Statistical Theory and Related Fields

ISSN: 2475-4269 (Print) 2475-4277 (Online) Journal homepage: www.tandfonline.com/journals/tstf20

Bias correction of partial-error in variables in a Poisson regression model

Kentarou Wada & Takeshi Kurosawa

STATISTICAL THEORY AND RELATED FIELDS
https://doi.org/10.1080/24754269.2025.2588849

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Bias correction of partial-error in variables in a Poisson regression model

Kentarou Wada and Takeshi Kurosawa

Department of Applied Mathematics, Tokyo University of Science, Tokyo, Japan

ABSTRACT

Previous studies on Poisson regression models with error-in-variables (EIV) assumed either a univariate EIV structure or multivariate EIV framework with all explanatory variables subject to error where the explanatory variable and error vectors are restricted to multivariate normal distributions. This study assumes that the explanatory variable and error vectors follow general distributions, with measurement error affecting only a subset of variables in the multivariate EIV framework. We define the partial-error naive estimator, derive its asymptotic bias and mean squared error, and propose a consistent estimator for the true parameter by correcting this bias. We also investigate a simplification of the new estimator when all components of the explanatory variable and error vectors are independent. This method is applicable even when the explanatory variable or error vectors follow a mixed distribution. Simulation studies and real data analysis are presented as illustrative examples to compare the performance of the partial-error naive estimator with that of the new estimator.

ARTICLE HISTORY
Received 13 July 2024
Revised 29 May 2025
Accepted 9 November 2025

KEYWORDS
Poisson regression; error-in-variables; multiple covariates; consistent estimator

1. Introduction

Data are frequently affected by measurement errors that stem from the inability to observe variables in their true form. Such errors can introduce bias into the estimation outcomes. Error-in-variables (EIV) models account for measurement errors in explanatory variables. Classical error models (Kukush & Schneeweiss, 2000; Shklyar et al., 2007) and Berkson error models (Burr, 1988; Huwang & Huang, 2000) are well-known examples of EIV models. This study focuses primarily on classical error models. Nonlinear models, particularly generalized linear models, offer greater flexibility than linear models. However, the estimation of generalized linear models from error-affected data is challenging. Various studies have explored nonlinear EIV models (e.g., Box, 1963; Geary, 1953), including the corrected score function developed by Nakamura (1990) for estimating generalized linear models using EIV. We mainly focussed on the Poisson regression model in the context of measurement errors. Among generalized linear models, the Poisson regression model is representative and analytically tractable.

CONTACT Kentarou Wada wadaken5269@gmail.com Department of Applied Mathematics, Tokyo University of Science, Tokyo, 162-8601, Japan

Kukush et al. (2004), Shklyar and Schneeweiss (2005), Jiang and Ma (2022), Guo and Li (2002) and Wada and Kurosawa (2023) have discussed approaches to Poisson regression models with classical errors. Kukush et al. (2004) demonstrated the statistical properties of the naive, corrected score, and structural quasi-score estimators of a Poisson regression model with a normal explanatory variable and error. Shklyar and Schneeweiss (2005), assuming a multivariate normal distribution for an explanatory variable and measurement error, compared the asymptotic covariance matrices of the corrected score, simple structural estimator, and structural quasi-score estimator of a Poisson regression model. Jiang and Ma (2022) assumed a high-dimensional explanatory variable with a multivariate normal error, proposing a new estimator for a Poisson regression model by combining Lasso regression and the corrected score function. From a Poisson regression model with classical errors, Guo and Li (2002) proposed an estimator that generalizes the corrected score function discussed by Nakamura (1990) for generally distributed errors, and derived the proposed estimator's asymptotic normality. Wada and Kurosawa (2023) generalized the naive estimator discussed by Kukush et al. (2004), derived the asymptotic bias of the naive estimator, and proposed a consistent estimator of the unknown parameter using the naive estimator.

Although various studies have been conducted on Poisson regression models with EIV, they have typically assumed a normal distribution for the explanatory variable. However, the explanatory variable is not always normally distributed. In addition, previous studies have assumed either a univariate EIV structure (Kukush et al., 2004; Wada & Kurosawa, 2023) or a multivariate EIV framework with all explanatory variables subject to error where the explanatory variable and error vectors are restricted to multivariate normal distributions (Shklyar & Schneeweiss, 2005). In contrast, this study assumes that the explanatory variable and error vectors follow general distributions, with measurement errors included in a subset of explanatory variables within the multivariate EIV framework. We propose a consistent estimator for the true parameter, and investigate a simplification of the new estimator when all components of the explanatory variable and error vectors are independent. This method is applicable, even when the explanatory variable or error vectors follow a mixed distribution.

Section 2 presents the Poisson regression model with measurement errors and defines the partial-error naive (PN) estimator. Section 3 considers the requirements for the existence of a PN estimator and derives its asymptotic bias and mean squared error (MSE), assuming that the explanatory variable and measurement error have general distributions. Section 4 proposes the corrected partial-error naive (CPN) estimator as a consistent estimator of the true parameter by correcting the bias of the PN estimator. It then investigates a simplification of the CPN estimator when all components of the explanatory variable and error vectors are independent. Additionally, we provide examples of the application of this CPN estimator to a number of cases: a multivariate normal explanatory variable with a normal error, Bernoulli and gamma explanatory variables with a gamma error, and gamma and normal explanatory variables with gamma and normal errors. Section 5 presents simulation studies that compare the performance of the PN and CPN estimators. Section 6 applies the PN and CPN estimators to real data.

2. Preliminary

In this section, we present a Poisson regression model with measurement errors and define the partial-error naive (PN) estimator.

