\boldsymbol{\beta})V_{jiik}(\boldsymbol{\beta})\right\},$$

and the linear system of equations consequently reduces to those $n(n-1)/2$ equations with $i < j$. When $\{i,j\} \cap \{k,\ell\} = \emptyset$, the pseudo observations I_{ij} and $\mathrm{I}_{k\ell}$ are uncorrelated, such that $V_{ijk\ell}(\boldsymbol{\beta}) = 0$. Using this, we conclude that the linear system of n^2 equations (2.8) is equivalent to the linear system of $n(n-1)/2$ equations

$$\mathbf{D}_{ij}(\boldsymbol{\beta}) = \frac{1}{n}\left[\sum_{k=1}^{n-1}\sum_{\ell=k+1}^{n}\left\{\widehat{\mathbf{B}}_{k\ell}^{\mathrm{EFF}}(\boldsymbol{\beta})V_{ijk\ell}(\boldsymbol{\beta})\right\} + \widehat{\mathbf{B}}_{ij}^{\mathrm{EFF}}(\boldsymbol{\beta})V_{ijij}(\boldsymbol{\beta})\right], \quad i < j. \quad (2.9)$$

A detailed calculation of this last step is provided in the Supplementary Material (Section 4). Now, define the $[n(n-1)/2 \times p]$-dimensional matrices $\mathbf{D}(\boldsymbol{\beta})$ and $\widehat{\mathbf{B}}^{\mathrm{EFF}}(\boldsymbol{\beta})$ such that the $[(i-1)(2n-i)/2+j-i]$th row corresponds to $\mathbf{D}_{ij}^{T}(\boldsymbol{\beta})$ and $\widehat{\mathbf{B}}_{ij}^{\mathrm{EFF},T}(\boldsymbol{\beta})$, respectively. Next, define the $[n(n-1)/2 \times n(n-1)/2]$-dimensional matrix $\mathbf{V}(\boldsymbol{\beta})$ such that $V_{ijk\ell}(\boldsymbol{\beta})$ is on the $[(i-1)(2n-i)/2+j-i]$th row and $[(k-1)(2n-k)/2+\ell-k]$th column. Finally, define the $[n(n-1)/2 \times n(n-1)/2]$-dimensional diagonal matrix $\mathbf{V}_{\mathrm{indep}}(\boldsymbol{\beta})$ such that the $[(i-1)(2n-i)/2+j-i]$th diagonal element is equal to $V_{ijij}(\boldsymbol{\beta})$. Using this notation, (2.9) can be written as the matrix equation $n\mathbf{D}(\boldsymbol{\beta}) = \widehat{\mathbf{B}}^{\mathrm{EFF},T}(\boldsymbol{\beta})\{\mathbf{V}(\boldsymbol{\beta}) + \mathbf{V}_{\mathrm{indep}}(\boldsymbol{\beta})\}$. It follows that

$$\widehat{\mathbf{B}}^{\mathrm{EFF}}(\boldsymbol{\beta}) = n\mathbf{D}^{T}(\boldsymbol{\beta})\{\mathbf{V}(\boldsymbol{\beta}) + \mathbf{V}_{\mathrm{indep}}(\boldsymbol{\beta})\}^{-1}. \quad (2.10)$$

A semiparametric efficient estimator $\widehat{\boldsymbol{\beta}}^{\mathrm{EFF}}$ can then be obtained by iteratively

solving the estimating equation $\widehat{\mathbf{B}}^{\mathrm{EFF},T}(\boldsymbol{\beta})\{\mathbf{I}-\mathbf{M}(\boldsymbol{\beta})\}=\mathbf{0}$, with $\mathbf{I}$ and $\mathbf{M}(\boldsymbol{\beta})$ both $n(n-1)/2$-dimensional vectors such that the $[(i-1)(2n-i)/2+j-i]$th element is given by I_{ij} and $M_{ij}(\boldsymbol{\beta})$, respectively.

Remark 1. Theorem 1 does not cover the setting in which $\widehat{\mathbf{B}}(\boldsymbol{\beta})$ is replaced by an estimator. However, doing so does not affect the asymptotic distribution of the resulting estimator of $\boldsymbol{\beta}$ because, by construction, the influence function of the estimator considered is orthogonal to the nuisance parameter space Tsiatis (2006). This implies that its asymptotic behavior is the same, regardless of whether it is evaluated at the true nuisance parameters or the estimated nuisance parameters, which converge at a rate faster than $n^{1/4}$ that of the truth.

Remark 2. Let $\mathbf{B}^{\mathrm{ST}}(\boldsymbol{\beta})$ denote the $[n(n-1)/2\times p]$ dimensional matrix with the $[(i-1)(2n-i)/2+j-i]$th row equal to $\mathbf{B}_{ij}^{\mathrm{ST},T}(\boldsymbol{\beta})$. Using the above notation, we have that $\mathbf{B}^{\mathrm{ST}}(\boldsymbol{\beta})=\mathbf{D}^{T}(\boldsymbol{\beta})\mathbf{V}_{\mathrm{indep}}^{-1}(\boldsymbol{\beta})$. It follows that $\widehat{\boldsymbol{\beta}}^{\mathrm{ST}}$ solves the estimating equation $\mathbf{D}^{T}(\boldsymbol{\beta})\mathbf{V}_{\mathrm{indep}}^{-1}(\boldsymbol{\beta})\{\mathbf{I}-\mathbf{M}(\boldsymbol{\beta})\}=\mathbf{0}$. That this estimator ignores the cross-correlation between the pseudo observations is clearly shown here, because $\mathbf{B}^{\mathrm{ST}}(\boldsymbol{\beta})$ can be obtained from $\widehat{\mathbf{B}}^{\mathrm{EFF}}(\boldsymbol{\beta})$ by forcing $\mathbf{V}(\boldsymbol{\beta})$ to be zero.

Remark 2 illustrates how the complex cross-correlation structure of the pseudo observations is taken into account by the semiparametric efficient estimator $\widehat{\boldsymbol{\beta}}^{\mathrm{EFF}}$. It is the $[n(n-1)/2\times n(n-1)/2]$-dimensional matrix $\mathbf{V}(\boldsymbol{\beta})$ in equation (2.10), consisting of the elements $V_{ijkl}(\boldsymbol{\beta})$ (the correlations between the pseudo observations), that shows how the information contained within these correlation coefficients is exploited by the semiparametric efficient estimator.

We still need to address one peculiarity. When solving the integral equation (2.5) numerically, we need reasonable estimators for the covariances $V_{ijkl}(\boldsymbol{\beta})$, because a nonparametric estimation is unstable or even unfeasible in small samples, owing to the curse of dimensionality (Robins and Ritov (1997)). We therefore need to impose an additional working model, which we show for a continuous outcome. In this case, the resulting estimator $\widehat{\boldsymbol{\beta}}^{\mathrm{EFF}}$ is only *locally efficient* and not globally efficient. That is, the semiparametric efficiency bound is only attained under a correctly specified working model for the covariance structure, but not necessarily otherwise. Under a misspecification of the working model, the consistency of the estimator is maintained.

When the outcome Y is continuous, in which case the conditional probabilistic index satisfies $\mathrm{P}\left(Y_i \preccurlyeq Y_j \mid \mathbf{X}_i,\mathbf{X}_j\right)=\mathrm{P}\left(Y_i<Y_j \mid \mathbf{X}_i,\mathbf{X}_j\right)$, the conditional covariance can be written as

$$V_{ijik}(\boldsymbol{\beta})=\mathrm{P}\left(Y_i<\min(Y_j,Y_k) \mid \mathbf{X}_i,\mathbf{X}_j,\mathbf{X}_k\right)-M_{ij}(\boldsymbol{\beta})M_{ik}(\boldsymbol{\beta})$$

Table 1. Relationship between the semiparametric transformation model and the probabilistic index model for two choices of error distribution F_ε.

Semiparametric Transformation Model $H(Y_i) = \mathbf{X}_i^T\boldsymbol{\alpha} + \varepsilon_i$	Probabilistic Index Model $\mathrm{P}(Y_i < Y_j \mid \mathbf{X}_i, \mathbf{X}_j) = g^{-1}\{(\mathbf{X}_j - \mathbf{X}_i)^T\boldsymbol{\beta}\}$
Normal error $F_\varepsilon(a) = \Phi(a)$	$g^{-1}(a) = \Phi(a)$ and $\boldsymbol{\beta} = \boldsymbol{\alpha}/\sqrt{2}$
Gumbel error $F_\varepsilon(a) = \exp\{-\exp(-a)\}$	$g^{-1}(a) = \mathrm{expit}(a)$ and $\boldsymbol{\beta} = \boldsymbol{\alpha}$

and

$$V_{ijjk}(\boldsymbol{\beta}) = \mathrm{P}(Y_i < Y_j < Y_k \mid \mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k) - M_{ij}(\boldsymbol{\beta})M_{jk}(\boldsymbol{\beta}).$$

Consequently, to compute the efficient estimator, one needs to model probabilities of the form $\mathrm{P}(Y_i < \min(Y_j, Y_k) \mid \mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k)$ and $\mathrm{P}(Y_i < Y_j < Y_k \mid \mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k)$. One flexible way of doing so, is to assume a semiparametric transformation model (STM)

$$H(Y) = \mathbf{X}^T\boldsymbol{\alpha} + \varepsilon, \tag{2.11}$$

where ε is a zero-mean random error term with a known cumulative distribution function $F_\varepsilon(\cdot)$, and $H(\cdot)$ is an unspecified strictly increasing function (e.g., Cuzick (1988); Cheng, Wei and Ying (1995)). Under such a model, it follows that

$$\mathrm{P}(Y_i < Y_j \mid \mathbf{X}_i, \mathbf{X}_j) = f^{-1}\{(\mathbf{X}_j - \mathbf{X}_i)^T\boldsymbol{\alpha}\}, \tag{2.12}$$

with $f^{-1}(a) = \int F_\varepsilon(a + b)dF_\varepsilon(b)$. For example, if ε follows a standard normal distribution, then $F_\varepsilon(a) = \Phi(a)$. From this, it follows that $f^{-1}(a) = \Phi(a/\sqrt{2})$, resulting in a probit PIM with $\boldsymbol{\beta} = \boldsymbol{\alpha}/\sqrt{2}$. Alternatively, if ε follows a Gumbel distribution with location parameter zero and scale parameter one,, $F_\varepsilon(a) = \exp\{-\exp(-a)\}$, then $f^{-1}(a) = \mathrm{expit}(a)$, resulting in a logistic PIM with $\boldsymbol{\beta} = \boldsymbol{\alpha}$. Table 1 summarizes these relationships.

Furthermore for the semiparametric transformation model,

$$\mathrm{P}(Y_i < \min(Y_j, Y_k) \mid \mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k) = h_1^{-1}\{(\mathbf{X}_i - \mathbf{X}_j)^T\boldsymbol{\alpha}, (\mathbf{X}_i - \mathbf{X}_k)^T\boldsymbol{\alpha}\}, \tag{2.13}$$

$$\mathrm{P}(Y_i < Y_j < Y_k \mid \mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k) = h_2^{-1}\{(\mathbf{X}_j - \mathbf{X}_i)^T\boldsymbol{\alpha}, (\mathbf{X}_j - \mathbf{X}_k)^T\boldsymbol{\alpha}\}, \tag{2.14}$$

where $h_1^{-1}(a, b) = \int\{1 - F_\varepsilon(a+c)\}\{1 - F_\varepsilon(b+c)\}dF_\varepsilon(c)$ and $h_2^{-1}(a, b) = \int F_\varepsilon(a + c)\{1 - F_\varepsilon(b + c)\}dF_\varepsilon(c)$. The functions $h_1^{-1}(\cdot, \cdot)$ and $h_2^{-1}(\cdot, \cdot)$ can be obtained by numerical integration.