2.1. Poisson regression model with errors

We assume a Poisson regression model of the response variable Y and vector of explanatory variables $\boldsymbol{X} = (X_1, \ldots, X_{p+q})^\top$.

$$Y \mid \boldsymbol{X} \sim \mathrm{Po}\left(\exp\left(\boldsymbol{\beta}^\top \begin{pmatrix} 1 \\ \boldsymbol{X} \end{pmatrix}\right)\right), \tag{1}$$

where $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_{p+q})^\top$. Under ordinary circumstances, $\boldsymbol{X}$ is assumed to be correctly observed. Here, we assume that the vector of explanatory variables $\boldsymbol{X}$ partially includes the measurement errors in the multivariate EIV framework. We define subvectors of $\boldsymbol{X}$ as

$$\boldsymbol{X}_o = (X_1, \ldots, X_p)^\top, \quad \boldsymbol{X}_e = (X_{p+1}, \ldots, X_{p+q})^\top. \tag{2}$$

While $\boldsymbol{X}_o$ represents a vector of explanatory variables that can be observed directly, $\boldsymbol{X}_e$ represents a vector of explanatory variables that cannot be observed directly. The values of $\boldsymbol{X}_e$ are observed with the unexpected measurement errors. Thus, $\boldsymbol{X}_e$ (but not $\boldsymbol{X}_o$) has a stochastic error $\boldsymbol{U}$ with

$$\boldsymbol{W} = \boldsymbol{X}_e + \boldsymbol{U},$$

where $\boldsymbol{U}$ is assumed to be independent of $(\boldsymbol{X}, Y)$, and $\boldsymbol{W}$ is observable. As $\boldsymbol{X}$ is also a random variable, it inherently includes the measurement error. However, we assume the existence of an exogenous measurement error U that is not due to the distribution of $\boldsymbol{X}$. For example, consider two measurement devices with their own intrinsic measurement errors. One device observes the value as $\boldsymbol{X}_o$, whereas the other, owing to product degradation or malfunction, includes an additional error $\boldsymbol{U}$ attached to the true measurement $\boldsymbol{X}_e$. This results in the observed value $\boldsymbol{W}$. Here, we assume that $\boldsymbol{X}$ is an $\mathbb{R}^{p+q}$-valued random vector, whereas $\boldsymbol{U}$ and $\boldsymbol{W}$ are $\mathbb{R}^q$-valued random vectors. We also assume that $(\boldsymbol{X}_i = (\boldsymbol{X}_{o,i}^\top, \boldsymbol{X}_{e,i}^\top)^\top, Y_i)$ $(i = 1, \ldots, n)$ are independently and identically distributed (i.i.d.) samples from the distribution of $(\boldsymbol{X}, Y)$. Furthermore, we assume that U_i $(i = 1, \ldots, n)$ are independent samples from the distribution of $\boldsymbol{U}$. Although we can observe $Y_i, \boldsymbol{X}_{o,i}, W_i, \boldsymbol{X}_{e,i} + \boldsymbol{U}_i$ $(i = 1, \ldots, n)$, we assume that $\boldsymbol{X}_e$ and $\boldsymbol{W}$ cannot be observed directly. Even when the distributions $\boldsymbol{X}$ and $\boldsymbol{W}$ are known, estimating model parameters from mismeasured data remains infeasible without additional information about the measurement error. Parameter estimation in this context requires at least partial knowledge of $\boldsymbol{U}$. EIV models rely on realistic assumptions about such error structures. Typical assumptions include a known mean and variance for $\boldsymbol{U}$ or a known mean for $\boldsymbol{U}$ along with the known ratio $\kappa_j = \boldsymbol{V}[X_j]/\boldsymbol{V}[W_j]$, where W_j is the component of $\boldsymbol{W}$ corresponding to X_j $(j = p+1, \ldots, p+q)$ (Fuller, 1987). Because $\boldsymbol{U}$ represents the measurement error, its mean is often assumed to be zero. Its variance may be estimated empirically. For instance, when a measuring device malfunctions during data collection, the data may be observed both before and after the introduction of errors, allowing the estimation of the mean and variance of $\boldsymbol{U}$. Based on this scenario, we assume that both the mean and variance of $\boldsymbol{U}$ are known. In the following definitions, the functions $M_{\boldsymbol{X}}$ and $K_{\boldsymbol{X}}$ represent the moment- and cumulant generating functions, respectively, for a random vector $\boldsymbol{X}$. We denote the subvectors of $\boldsymbol{\beta}$ as $\boldsymbol{\beta}_1 = (\beta_1, \ldots, \beta_p)^\top$ and $\boldsymbol{\beta}_2 = (\beta_{p+1}, \ldots, \beta_{p+q})^\top$. These are regression parameters corresponding to the subvectors of explanatory variables $\boldsymbol{X}_o$ and $\boldsymbol{X}_e$, respectively.

2.2. Partial-error naive estimator

Definition 2.1: We define the PN estimator $\hat{\boldsymbol{\beta}}^{(\mathrm{PN})} = (\hat{\beta}_0^{(\mathrm{PN})}, \ldots, \hat{\beta}_{p+q}^{(\mathrm{PN})})^\top$ for $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_{p+q})^\top$ as the solution of the following equation:

$$S_n\left(\hat{\boldsymbol{\beta}}^{(\mathrm{PN})}\right) = \mathbf{0}_{p+q+1},$$

where

$$\begin{aligned} S_n(\tilde{\boldsymbol{b}}) &= \frac{1}{n}\sum_{i=1}^{n}\left\{Y_i - \exp\left(\tilde{b}_0 + \tilde{\boldsymbol{b}}_1^\top \boldsymbol{X}_{o,i} + \tilde{\boldsymbol{b}}_2^\top \boldsymbol{W}_i\right)\right\}(1, \boldsymbol{X}_{o,i}^\top, \boldsymbol{W}_i^\top)^\top, \\ \tilde{\boldsymbol{b}} &= \left(\tilde{b}_0, \ldots, \tilde{b}_{p+q}\right)^\top, \end{aligned} \tag{3}$$

$\tilde{\boldsymbol{b}}_1 = (\tilde{b}_1, \ldots, \tilde{b}_p)^\top, \tilde{\boldsymbol{b}}_2 = (\tilde{b}_{p+1}, \ldots, \tilde{b}_{p+q})^\top$ are subvectors of $\tilde{\boldsymbol{b}}$ and $\mathbf{0}_{p+q+1}$ is a $(p+q+1)$-dimensional vector with zeros.

This definition is a natural extension of the naive estimator in Kukush et al. (2004). Following the argument in Kukush and Shklyar (2002), we obtain the convergence of the PN estimator:

$$\hat{\boldsymbol{\beta}}^{(\mathrm{PN})} \xrightarrow{\text{a.s.}} \boldsymbol{b} \neq \boldsymbol{\beta}, \tag{4}$$

where $\boldsymbol{b} = (b_0, \boldsymbol{b}_1^\top, \boldsymbol{b}_2^\top)^\top$ is a solution to the following estimating equation:

$$\mathbf{E}_{\boldsymbol{X},\boldsymbol{W}}\left[\mathbf{E}_{Y|(\boldsymbol{X},\boldsymbol{W})}\left[\left\{Y - \exp(b_0 + \boldsymbol{b}_1^\top \boldsymbol{X}_o + \boldsymbol{b}_2^\top \boldsymbol{W})\right\}(1, \boldsymbol{X}_o^\top, \boldsymbol{W}^\top)^\top\right]\right] = \mathbf{0}_{p+q+1}. \tag{5}$$

3. Properties of the partial-error naive estimator

In this section, we consider the requirements for the existence of a PN estimator and derive its asymptotic bias and mean square error (MSE), assuming that the explanatory variable and measurement error each follow a general distribution.