2.3. A bias-reduced estimator

Obtaining the locally efficient estimator $\widehat{\boldsymbol{\beta}}^{\mathrm{EFF}}$ is computationally expensive, especially when the covariances $V_{ijkl}(\boldsymbol{\beta})$ depend on the parameter $\boldsymbol{\beta}$. Here, we propose a simplification to address this problem. Instead of using $V_{ijkl}(\boldsymbol{\beta})$, where $\boldsymbol{\beta}$ is treated as a running parameter, we fix the value of $\boldsymbol{\beta}$ to a prespecified value $\boldsymbol{\beta}^*$. This results in the covariances $V_{ijkl}(\boldsymbol{\beta}^*)$ and the estimating equation $\sum_{i=1}^n \sum_{j=1}^n \mathbf{U}_{ij}^*(\boldsymbol{\beta}) = \mathbf{0}$, with $\mathbf{U}_{ij}^*(\boldsymbol{\beta}) = \widehat{\mathbf{B}}_{ij}^{*}(\boldsymbol{\beta})\{\mathrm{I}_{ij} - M_{ij}(\boldsymbol{\beta})\}$, in which $\widehat{\mathbf{B}}_{ij}^{*}(\boldsymbol{\beta})$ solves

$$\mathbf{D}_{ij}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{k=1}^{n}\left\{\widehat{\mathbf{B}}_{ik}^{*}(\boldsymbol{\beta})V_{ijik}(\boldsymbol{\beta}^*) + \widehat{\mathbf{B}}_{jk}^{*}(\boldsymbol{\beta})V_{ijjk}(\boldsymbol{\beta}^*)\right\}, \tag{2.15}$$

yielding $\widehat{\mathbf{B}}^{*}(\boldsymbol{\beta}) = n\mathbf{D}^T(\boldsymbol{\beta})\{\mathbf{V}(\boldsymbol{\beta}^*)+\mathbf{V}_{\mathrm{indep}}(\boldsymbol{\beta}^*)\}^{-1}$. The quantity $\widehat{\mathbf{B}}_{ij}^{*}(\boldsymbol{\beta})$ serves as the numerical approximation of $\mathbf{B}_{ij}^{*}(\boldsymbol{\beta})$, the solution to the integral equation (2.5), including the aforementioned simplification. This procedure requires that we invert the (model-based) variance-covariance matrix of the pseudo observations only once, in contrast to the calculation of the locally efficient estimator, which requires an inversion of this matrix in every step of the iterative procedure. This results in a computational gain.

The solution to $\sum_{i=1}^n \sum_{j=1}^n \mathbf{U}_{ij}^*(\boldsymbol{\beta}) = \mathbf{0}$ gives the estimator $\widehat{\boldsymbol{\beta}}^*$. Because this estimation procedure constitutes a special case of (2.4) (for any choice of $\boldsymbol{\beta}^*$), where the function $\mathbf{B}_{ij}(\boldsymbol{\beta})$ is set to $\widehat{\mathbf{B}}_{ij}^{*}(\boldsymbol{\beta})$, the estimator $\widehat{\boldsymbol{\beta}}^*$ is a consistent and asymptotically normal estimator of $\boldsymbol{\beta}$.

An important question that remains is how to choose the value $\boldsymbol{\beta}^*$. This value was selected to minimize the second-order finite-sample bias of $\hat{\boldsymbol{\beta}}^*$; see Proposition 1. A proof and regularity conditions are given in the Supplementary Material (Section 5).

Proposition 1. *Assume that the semiparametric transformation model* (2.11) *holds, and let* ε *follow a symmetric distribution about zero. Then, under regularity conditions, the second-order finite-sample bias of* $\widehat{\boldsymbol{\beta}}^*$ *is minimized at* $\boldsymbol{\beta}^* = \mathbf{0}$.

When we set $\boldsymbol{\beta}^* = \mathbf{0}$, we denote the estimator $\widehat{\boldsymbol{\beta}}^*$ by $\widehat{\boldsymbol{\beta}}^{\mathrm{BR}}$, the bias-reduced estimator (hence the superscript BR). When the true value $\boldsymbol{\beta}_0$ is equal to $\mathbf{0}$, the covariances $V_{ijkl}(\boldsymbol{\beta})$ are correctly modeled, given a correct specification of the STM. In this case (under the null), the estimator $\widehat{\boldsymbol{\beta}}^{\mathrm{BR}}$ is also semiparametric efficient under model $\mathcal{M}_{\mathrm{PIM}}$.

Remark 3. The working model (2.11) is parameterized by the nuisance parameter $\boldsymbol{\alpha}$, which is a function of the parameter of interest $\boldsymbol{\beta}$; say, $\boldsymbol{\alpha} = \mathbf{k}(\boldsymbol{\beta})$. For the locally efficient estimator, $\boldsymbol{\alpha}$ is treated as a running parameter, and for the

biased-reduced estimator, it is set to $\boldsymbol{\alpha} = \mathbf{0}$. Other estimators for $\boldsymbol{\beta}$ might be constructed by estimating $\boldsymbol{\alpha}$ in (2.11) directly, for example using the rank likelihood (Cuzick (1988)), and treating it as fixed while solving (2.4) for $\boldsymbol{\beta}$.

2.4. Computational issues and solutions

Before empirically studying the theoretical properties of the different estimators, we first focus on their computational properties. All three estimators require solving a system of equations of the form (2.4). The difference between the estimators lies in the computation of $\mathbf{B}_{ij}(\boldsymbol{\beta})$. Equation (2.10) gives an expression for $\widehat{\boldsymbol{\beta}}^{\mathrm{EFF}}$, and requires inverting a matrix of dimension $n(n-1)/2 \times n(n-1)/2$. When an iterative algorithm (e.g. the Newton method) is used to solve (2.4), this matrix has to be inverted at each iteration. The estimator $\widehat{\boldsymbol{\beta}}^{\mathrm{BR}}$ is computationally less expensive, because the inversion occurs only once and not at each iteration. Here $\mathbf{B}_{ij}(\boldsymbol{\beta}^*)$ is not a function of $\boldsymbol{\beta}$, but is instead kept fixed at $\boldsymbol{\beta}^* = \mathbf{0}$. Computationally, the estimator $\widehat{\boldsymbol{\beta}}^{\mathrm{ST}}$ is the least demanding of the three because it does not require inverting an $n(n-1)/2 \times n(n-1)/2$ matrix because $\mathbf{B}_{ij}(\boldsymbol{\beta})$ has a simple expression of the form $\mathbf{B}_{ij}(\boldsymbol{\beta}) = \{\partial M_{ij}(\boldsymbol{\beta})/\partial \boldsymbol{\beta}\}/[M_{ij}(\boldsymbol{\beta})\{1 - M_{ij}(\boldsymbol{\beta})\}]$.

See Section 6 of the Supplementary Material for the computation times of the three estimators for several sample sizes.

Despite its computational superiority, $\widehat{\boldsymbol{\beta}}^{\mathrm{ST}}$ is still computationally demanding because of the double summation in (2.4), which can be problematic for large n. We therefore propose a *partition estimator* that is computationally less demanding, but asymptotically equivalent.

For this purpose, we partition the data into k distinct parts S_i $(i = 1, \ldots, k)$ of size $|S_i| = m_i$, such that $\sum_{i=1}^{k} m_i = n$, $m_i \to \infty$ as $n \to \infty$ and $k \to \infty$ and $k/\sqrt{n} \to 0$ as $n \to \infty$. Let $\widehat{\boldsymbol{\beta}}_i$ denote any RAL estimator applied to part i of the data. The partition estimator is then given by

$$\widetilde{\boldsymbol{\beta}} := \frac{1}{n} \sum_{i=1}^{k} m_i \cdot \widehat{\boldsymbol{\beta}}_i.$$

Theorem 3 shows the first-order asymptotic equivalence of the partition estimator $\widetilde{\boldsymbol{\beta}}$ and the corresponding RAL estimator $\widehat{\boldsymbol{\beta}}$ applied to the entire data set without partitioning.

Theorem 3. *Under model $\mathcal{M}_{\mathrm{PIM}}$, the partition estimator $\widetilde{\boldsymbol{\beta}} = n^{-1} \sum_{i=1}^{k} m_i \cdot \widehat{\boldsymbol{\beta}}_i$ ($\widehat{\boldsymbol{\beta}}_i$ is a consistent estimator applied to part i of the data, for $i = 1, \ldots, k$) is a consistent estimator for $\boldsymbol{\beta}_0$. It further holds that $\sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o_p(1)$, with $\widehat{\boldsymbol{\beta}}$ the estimator applied to the entire data set without partitioning.*

A proof of Theorem 3 is given in Section 7 of the Supplementary Material. It thus follows from Theorem 3 that $\sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ and $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ have the same limiting distribution. A consistent estimator of the variance of $\widetilde{\boldsymbol{\beta}}$ is obtained from

$$\text{Var}\left(\widetilde{\boldsymbol{\beta}}\right) = \frac{1}{n^2}\sum_{i=1}^{k} m_i^2 \cdot \text{Var}\left(\widehat{\boldsymbol{\beta}}_i\right),$$

with $\text{Var}\left(\widehat{\boldsymbol{\beta}}_i\right)$ replaced by the sandwich estimator from Theorem 1.

In practice, we propose using $k = \lfloor n^{0.5-\delta} \rfloor$ with $0 < \delta < 0.5$ and $\lfloor x \rfloor$ the integer part of x, and partitioning the data into k groups such that $\max_{i=1,\ldots,k}(m_i) - \min_{i=1,\ldots,k}(m_i) \leq 1$. The partition estimator is then substantially faster, because it requires k estimates on a subset of size m_i instead of one estimate on the entire data set of size n. In practice, δ can be chosen to minimize the computational complexity.

3. Empirical Evaluation

To study the empirical performance of the estimators of Section 2, data are generated under the linear transformation model (2.11). The strictly increasing function $H(\cdot)$ is set as the identity function. Table 1 summarizes the relationship between the data-generating model and the PIM.

All simulations are performed in R (R Core Team (2018)); the R code is available at `goo.gl/UA4mFV`.

3.1. Normally distributed error

We start by considering a normally distributed error in model (2.11), which corresponds to a probit PIM. We consider a univariate covariate X that follows a discrete uniform distribution, with K support values spaced equally between (a, u). For the following simulation experiments, we set a to 0.1, K to 10, and evaluate different values of the upper limit u. By using a discrete covariate, equation (2.5) reduces to a summation that, following steps (2.8)–(2.10), allows a closed-form solution for $\mathbf{B}^{\text{EFF}}(\boldsymbol{\beta})$. In this way, we avoid having to approximate equation (2.5) so that differences in efficiency, if any, are due solely to the estimators themselves. We can study the range of the potential efficiency gain empirically by considering specific simulation scenarios in which small and larger gains in efficiency of $\widehat{\boldsymbol{\beta}}^{\text{EFF}}$ over $\widehat{\boldsymbol{\beta}}^{\text{ST}}$ are theoretically expected. We do this by selecting values of the data-generating model so that the difference (in terms of the Frobenius norm) between the estimating equations of $\widehat{\boldsymbol{\beta}}^{\text{EFF}}$ and $\widehat{\boldsymbol{\beta}}^{\text{ST}}$ is

maximized (leading to a scenario in which we expect $\widehat{\boldsymbol{\beta}}^{\mathrm{EFF}}$ to perform better) or minimized (leading to a scenario in which we expect similar performance for both estimators). See the Supplementary Material (Section 8) for more information on how to obtain these simulation scenarios.