3.1. Existence of the partial-error naive estimator

The PN estimator does not always exist for $\boldsymbol{X}$ and $\boldsymbol{U}$ when they are general random variables. Therefore, we assume the existence of the following expectation:

$$\mathbf{E}_{\boldsymbol{X},\boldsymbol{W}}\left[\mathbf{E}_{Y|(\boldsymbol{X},\boldsymbol{W})}\left[\left\{Y - \exp(b_0 + \boldsymbol{b}_1^\top \boldsymbol{X}_o + \boldsymbol{b}_2^\top \boldsymbol{W})\right\}(1, \boldsymbol{X}_o^\top, \boldsymbol{W}^\top)^\top\right]\right].$$

This expectation is assumed to be a requirement for the existence of the PN estimator. Consequently, the following six expectations are met:

$$
\begin{aligned}
\mathbf{E}[Y] &= \mathbf{E}_{\boldsymbol{X}}\left[\mathbf{E}[Y \mid \boldsymbol{X}]\right] = \mathbf{E}_{\boldsymbol{X}}\left[\exp(\beta_0 + \boldsymbol{\beta}_1^\top \boldsymbol{X}_o + \boldsymbol{\beta}_2^\top \boldsymbol{X}_e)\right] = \mathrm{e}^{\beta_0} M_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix},\\
\mathbf{E}[Y\boldsymbol{X}_o] &= \mathbf{E}_{\boldsymbol{X}}\left[\boldsymbol{X}_o \mathbf{E}[Y \mid \boldsymbol{X}]\right] = \mathbf{E}_{\boldsymbol{X}}\left[\boldsymbol{X}_o \exp(\beta_0 + \boldsymbol{\beta}_1^\top \boldsymbol{X}_o + \boldsymbol{\beta}_2^\top \boldsymbol{X}_e)\right]\\
&= \mathrm{e}^{\beta_0} \frac{\partial}{\partial \boldsymbol{\beta}_1} M_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix},\\
\mathbf{E}[Y\boldsymbol{W}] &= \mathbf{E}_{\boldsymbol{X}}\left[\mathbf{E}[Y \mid \boldsymbol{X}]\mathbf{E}[\boldsymbol{W} \mid \boldsymbol{X}]\right]\\
&= \mathbf{E}_{\boldsymbol{X}}\left[(\boldsymbol{X}_e + \mathbf{E}[\boldsymbol{U}]) \exp(\beta_0 + \boldsymbol{\beta}_1^\top \boldsymbol{X}_o + \boldsymbol{\beta}_2^\top \boldsymbol{X}_e)\right]\\
&= \mathrm{e}^{\beta_0} \mathbf{E}[\boldsymbol{U}] M_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix} + \mathrm{e}^{\beta_0} \frac{\partial}{\partial \boldsymbol{\beta}_2} M_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix},\\
\mathbf{E}[\exp(b_0 + \boldsymbol{b}_1^\top \boldsymbol{X}_o + \boldsymbol{b}_2^\top \boldsymbol{W})] &= \mathrm{e}^{b_0} \mathbf{E}[\exp(\boldsymbol{b}_1^\top \boldsymbol{X}_o + \boldsymbol{b}_2^\top \boldsymbol{X}_e + \boldsymbol{b}_2^\top \boldsymbol{U})]\\
&= \mathrm{e}^{b_0} M_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{pmatrix} M_{\boldsymbol{U}}(\boldsymbol{b}_2),\\
\mathbf{E}[\boldsymbol{X}_o \exp(b_0 + \boldsymbol{b}_1^\top \boldsymbol{X}_o + \boldsymbol{b}_2^\top \boldsymbol{W})] &= \mathbf{E}_{\boldsymbol{X}}\left[\mathbf{E}_{\boldsymbol{U}}\left[\boldsymbol{X}_o \exp(b_0 + \boldsymbol{b}_1^\top \boldsymbol{X}_o + \boldsymbol{b}_2^\top \boldsymbol{X}_e + \boldsymbol{b}_2^\top \boldsymbol{U})\right]\right]\\
&= \mathrm{e}^{b_0} \frac{\partial}{\partial \boldsymbol{b}_1} M_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{pmatrix} M_{\boldsymbol{U}}(\boldsymbol{b}_2),\\
\mathbf{E}[\boldsymbol{W} \exp(b_0 + \boldsymbol{b}_1^\top \boldsymbol{X}_o + \boldsymbol{b}_2^\top \boldsymbol{W})] &= \mathbf{E}_{\boldsymbol{X}}\left[\mathbf{E}_{\boldsymbol{U}}\left[(\boldsymbol{X}_e + \boldsymbol{U}) \exp(b_0 + \boldsymbol{b}_1^\top \boldsymbol{X}_o + \boldsymbol{b}_2^\top \boldsymbol{X}_e + \boldsymbol{b}_2^\top \boldsymbol{U})\right]\right]\\
&= \mathrm{e}^{b_0} \frac{\partial}{\partial \boldsymbol{b}_2} M_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{pmatrix} M_{\boldsymbol{U}}(\boldsymbol{b}_2)\\
&\quad + \mathrm{e}^{b_0} M_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{pmatrix} \frac{\partial}{\partial \boldsymbol{b}_2} M_{\boldsymbol{U}}(\boldsymbol{b}_2).
\end{aligned}
\tag{6}
$$

We use the conditional independence of Y and $\boldsymbol{W}$ under a given $\boldsymbol{X}$ to calculate these expectations (see Lemma A.3). These expectations require the existence of the following condition:

$$
M_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix}, \quad M_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{pmatrix}, \quad M_{\boldsymbol{U}}(\boldsymbol{b}_2). \tag{7}
$$

This condition is required for the existence of a PN estimator.

3.2. Asymptotic bias of the partial-error naive estimator

The PN estimator satisfies (4) and has an asymptotic bias for the true $\boldsymbol{\beta}$. Here, we derive the asymptotic bias under general conditions. Let $\boldsymbol{G}\left(\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix}, \begin{pmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{pmatrix}\right)$ be a function by

$$
\begin{pmatrix}
\frac{\partial}{\partial \boldsymbol{b}_1} K_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{pmatrix} - \frac{\partial}{\partial \boldsymbol{\beta}_1} K_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix}\\
\frac{\partial}{\partial \boldsymbol{b}_2} K_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{pmatrix} + \frac{\partial}{\partial \boldsymbol{b}_2} K_{\boldsymbol{U}}(\boldsymbol{b}_2) - \mathbf{E}[\boldsymbol{U}] - \frac{\partial}{\partial \boldsymbol{\beta}_2} K_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix}
\end{pmatrix}. \tag{8}
$$

Theorem 3.1: *Let $Y \mid \boldsymbol{X}$ be a Poisson regression in* (1) *with* (2). *We assume conditions* (C1)–(C3).