Specifically, we consider $u = 2$ and $\alpha = 0$ (leading to a scenario in which the efficient estimator is expected to perform better than $\widehat{\boldsymbol{\beta}}^{\mathrm{ST}}$) or $\alpha = 2$ (leading to a scenario in which $\widehat{\boldsymbol{\beta}}^{\mathrm{EFF}}$ is expected to perform similarly to $\widehat{\boldsymbol{\beta}}^{\mathrm{ST}}$); corresponding to the true values $\beta_0 = 0$ and $\beta_0 = \sqrt{2}$, respectively. The simulation results are available in Section 8 of the Supplementary Material. Next, we describe our main findings.

The simulation results indicate that $\widehat{\boldsymbol{\beta}}^{\mathrm{ST}}$ is nearly as efficient as $\widehat{\boldsymbol{\beta}}^{\mathrm{EFF}}$ for both choices of β_0, suggesting that the information lost by not considering the cross-correlation of the pseudo observations is negligible, especially when the sample size increases. This is interesting, especially from a computational point of view: computationally $\widehat{\boldsymbol{\beta}}^{\mathrm{ST}}$ is substantially less intensive than the semiparametric efficient estimator $\widehat{\boldsymbol{\beta}}^{\mathrm{EFF}}$, because the latter requires calculating and inverting a covariance matrix of dimension $n(n-1)/2$. This behavior can be explained intuitively as follows. The number of nonzero elements of $\{\mathbf{V}(\boldsymbol{\beta}) + \mathbf{V}_{\mathrm{indep}}(\boldsymbol{\beta})\}$ in expression (2.10) is equal to $n(n-1)(n-3/2)$. The sparsity of this matrix is equal to $(4n-6)/[n(n-1)]$; which converges to zero as the sample size increases. Hence, with an increasing sample size, the variance-covariance matrix of the pseudo observations becomes sparser, with most of the significant information on the diagonal. This supports using the standard estimator in larger samples, rather than the locally efficient or even the bias-reduced estimator, resulting in the finite-sample bias becoming less of a concern.

The simulation results also show the local efficiency property of $\widehat{\boldsymbol{\beta}}^{\mathrm{BR}}$. Its relative efficiency, compared with that of $\widehat{\boldsymbol{\beta}}^{\mathrm{EFF}}$, is close to one when $\beta_0 = 0$, but increases, that is, $\widehat{\boldsymbol{\beta}}^{\mathrm{BR}}$ becomes less efficient than $\widehat{\boldsymbol{\beta}}^{\mathrm{EFF}}$, when the true β_0 deviates from zero. This is because the covariance structure of $\widehat{\boldsymbol{\beta}}^{\mathrm{BR}}$ is no longer specified correctly when $\beta_0 \neq 0$, leading to inefficient, but still consistent estimates. Its bias reduction property is also noticeable, especially when $\beta_0 = \sqrt{2}$, at small sample sizes. This bias reduction comes at a price of higher standard errors.

For small sample sizes, especially for $n = 25$, the empirical variance is underestimated by the sandwich estimator of the standard error (for all estimators). In these small sample settings, one could use resampling techniques, such as the adjusted jackknife empirical likelihood method (see Amorim et al. (2018)), to obtain appropriate confidence intervals with correct coverage. The coverage here using the asymptotic results is better for $n = 100$.

3.2. Gumbel distributed error

We now consider the case where the error term ε in model (2.11) follows a Gumbel distribution with location parameter zero and scale parameter one, which leads to a logistic PIM; see Table 1. As before, we consider specific simulation scenarios in which small and larger gains in efficiency are expected of $\widehat{\boldsymbol{\beta}}^{\mathrm{EFF}}$ over $\widehat{\boldsymbol{\beta}}^{\mathrm{ST}}$. The rationale behind these settings is discussed in the Supplementary Material (Section 8). Because the Gumbel distribution of the error is not symmetric around zero, this scenario also allows us to investigate the extent to which the result of Proposition 1 is retained here.

To better understand the restrictions imposed by PIMs, we also add $\widehat{\boldsymbol{\beta}}^{\mathrm{PH}}$ (the Cox partial likelihood estimator) to the simulation study. A semiparametric transformation model with a Gumbel error is equivalent to the Cox proportional hazards model, such that $\widehat{\boldsymbol{\beta}}^{\mathrm{PH}}$ is the efficient estimator under this more restrictive semiparametric transformation model. The simulation results can be found in Section 8 of the Supplementary Material. We discuss the main findings below.

The estimator $\widehat{\boldsymbol{\beta}}^{\mathrm{PH}}$ is more efficient than all competitors; it is around 20% more efficient than $\widehat{\boldsymbol{\beta}}^{\mathrm{EFF}}$, regardless of the sample size. This is because proportional hazard models or, more generally, semiparametric transformation models, are more restrictive than PIMs. In the Supplementary Material (Section 9), we explain in more detail why this is the case, and that this does not contradict the semiparametric theory. The relationship between these two approaches is discussed further in Section 5.

For the three PIM estimators (those that are within the class of RAL estimators given in Theorem 1), note that for $\beta_0 = 0$, there is almost no gain in efficiency, with all three estimators exhibiting similar performance. When $\beta_0 = 2$ and for a sample of size 25, both $\widehat{\boldsymbol{\beta}}^{\mathrm{BR}}$ and $\widehat{\boldsymbol{\beta}}^{\mathrm{EFF}}$ have a lower mean squared error (MSE) than that of $\widehat{\boldsymbol{\beta}}^{\mathrm{ST}}$. Note that the covariance structure used in the construction of $\widehat{\boldsymbol{\beta}}^{\mathrm{BR}}$ is misspecified when $\beta_0 = 2$, but its smaller bias contributes to a smaller MSE. Furthermore, even though the assumptions of Proposition 1 are not fulfilled (the Gumbel distribution is not symmetric around zero), $\widehat{\boldsymbol{\beta}}^{\mathrm{BR}}$ shows a smaller bias than the other estimators do, contributing to a lower MSE when the sample size is small.

As the sample size increases, all three estimators are nearly unbiased and the estimator $\widehat{\boldsymbol{\beta}}^{\mathrm{ST}}$ shows the largest MSE. The efficient estimator $\widehat{\boldsymbol{\beta}}^{\mathrm{EFF}}$ outperforms both competitors, although the improvement is modest. However, this gain in efficiency comes with a considerable cost in terms of computing time. Finally, the empirical coverage of the 95% confidence intervals is again similar.

3.3. Partition estimator

We also examine the finite-sample performance of the partition estimator $\widetilde{\boldsymbol{\beta}}$ of Section 2.4. We consider only the standard estimator. This is because the partition estimator is only of interest for large sample sizes, and we want to compare $\widetilde{\boldsymbol{\beta}}$ with $\widehat{\boldsymbol{\beta}}$. Furthermore, the calculations for $\widehat{\boldsymbol{\beta}}^{\mathrm{BR}}$ and $\widehat{\boldsymbol{\beta}}^{\mathrm{EFF}}$ (non-partitioned) become infeasible for large sample sizes. Section 8 of the Supplementary Material shows the empirical results of the partition estimator. For $n \geq 500$, the partition estimator $\widetilde{\boldsymbol{\beta}}^{\mathrm{ST}}$ is almost as efficient as $\widehat{\boldsymbol{\beta}}^{\mathrm{ST}}$. The partition variance estimator exhibits a slight underestimation, but this decreases as the sample size increases. Overall, we can say that the distributions of $\widetilde{\boldsymbol{\beta}}^{\mathrm{ST}}$ and $\widehat{\boldsymbol{\beta}}^{\mathrm{ST}}$ are approximately equal for $n \geq 500$.

3.4. Conclusion

The results of this empirical evaluation suggest that in many settings, there is no practical difference in efficiency between $\widehat{\boldsymbol{\beta}}^{\mathrm{ST}}$ and both $\widehat{\boldsymbol{\beta}}^{\mathrm{EFF}}$ and $\widehat{\boldsymbol{\beta}}^{\mathrm{BR}}$, and when there is a difference, its magnitude is modest. Therefore, $\widehat{\boldsymbol{\beta}}^{\mathrm{ST}}$ is the preferred estimator in practice, given its computational superiority over the other estimators.

4. Illustration

The Health Evaluation and Linkage to Primary Care study is a clinical trial for adult inpatients recruited from a detoxification unit. The data are made available in Appendix B of Horton and Kleinman (2010). To show how PIMs can supplement conventional analyses, we consider a cross-sectional part of the original study with $n = 453$, and focus on the association between the consequences of substance abuse and depression symptoms, while controlling for gender (1: female, 0: male) and homelessness (1: homeless at least one night in the last six months, 0: otherwise). The analyses are performed using R (R Core Team (2018)), and all R code for this analysis can be downloaded from `goo.gl/UA4mFV`.

Substance abuse consequence is the primary outcome, and is measured using the Inventory of Drug Use Consequences (InDUC, range 0–50) based on 50 items, in which higher scores indicate worse life consequences (Blanchard et al. (2001)). Because the InDUC score is an ordinal outcome, the probabilistic index is a relevant summary measure. Depression is measured using the Center for Epidemiologic Studies Depression Scale (CESD, range 0–60), with higher scores indicating more symptoms of depression. To visualize the association, Figure 1 (left panel) shows a scatter plot of the InDUC score as a function of the CESD,

Table 2. Estimates and standard errors of the three estimators when data are partitioned for computational reasons. The two columns on the right show the estimates and the standard errors, respectively when fitted to the data set without partitioning.

	With partitioning						**Without partitioning**	
	$\widetilde{\boldsymbol{\beta}}^{\mathrm{ST}}$	St. Error	$\widetilde{\boldsymbol{\beta}}^{\mathrm{BR}}$	St. Error	$\widetilde{\boldsymbol{\beta}}^{\mathrm{EFF}}$	St. Error	$\widehat{\boldsymbol{\beta}}^{\mathrm{ST}}$	St. Error
CESD	0.09708	0.03155	0.09657	0.03284	0.10037	0.03220	0.07704	0.03735
CESD^2	−0.00249	0.00112	−0.00232	0.00119	−0.00260	0.00117	−0.00197	0.00127
CESD^3	0.00003	0.00001	0.00002	0.00001	0.00003	0.00001	0.00002	0.00001
homeless	0.34356	0.06994	0.33512	0.07621	0.33742	0.07658	0.32386	0.06951
gender	−0.59627	0.08633	−0.59858	0.09279	−0.59578	0.09255	−0.61094	0.08684

together with the fit of a linear regression model, controlling for gender and homeless status. To allow for some flexibility, the CESD is modeled as a third-order polynomial. There is a positive association between the depression score and the mean substance abuse consequence score. The plot further shows a decreasing outcome variability with an increasing CESD score. Transforming the outcome using a Box–Cox power function stabilizes the variance and makes the residuals approximately normal; see the Supplementary Material for details (Section 10). The data-generating model can therefore be approximated using a probit link PIM. See Section 2.2 for more details on the connection between a transformation model and a PIM.