(C1) (7) *exists.*
(C2) $\det \dfrac{\partial \boldsymbol{G}}{\partial \begin{pmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \end{pmatrix}^\top} \neq 0$ *is satisfied where* $\mathbf{G}$ *is given in* (8).
(C3) *Each component of* $\hat{\boldsymbol{\beta}}^{(\mathrm{PN})}(\hat{\boldsymbol{\beta}}^{(\mathrm{PN})})^\top$ *is uniformly integrable.*

Then, following conditions (C1)–(C3), *the asymptotic bias of the PN estimator* $\hat{\boldsymbol{\beta}}^{(\mathrm{PN})} = (\hat{\beta}_0^{(\mathrm{PN})}, (\hat{\boldsymbol{\beta}}_1^{(\mathrm{PN})})^\top, (\hat{\boldsymbol{\beta}}_2^{(\mathrm{PN})})^\top)^\top$ *is represented as*

$$\lim_{n\to\infty} \mathbf{E}[\hat{\beta}_0^{(\mathrm{PN})} - \beta_0] = b_0 - \beta_0 = \log\left(\frac{M_X\begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}}{M_X\left(\boldsymbol{g}\begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}\right) M_U\left(\boldsymbol{g}_2\begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}\right)} \right),$$

$$\lim_{n\to\infty} \mathbf{E}\left[\begin{pmatrix} \hat{\boldsymbol{\beta}}_1^{(\mathrm{PN})} - \boldsymbol{\beta}_1 \\ \hat{\boldsymbol{\beta}}_2^{(\mathrm{PN})} - \boldsymbol{\beta}_2 \end{pmatrix}\right] = \begin{pmatrix} \boldsymbol{b}_1 - \boldsymbol{\beta}_1 \\ \boldsymbol{b}_2 - \boldsymbol{\beta}_2 \end{pmatrix} = \boldsymbol{g}\begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} - \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix},$$

where $\boldsymbol{g}$ *is a continuously differentiable implicit function, with* $\begin{pmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \end{pmatrix} = \boldsymbol{g}\begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}$ *in the neighbourhood of* $\left(\begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \end{pmatrix}\right)$ *satisfying* $\mathbf{G} = \mathbf{0}$. *Furthermore,* $\boldsymbol{g}_2(\boldsymbol{x}) = (g_{p+1}(\boldsymbol{x}), \ldots, g_{p+q}(\boldsymbol{x}))^\top$ *is a subvector of* $\boldsymbol{g}$. *The asymptotic MSE of the PN estimator is then given by the squared asymptotic bias.*

Remark 3.1: In Wada and Kurosawa (2023), the asymptotic bias and MSE of the naive estimator were derived under the condition that the limit and expectation were exchangeable without noting. These results are valid under the condition. Regarding exchangeability, by referring to the discussion in Kukush and Shklyar (2002), we can replace the convergence in probability of the naive estimator with almost sure convergence. Then, assuming uniform integrability additionally, the Vitali convergence theorem (Rosenthal, 2025) can be applied, which justifies the exchangeability of the limit and expectation.

4. Bias correction

In this section, we propose the corrected partial-error naive (CPN) estimator as a consistent estimator of the true parameter by correcting the bias of the PN estimator. We investigate a simplification of the CPN estimator when all components of the explanatory variable and error vectors are independent. Additionally, we provide examples of the application of the CPN estimator for a number of cases, including a multivariate normal explanatory variable with a normal error, Bernoulli and gamma explanatory variables with a gamma error, and gamma and normal explanatory variables with gamma and normal errors.

4.1. Corrected partial-error naive estimator

The exact distribution of $Y \mid \boldsymbol{W}$ is given by

$$f_{Y|\boldsymbol{W}}(y \mid \boldsymbol{w}) = \frac{1}{f_{\boldsymbol{W}}(\boldsymbol{w})} \int_{\mathrm{supp}(f_{\boldsymbol{X}})} f_{Y|\boldsymbol{X}}(y \mid \boldsymbol{x}) f_{\boldsymbol{U}}(\boldsymbol{w} - \boldsymbol{x}_e) f_{\boldsymbol{X}}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}$$
$$= \frac{1}{f_{\boldsymbol{W}}(\boldsymbol{w})} \int_{\mathrm{supp}(f_{\boldsymbol{X}})} \mathrm{Po}\left(\exp\left(\boldsymbol{\beta}^\top \begin{pmatrix} 1 \\ \boldsymbol{x} \end{pmatrix}\right)\right) f_{\boldsymbol{U}}(\boldsymbol{w} - \boldsymbol{x}_e) f_{\boldsymbol{X}}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}. \qquad (9)$$

In general, we cannot calculate a right-hand side integral of (9). Therefore, we cannot use the maximum likelihood estimator of $Y \mid \boldsymbol{W}$. In this study, we construct a consistent estimator of the true parameter by correcting the bias of the PN estimator. This is proposed as a corrected partial-error naive (CPN) estimator using the following theorem.

Theorem 4.1: *Let $Y \mid \boldsymbol{X}$ be a Poisson regression in* (1) *with* (2). *We assume* (C1) *in Theorem* 3.1 *and Condition* (C'2).

(C'2) $\det \dfrac{\partial \boldsymbol{G}}{\partial \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}^\top} \neq 0$ *is satisfied where* $\boldsymbol{G}$ *is given in* (8).