For all estimators, we consider the partition estimator of Section 2.4 with $k = \lfloor n^{0.25} \rfloor = \lfloor 453^{0.25} \rfloor = 4$ partitions, and $\widehat{\boldsymbol{\beta}}^{\mathrm{ST}}$ is also applied to the data set without partitioning.

Table 2 displays the results. All estimates are quite similar, which is in line with the findings of the simulation study. Furthermore note that there is only a small difference between these and the estimates obtained by fitting the standard estimator on the full data set. This demonstrates the strength of this estimator: it is computationally superior and nearly efficient. The partition standard errors are smaller than those of the estimator applied to the data set without partitioning, but this is the result of the variance estimator underestimation, as discussed in Section 3.3.

The regression coefficients of the CESD, gender and homeless status, are significantly different from zero ($p < 0.0001$). There is no evidence that the association between the CESD and the InDUC scores (while controlling for gender and homeless status) deviates significantly from linearity ($p = 0.22$).

To illustrate the interpretation, consider two patients of the same gender and with the same homeless status, but with a difference of 10 on the depression scale. The probability that the person with the lower CESD score will have a

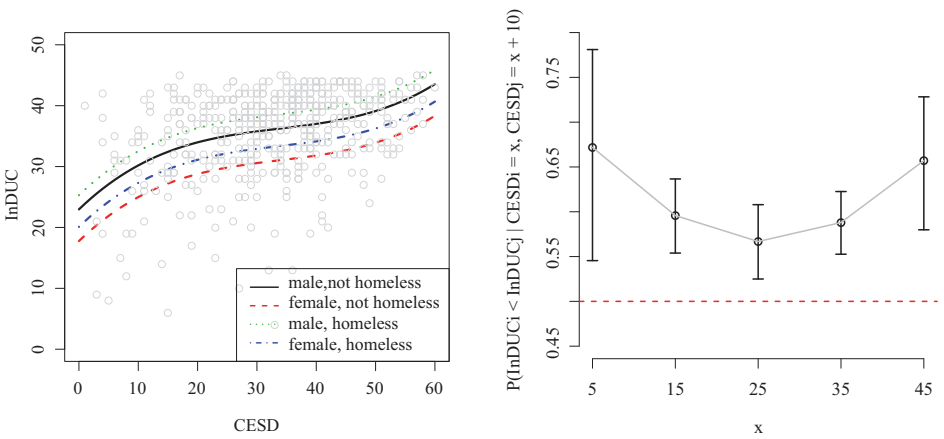

Figure 1. **Left**: scatter plot of the InDUC score as a function of the CESD, together with the fit of a linear regression model controlling for gender and homeless status. **Right**: estimated probability that the InDUC score is lower for a patient with a CESD of x, as compared with a patient with a CESD that is 10 units higher (both patients have the same gender and homelessness status) as a function of x. The vertical lines denote the pointwise 95% confidence intervals.

lower InDUC score is displayed in Figure 1 (right panel) as a function of the lowest CESD score. From this panel, we see that lower depression scores are associated with lower substance abuse consequence scores, because the PI is above 0.5. For example, consider a patient with a score of 25 and a patient with a score of 35. The probability that the patient with the lower CESD score will also have the lowest InDUC score is

$$\Phi[0.07704 \cdot (35 - 25) - 0.00197 \cdot (35^2 - 25^2) + 0.00002 \cdot (35^3 - 25^3)] = 56.7\%,$$

where the percentage is calculated using the unrounded values of $\widehat{\boldsymbol{\beta}}^{\mathrm{ST}}$. Hence, it is more likely that having fewer symptoms of depression is associated with fewer substance abuse consequences. However, the effect is modest, because the probability is close to 50%. From the plot, we also see that this effect is not linear: the impact of a 10 unit difference in the CESD score on the InDUC score depends on the CESD scores. For low and high CESD scores, the estimated PI is the largest, although it is less precise.

Note that when interpreting the PI, we are comparing two different subpopulations. In the former example, we are comparing two populations of the same gender and with the same homeless status, but with a difference of 10 points on the depression scale and with potentially different values for all other (possibly

unmeasured) variables. Depending on how heterogeneous these populations are, the PI will be closer to 0.5 when both subpopulations are more heterogeneous, and will deviate more from 0.5 when both subpopulations are more homogeneous. The PI does not indicate how much a specific patient would benefit from/be hurt by a 10 point increase on the depression scale. Similarly, in a randomized design, the PI refers to the interpretation of the probability that a randomly selected treated subject has a higher outcome than an independently randomly selected untreated subject. This should not be interpreted as the probability of benefiting from the treatment. For a more detailed discussion, see for example, Senn (2006) and Greenland et al. (2020).

5. Discussion

We have derived solid semiparametric theory for PIMs and the (locally) efficient estimator $\widehat{\boldsymbol{\beta}}^{\mathrm{EFF}}$ of the parameter $\boldsymbol{\beta}$ indexing these models, where efficiency is attained under an additional correct specification of an STM. We proposed a second estimator, $\widehat{\boldsymbol{\beta}}^{\mathrm{BR}}$, which has a local efficiency property and reduced second-order finite-sample bias.

Our results have shown that the standard estimator is nearly efficient under several data-generating mechanisms. This is surprising, considering the correlation between the pseudo observations, but can be explained by the sparsity of their covariance matrix. This degree of sparsity increases with the sample size n. An intuitive explanation for this behavior is given in Section 11 of the Supplementary Material. In view of this and its computational efficiency, we recommend using the standard estimator. We have further extended the standard estimator by providing a computationally improved partition estimation strategy. Techniques for sparse matrices will likely result in even better computational properties, and will be explored in future research.

PIMs might also be used to analyze composite outcomes. Pocock et al. (2012) proposed the win ratio, which is related to the PI, as a meaningful effect size. Recently, several authors have proposed methods for modeling the win ratio as a function of covariates (e.g., Follmann et al. (2020); Mao and Wang (2021)). The estimators of the model parameters are related to the PIM parameter estimators of Thas et al. (2012). However, because composite endpoints often involve time-to-event outcomes (e.g., survival times), censoring is an important issue when estimating the win ratio. These insights may perhaps be transferred to PIMs so that the estimation theory can be extended to censoring. We can also consider the approach of Cheng, Wei and Ying (1995) in the context of transformation

models for censored data, which relies on the inverse probability of censoring weighting.

Supplementary Material

This online Supplementary Material contains the development of the semiparametric efficiency theory, detailed proofs and calculations, explanations concerning the simulation setup, the tables of the simulation study, and additional figures for the data analysis.

Acknowledgments

The authors thank the Flemish Research Council (FWO) for financial support (Grant G.0202.14N) and the editor, associate editor, and referees for their insightful and constructive comments.

References

Amorim, G., Thas, O., Vermeulen, K., Vansteelandt, S. and De Neve, J. (2018). Small sample inference for probabilistic index models. *Computational Statistics and Data Analysis* **121**, 137–148.

Blanchard, K., Morgenstern, J., Morgan, T. and Labourie, E. (2001). Consequences of substance use: Psychometric properties of the Inventory of Drug Use Consequences (InDUC). *Alcohol and Clinical Experimentation and Research* **25**, 136A.

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics* **34**, 305–334.

Cheng, S., Wei, L. and Ying, Z. (1995). Analysis of transformation models with censored data. *Biometrika* **82**, 835–845.

Cuzick, J. (1988). Rank regression. *The Annals of Statistics* **16**, 1369–1389.

De Neve, J. and Thas, O. (2015). A regression framework for rank tests based on the probabilistic index model. *Journal of the American Statistical Association* **110**, 1276–1283.

Follmann, D., Fay, M. P., Hamasaki, T. and Evans, S. (2020). Analysis of ordered composite endpoints. *Statistics in Medicine* **39**, 602–616.

Greenland, S., Fay, M. P., Brittain, E. H., Shih, J. H., Follmann, D. A., Gabriel, E. E. et al. (2020). On causal inferences for personalized medicine: How hidden causal assumptions led to erroneous causal claims about the *D*-value. *The American Statistician* **74**, 243–248.

Horton, N. J. and Kleinman, K. (2010). *Using R for Data Management, Statistical Analysis, and Graphics*. CRC Press, Boca Raton.

Leng, C. and Cheng, G. (2012). Discussion of "Probabilistic Index Models" by O.Thas, J. De Neve, L. Clement and J.-P. Ottoy. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **74**, 661–662.

Mao, L. and Wang, T. (2021). A class of proportional win-fractions regression models for composite outcomes. *Biometrics* **77**, 1265–1275.

Newey, W. (1988). Adaptive estimation of regression models via moment restrictions. *Journal of Econometrics* **38**, 301–339.

Newey, W. (1990). Semiparametric efficiency bounds. *Journal of Applied Econometrics* **5**, 99–135.

Oja, H. (2012). Discussion of "Probabilistic Index Models" by O. Thas, J. De Neve, L. Clement and J.-P. Ottoy. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **74**, 663–664.

Pocock, S. J., Ariti, C. A., Collier, T. J. and Wang, D. (2012). The win ratio: A new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *European Heart Journal* **33**, 176–182.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Robins, J. M. and Ritov, Y. (1997). Toward a curse of dimensionality appropriate (CODA) asymptotic theory for semi-parametric models. *Statistics in Medicine* **16**, 285–319.

Senn, S. (2006). Letter to the Editor. Probabilistic index: An intuitive non-parametric approach to measuring the size of the treatment effects by L. Acion, J. J. Peterson, S. Temple and S. Arndt, Statistics in Medicine. *Statistics in Medicine* **25**, 3944–3946.

Thas, O., De Neve, J., Clement, L. and Ottoy, J.-P. (2012). Probabilistic index models (with discussion). *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **74**, 623–671.

Tsiatis, A. A. (2006). *Semiparametric Theory and Missing Data*. Springer, New York.

Van Keilegom, I. (2012). Discussion of "Probabilistic Index Models" by O. Thas, J. De Neve, L. Clement and J.-P. Ottoy. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **74**, 654.

Vermeulen, K., Thas, O. and Vansteelandt, S. (2015). Increasing the power of the Mann-Whitney test in randomized experiments through flexible covariate adjustment. *Statistics in Medicine* **34**, 1012–1030.

Karel Vermeulen

Department of Data Analysis and Mathematical Modelling, Ghent University, Ghent, Belgium.

E-mail: Karelb.Vermeulen@UGent.be

Jan De Neve

Department of Data Analysis, Ghent University, Ghent, Belgium.

E-mail: Jan.DeNeve@UGent.be

Gustavo Amorim

Department of Data Analysis and Mathematical Modelling, Ghent University, Ghent, Belgium and Department of Biostatistics, Vanderbilt University Medical Center, Nashville, US.

E-mail: ggca@outlook.com

Olivier Thas

Department of Data Analysis and Mathematical Modelling, Ghent University, Ghent, Belgium and Center for Statistics, Hasselt University, Hasselt, Belgium and National Institute for Applied Statistics Research Australia (NIASRA), School of Mathematics and Applied Statistics, University of Wollongong, Wollongong, Australia.

E-mail: Olivier.Thas@UGent.be

Stijn Vansteelandt

Department of Applied Mathematics, Computer Sciences and Statistics, Ghent University, Ghent, Belgium and Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, United Kingdom.

E-mail: Stijn.Vansteelandt@UGent.be

(Received March 2021; accepted August 2021)