Then, the CPN estimator $\hat{\boldsymbol{\beta}}^{(\mathrm{CPN})} = (\hat{\beta}_0^{(\mathrm{CPN})}, (\hat{\boldsymbol{\beta}}_1^{(\mathrm{CPN})})^\top, (\hat{\boldsymbol{\beta}}_2^{(\mathrm{CPN})})^\top)^\top$ *of* $\boldsymbol{\beta}$, *which is strongly consistent, is represented as*

$$\hat{\beta}_0^{(\mathrm{CPN})} = \hat{\beta}_0^{(\mathrm{PN})} + \log\left(\frac{M_{\boldsymbol{X}}\begin{pmatrix} \hat{\boldsymbol{\beta}}_1^{(\mathrm{PN})} \\ \hat{\boldsymbol{\beta}}_2^{(\mathrm{PN})} \end{pmatrix} M_{\boldsymbol{U}}\left(\hat{\boldsymbol{\beta}}_2^{(\mathrm{PN})}\right)}{M_{\boldsymbol{X}}\begin{pmatrix} \hat{\boldsymbol{\beta}}_1^{(\mathrm{CPN})} \\ \hat{\boldsymbol{\beta}}_2^{(\mathrm{CPN})} \end{pmatrix}} \right),$$

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}_1^{(\mathrm{CPN})} \\ \hat{\boldsymbol{\beta}}_2^{(\mathrm{CPN})} \end{pmatrix} = \boldsymbol{h}\begin{pmatrix} \hat{\boldsymbol{\beta}}_1^{(\mathrm{PN})} \\ \hat{\boldsymbol{\beta}}_2^{(\mathrm{PN})} \end{pmatrix},$$

where $\boldsymbol{h}$ *is a continuously differentiable implicit function with* $\begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} = \boldsymbol{h}\begin{pmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \end{pmatrix}$ *in the neighbourhood of* $\left(\begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \end{pmatrix}\right)$ *satisfying* $\boldsymbol{G} = \boldsymbol{0}$.

Note that $\boldsymbol{h}$ in Theorem 4.1 is the inverse function of $\boldsymbol{g}$ in Theorem 3.1. Here, we consider a situation in which the components of random vectors $\boldsymbol{X}$ and $\boldsymbol{U}$ are independent. We obtain the CPN estimator in Theorem 4.2 where the components of $\boldsymbol{X} = (X_1, \ldots, X_{p+q})^\top$ and $\boldsymbol{U} = (U_1, \ldots, U_q)^\top$ are independent.

Theorem 4.2: *Let $Y \mid \boldsymbol{X}$ be a Poisson regression in* (1) *with* (2). *We assume the following conditions* (D1)–(D3).

(D1) *Assume the existence of* $M_{X_i}(b_i)$ $(i = 1, \ldots, p+q)$, $M_{U_j}(b_j)$ $(j = 1, \ldots, q)$.
(D2) $K''_{X_j}(\beta_j) \neq 0$ $(j = 1, \ldots, p+q)$ *is satisfied.*

(D3) *The components of* $\boldsymbol{X} = (X_1, \ldots, X_{p+q})^\top$ *and* $\boldsymbol{U} = (U_1, \ldots, U_q)^\top$ *are independent.*

Then, the CPN estimator $\hat{\boldsymbol{\beta}}^{(\mathrm{CPN})} = (\hat{\beta}_0^{(\mathrm{CPN})}, \ldots, \hat{\beta}_{p+q}^{(\mathrm{CPN})})^\top$ *of* $\boldsymbol{\beta}$*, which is strongly consistent, is represented as*

$$\hat{\beta}_0^{(\mathrm{CPN})} = \hat{\beta}_0^{(\mathrm{PN})} + \sum_{i=p+1}^{p+q} K_{X_i}\left(\hat{\beta}_i^{(\mathrm{PN})}\right) + \sum_{i=1}^{q} K_{U_i}\left(\hat{\beta}_{p+i}^{(\mathrm{PN})}\right) - \sum_{i=p+1}^{p+q} K_{X_i}\left(\hat{\beta}_i^{(\mathrm{CPN})}\right),$$

$$\hat{\beta}_j^{(\mathrm{CPN})} = \hat{\beta}_j^{(\mathrm{PN})} \quad (j = 1, \ldots, p),$$

$$\hat{\beta}_j^{(\mathrm{CPN})} = h_j\left(\hat{\beta}_j^{(\mathrm{PN})}\right) \quad (j = p+1, \ldots, p+q),$$

where h_j $(j = p+1, \ldots, p+q)$ *is a continuously differentiable implicit function with* $\beta_j = h_j(b_j)$ *in the neighbourhood of* (β_j, b_j) *satisfying*

$$G_j(\beta_j, b_j) = K'_{X_j}(b_j) + K'_{U_{j-p}}(b_j) - \mathbf{E}[U_{j-p}] - K'_{X_j}(\beta_j) = 0.$$

Remark 4.1: The implicit function h_j in Theorem 4.2 is equivalent to the formula for the corrected naive (CN) estimator proposed in Wada and Kurosawa (2023) for univariate EIV models. Thus, we can use the CN estimator for multivariate EIV models when the components of $\boldsymbol{X}$ and $\boldsymbol{U}$ are independent.

4.2. Application examples

Example 4.3: We derive the CPN estimator assuming that

$$Y \mid \boldsymbol{X} \sim \mathrm{Po}\left(\exp\left(\boldsymbol{\beta}^\top \begin{pmatrix} 1 \\ \boldsymbol{X} \end{pmatrix}\right)\right), \quad \boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^\top, \quad \boldsymbol{X} = (X_1, X_2, X_3)^\top,$$

$$\boldsymbol{X}_o = (X_1, X_2)^\top, \quad X_e = X_3, \quad W = X_3 + U,$$

$$\boldsymbol{X} \sim N_3(\boldsymbol{\mu}, \Sigma), \quad U \sim N(0, \sigma^2),$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)^\top \in \mathbb{R}^3, \Sigma = [\sigma_{ij}] \in \mathbb{R}^{3\times 3}, 0 < \sigma^2 < \infty$. We use the following partition expressions for $\boldsymbol{\mu}$ and Σ:

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \mu_3 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_1 & \boldsymbol{\sigma}_3 \\ \boldsymbol{\sigma}_3^\top & \sigma_{33} \end{pmatrix},$$

where

$$\boldsymbol{\mu}_1 = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma_1 = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}, \quad \boldsymbol{\sigma}_3 = \begin{pmatrix} \sigma_{13} \\ \sigma_{23} \end{pmatrix}.$$

We obtain

$$\boldsymbol{G}\left(\begin{pmatrix} \boldsymbol{\beta}_1 \\ \beta_3 \end{pmatrix}, \begin{pmatrix} \boldsymbol{b}_1 \\ b_3 \end{pmatrix}\right) = \begin{pmatrix} \Sigma_1(\boldsymbol{b}_1 - \boldsymbol{\beta}_1) + (b_3 - \beta_3)\boldsymbol{\sigma}_3 \\ (\boldsymbol{b}_1 - \boldsymbol{\beta}_1)'\boldsymbol{\sigma}_3 + b_3(\sigma_{33} + \sigma^2) - \beta_3\sigma_{33} \end{pmatrix}$$

and $\boldsymbol{G}$ satisfies

$$\det \frac{\partial \boldsymbol{G}}{\partial \begin{pmatrix} \boldsymbol{\beta}_1 \\ \beta_3 \end{pmatrix}^\top} = \det \begin{pmatrix} -\Sigma_1 & -\boldsymbol{\sigma}_3 \\ -\boldsymbol{\sigma}_3^\top & -\sigma_{33} \end{pmatrix} = (-1)^3 \det(\Sigma) \neq 0.$$

From $\boldsymbol{G} = \boldsymbol{0}_3$, we obtain the following implicit function:

$$\boldsymbol{\beta}_1 = \boldsymbol{b}_1 - \frac{\sigma_2^2 b_3}{\sigma_{33} - \sigma_3^\top \Sigma_1^{-1} \boldsymbol{\sigma}_3},$$
$$\beta_3 = \frac{\sigma_{33} + \sigma^2 - \boldsymbol{\sigma}_3^\top \Sigma_1^{-1} \boldsymbol{\sigma}_3}{\sigma_{33} - \sigma_3^\top \Sigma_1^{-1} \boldsymbol{\sigma}_3} b_3.$$

Thus, according to Theorem 4.1, the CPN estimator is represented as follows:

$$\begin{aligned}
\hat{\beta}_0^{(\mathrm{CPN})} &= \hat{\beta}_0^{(\mathrm{PN})} + \left(\hat{\boldsymbol{\beta}}_1^{(\mathrm{PN})} - \hat{\boldsymbol{\beta}}_1^{(\mathrm{CPN})}\right)^\top \boldsymbol{\mu}_1 + \left(\hat{\beta}_3^{(\mathrm{PN})} - \hat{\beta}_3^{(\mathrm{CPN})}\right) \mu_3 \\
&\quad - \frac{1}{2}\left(\left(\hat{\boldsymbol{\beta}}_1^{(\mathrm{CPN})}\right)^\top \Sigma_1 \hat{\boldsymbol{\beta}}_1^{(\mathrm{CPN})} + \hat{\beta}_3^{(\mathrm{CPN})}\left(2\boldsymbol{\sigma}_3^\top \hat{\boldsymbol{\beta}}_1^{(\mathrm{CPN})} + \sigma_{33}\hat{\beta}_3^{(\mathrm{CPN})}\right)\right) \\
&\quad + \frac{1}{2}\left(\left(\hat{\boldsymbol{\beta}}_1^{(\mathrm{PN})}\right)^\top \Sigma_1 \hat{\boldsymbol{\beta}}_1^{(\mathrm{PN})} + \hat{\beta}_3^{(\mathrm{PN})}\left(2\boldsymbol{\sigma}_3^\top \hat{\boldsymbol{\beta}}_1^{(\mathrm{PN})} + \sigma_{33}\hat{\beta}_3^{(\mathrm{PN})}\right)\right) \\
&\quad + \frac{1}{2}\sigma^2\left(\hat{\beta}_3^{(\mathrm{PN})}\right)^2, \\
\hat{\boldsymbol{\beta}}_1^{(\mathrm{CPN})} &= \hat{\boldsymbol{\beta}}_1^{(\mathrm{PN})} - \frac{\sigma^2 \hat{\beta}_3^{(\mathrm{PN})}}{\sigma_{33} - \boldsymbol{\sigma_3}^\top \Sigma_1^{-1} \boldsymbol{\sigma}_3} \Sigma_1^{-1} \boldsymbol{\sigma}_3, \\
\hat{\beta}_3^{(\mathrm{CPN})} &= \frac{\sigma_{33} + \sigma^2 - \boldsymbol{\sigma}_3^\top \Sigma_1^{-1} \boldsymbol{\sigma}_3}{\sigma_{33} - \boldsymbol{\sigma}_3^\top \Sigma_1^{-1} \boldsymbol{\sigma}_3} \hat{\beta}_3^{(\mathrm{PN})}.
\end{aligned}$$

Example 4.4: We derive the CPN estimator assuming that

$$\begin{aligned}
&Y \mid \boldsymbol{X} \sim \mathrm{Po}\left(\exp\left(\boldsymbol{\beta}^\top \begin{pmatrix} 1 \\ \boldsymbol{X} \end{pmatrix}\right)\right), \quad \boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^\top, \quad \boldsymbol{X} = (X_1, X_2, X_3)^\top, \\
&\boldsymbol{X}_o = (X_1, X_2)^\top, \quad X_e = X_3, \quad W = X_3 + U, \\
&X_1 \sim \mathrm{Be}(p_1), \quad X_2 \sim \mathrm{Be}(p_2), \quad X_3 \sim \Gamma(k_1, \lambda), \quad U \sim \Gamma(k_2, \lambda),
\end{aligned}$$

where $0 < p_1 < 1, 0 < p_2 < 1, k_1 > 0, \lambda > 0, k_2 > 0$. We also assume X_1, X_2 and X_3 are mutually independent. We obtain

$$G_3(\beta_3, b_3) = \frac{k_1 + k_2}{\lambda - b_3} - \frac{k_2}{\lambda} - \frac{k_1}{\lambda - \beta_3}.$$

From $G_3 = 0$, we obtain the implicit function as

$$\beta_3 = \frac{(k_1 + k_2)\lambda b_3}{k_1 \lambda + k_2 b_3}.$$

Thus, according to Theorem 4.2, the CPN estimator is represented as follows:

$$\begin{aligned}
\hat{\beta}_0^{(\mathrm{CPN})} &= \hat{\beta}_0^{(\mathrm{PN})} - k_2 \log\left(1 - \hat{\beta}_3^{(\mathrm{PN})}/\lambda\right) + k_1 \log\left(\frac{\lambda - \hat{\beta}_3^{(\mathrm{CPN})}}{\lambda - \hat{\beta}_3^{(\mathrm{PN})}}\right), \\
\hat{\beta}_1^{(\mathrm{CPN})} &= \hat{\beta}_1^{(\mathrm{PN})},
\end{aligned}$$

$$\hat{\beta}_2^{(\mathrm{CPN})} = \hat{\beta}_2^{(\mathrm{PN})},$$
$$\hat{\beta}_3^{(\mathrm{CPN})} = \frac{(k_1+k_2)\lambda\hat{\beta}_3^{(\mathrm{PN})}}{k_1\lambda + k_2\hat{\beta}_3^{(\mathrm{PN})}}.$$

Example 4.5: We derive the CPN estimator assuming that

$$\begin{aligned}
&Y \mid \boldsymbol{X} \sim \mathrm{Po}\left(\exp\left(\boldsymbol{\beta}^\top \begin{pmatrix}1\\ \boldsymbol{X}\end{pmatrix}\right)\right), \quad \boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^\top, \quad \boldsymbol{X} = (X_1, X_2)^\top,\\
&\boldsymbol{X}_e = (X_1, X_2)^\top = \boldsymbol{X}, \quad W_1 = X_1 + U_1, \quad W_2 = X_2 + U_2,\\
&X_1 \sim \Gamma(k_1, \lambda), \quad X_2 \sim N(\mu, \sigma_x^2), \quad U_1 \sim \Gamma(k_2, \lambda), \quad U_2 \sim N(0, \sigma_u^2),\\
&X_1 \perp X_2, \quad U_1 \perp U_2,
\end{aligned}$$

where $k_1 > 0, \lambda > 0, k_2 > 0, 0 < \sigma_x^2 < \infty, 0 < \sigma_u^2 < \infty$. We obtain

$$G_j(\beta_j, b_j) = \begin{cases} \dfrac{k_1+k_2}{\lambda - b_j} - \dfrac{k_2}{\lambda} - \dfrac{k_1}{\lambda - \beta_j}, & j = 1,\\ (\sigma_x^2 + \sigma_u^2) b_j - \sigma_x^2 \beta_j, & j = 2. \end{cases}$$

From $G_j = 0$ $(j = 1, 2)$, we obtain the following implicit functions:

$$\beta_1 = \frac{(k_1+k_2)\lambda b_1}{k_1\lambda + k_2 b_1}, \quad \beta_2 = \frac{\sigma_x^2 + \sigma_u^2}{\sigma_x^2} b_2.$$

Thus, according to Theorem 4.2, the CPN estimator is represented as follows:

$$\begin{aligned}
\hat{\beta}_0^{(\mathrm{CPN})} &= \hat{\beta}_0^{(\mathrm{PN})} + k_1 \log\left(\frac{\lambda - \hat{\beta}_1^{(\mathrm{CPN})}}{\lambda - \hat{\beta}_1^{(\mathrm{PN})}}\right)\\
&\quad - k_2 \log\left(1 - \hat{\beta}_1^{(\mathrm{PN})}/\lambda\right) + \mu\left(\hat{\beta}_2^{(\mathrm{PN})} - \hat{\beta}_2^{(\mathrm{CPN})}\right)\\
&\quad + \frac{1}{2}(\sigma_x^2 + \sigma_u^2)\left(\hat{\beta}_2^{(\mathrm{PN})}\right)^2 - \frac{1}{2}\sigma_x^2\left(\hat{\beta}_2^{(\mathrm{CPN})}\right)^2,\\
\hat{\beta}_1^{(\mathrm{CPN})} &= \frac{(k_1+k_2)\lambda\hat{\beta}_1^{(\mathrm{PN})}}{k_1\lambda + k_2\hat{\beta}_1^{(\mathrm{PN})}},\\
\hat{\beta}_2^{(\mathrm{CPN})} &= \frac{\sigma_x^2 + \sigma_u^2}{\sigma_x^2}\hat{\beta}_2^{(\mathrm{PN})}.
\end{aligned}$$

5. Simulation studies

In this section, we present simulation studies that compare the performance of the PN and CPN estimators. We denote the sample size as n and the number of simulations as MC. We calculate the estimated bias for $\hat{\boldsymbol{\beta}}^{(\mathrm{PN})}$ and $\hat{\boldsymbol{\beta}}^{(\mathrm{CPN})}$ as follows:

$$\widehat{\mathrm{BIAS}}\left(\hat{\boldsymbol{\beta}}^{(\mathrm{PN})}\right) = \frac{1}{\mathrm{MC}} \sum_{i=1}^{\mathrm{MC}} \hat{\boldsymbol{\beta}}_i^{(\mathrm{PN})} - \boldsymbol{\beta},$$

$$\widehat{\mathrm{BIAS}}\left(\hat{\boldsymbol{\beta}}^{(\mathrm{CPN})}\right) = \frac{1}{\mathrm{MC}} \sum_{i=1}^{\mathrm{MC}} \hat{\boldsymbol{\beta}}_i^{(\mathrm{CPN})} - \boldsymbol{\beta},$$

where $\hat{\boldsymbol{\beta}}_i^{(\mathrm{PN})}$ and $\hat{\boldsymbol{\beta}}_i^{(\mathrm{CPN})}$ represent the PN and CPN estimators in the ith time simulation, respectively. Sampling from the joint distribution of $(\boldsymbol{X}, Y)$ involves first sampling $\boldsymbol{X}_i$ $(i = 1, \ldots, n)$ from the distribution of $\boldsymbol{X}$, and then sampling Y_i $(i = 1, \ldots, n)$ from the conditional distribution of $Y|\boldsymbol{X}_i$.

5.1. Case 1

We assume

$$\begin{aligned} &Y \mid \boldsymbol{X} \sim \mathrm{Po}\left(\exp\left(\boldsymbol{\beta}^\top \begin{pmatrix} 1 \\ \boldsymbol{X} \end{pmatrix}\right)\right), \quad \boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^\top, \quad \boldsymbol{X} = (X_1, X_2, X_3)^\top, \\ &\boldsymbol{X}_e = (X_1, X_2, X_3)^\top = \boldsymbol{X}, \quad \boldsymbol{W} = \boldsymbol{X} + \boldsymbol{U}, \\ &\boldsymbol{X} \sim N_3(\boldsymbol{\mu}, \Sigma_{\boldsymbol{X}}), \quad \boldsymbol{U} \sim N_3(\boldsymbol{0}_3, \Sigma_{\boldsymbol{U}}), \end{aligned}$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)^\top \in \mathbb{R}^3$, $\Sigma = [\sigma_{ij}] \in \mathbb{R}^{3\times 3}$. Let

$$\boldsymbol{\beta} = \begin{pmatrix} 0.1 \\ 0.1 \\ 0.2 \\ 0.3 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} 1 \\ 1.2 \\ 0.5 \end{pmatrix}, \quad \Sigma_{\boldsymbol{X}} = \begin{pmatrix} 1 & 0.2 & -0.5 \\ 0.2 & 1.1 & 0.3 \\ -0.5 & 0.3 & 1.2 \end{pmatrix}, \quad \Sigma_{\boldsymbol{U}} = I_3,$$

$n = 1000, \mathrm{MC} = 10{,}000$. As mentioned in Section 2.1, we assume that the true value of Σ_u is known. We estimate $\boldsymbol{\mu}$ and Σ_x using the method of moments in terms of $\boldsymbol{W}$:

$$\hat{\boldsymbol{\mu}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{W}_i, \quad \widehat{\Sigma}_{\boldsymbol{X}} = \frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{W}_i - \overline{\boldsymbol{W}})(\boldsymbol{W}_i - \overline{\boldsymbol{W}})^\top - \Sigma_{\boldsymbol{U}}.$$

In this case, we compare the performance of the PN and CPN estimators to that of corrected score (CS), simple structural (SS) and quasi-score (QS) estimators. These are consistent estimators for the true parameter discussed in Shklyar and Schneeweiss (2005). Table 1 lists the estimated and asymptotic biases of the estimators for the true $\boldsymbol{\beta}$. The bias of the PN estimator is corrected using the CPN estimator. The performance of the CPN estimator is non-inferior to the CS, SS and QS estimators. The existing methods discussed in Shklyar and Schneeweiss (2005) are limited to EIV models, in which all explanatory variables are measured with error, and $\boldsymbol{X}$ and $\boldsymbol{U}$ are restricted to a multivariate normal distribution. The more general CPN estimator demonstrates a comparable performance even in such a special case, indicating its effectiveness and broader applicability compared to the CS, SS and QS estimators.

5.2. Case 2

We assume

$$Y \mid \boldsymbol{X} \sim \mathrm{Po}\left(\exp\left(\boldsymbol{\beta}^\top \begin{pmatrix} 1 \\ \boldsymbol{X} \end{pmatrix}\right)\right), \quad \boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^\top, \quad \boldsymbol{X} = (X_1, X_2, X_3)^\top,$$

Table 1. Estimated and asymptotic theoretical bias for a multivariate normal distribution with multivariate normal error.

		PN	CPN	CS	SS	QS
$\hat{\beta}_0$	$\widehat{\text{BIAS}}$	0.2700	−0.0020	−0.0035	−0.0020	−0.0020
	Asy.Bias	0.2701	0	0	0	0
$\hat{\beta}_1$	$\widehat{\text{BIAS}}$	−0.0797	0.0008	0.0014	0.0008	0.0008
	Asy.Bias	−0.0796	0	0	0	0
$\hat{\beta}_2$	$\widehat{\text{BIAS}}$	−0.0670	−0.0003	−0.0001	−0.0003	−0.0003
	Asy.Bias	−0.0669	0	0	0	0
$\hat{\beta}_3$	$\widehat{\text{BIAS}}$	−0.1453	0.0012	0.0021	0.0012	0.0012
	Asy.Bias	−0.1453	0	0	0	0

$$\boldsymbol{X}_o = (X_1, X_2)^\top, \quad X_e = X_3, \quad W = X_3 + U,$$
$$\boldsymbol{X} \sim N_3(\boldsymbol{\mu}, \Sigma), \quad U \sim N(0, \sigma^2),$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)^\top \in \mathbb{R}^3$, $\Sigma = [\sigma_{ij}] \in \mathbb{R}^{3\times3}$, $0 < \sigma^2 < \infty$. Let

$$\boldsymbol{\beta} = \begin{pmatrix} 0.1 \\ 0.1 \\ 0.2 \\ 0.3 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} 1 \\ 1.2 \\ 0.5 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & 0.2 & -0.5 \\ 0.2 & 1.1 & 0.3 \\ -0.5 & 0.3 & 1.2 \end{pmatrix},$$

$n = 5000$, MC $= 10{,}000$. We performed the simulation using three different values of σ^2: 0.25, 0.5, 1. As in Case 1, we assume that the true value of σ^2 is known. We estimate $\boldsymbol{\mu}$ and Σ within the CPN estimator using the method of moments in terms of $\boldsymbol{X}_o = (X_1, X_2)^\top$ and W because the value of X_3 is not directly observable.

$$\hat{\boldsymbol{\mu}}_1 = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_{o,i}, \quad \hat{\mu}_3 = \frac{1}{n}\sum_{i=1}^{n} w_i,$$
$$\widehat{\Sigma}_1 = \frac{1}{n}\sum_{i=1}^{n} (\boldsymbol{x}_{o,i} - \bar{\boldsymbol{x}}_o)(\boldsymbol{x}_{o,i} - \bar{\boldsymbol{x}}_o)^\top,$$
$$\hat{\boldsymbol{\sigma}} = \frac{1}{n}\sum_{i=1}^{n} (\boldsymbol{x}_{o,i} - \bar{\boldsymbol{x}}_o)(w_i - \bar{w}), \quad \hat{\sigma}_{33} = \frac{1}{n}\sum_{i=1}^{n} (w_i - \bar{w})^2 - \sigma^2,$$

where $(\boldsymbol{x}_{o,i}, w_i)$ $(i = 1, \ldots, n)$ are samples of the distributions of $(\boldsymbol{X}_o, W)$.

Table 2 lists the estimated and asymptotic biases of the estimators for the true $\boldsymbol{\beta}$. Bias correction of the PN estimator was performed using the CPN estimator. Its bias increases with increasing σ^2. However, the bias of the CPN estimator is small for large σ^2.

5.3. Case 3

We assume

$$Y \mid \boldsymbol{X} \sim \text{Po}\left(\exp\left(\boldsymbol{\beta}^\top \begin{pmatrix} 1 \\ \boldsymbol{X} \end{pmatrix}\right)\right), \quad \boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^\top, \quad \boldsymbol{X} = (X_1, X_2, X_3)^\top,$$
$$\boldsymbol{X}_o = (X_1, X_2)^\top, \quad X_e = X_3, \quad W = X_3 + U,$$

Table 2. Estimated and asymptotic theoretical bias for a multivariate normal distribution with a normal error.

		0.25		0.5		1	1
	σ^2	PN	CPN	PN	CPN	PN	CPN
$\hat{\beta}_0$	$\widehat{\text{BIAS}}$	0.0534	0.0003	0.0856	−0.0003	0.1234	−0.0008
	Asy.Bias	0.0531	0	0.0857	0	0.1238	0
$\hat{\beta}_1$	$\widehat{\text{BIAS}}$	−0.0410	0.0001	−0.0664	0.0001	−0.0959	0.0004
	Asy.Bias	−0.0411	0	−0.0664	0	−0.0960	0
$\hat{\beta}_2$	$\widehat{\text{BIAS}}$	0.0268	−0.0002	0.0435	−0.0002	0.0629	−0.0004
	Asy.Bias	0.0270	0	0.0436	0	0.0629	0
$\hat{\beta}_3$	$\widehat{\text{BIAS}}$	−0.0716	−0.0001	−0.1155	0.0002	−0.1666	0.0008
	Asy.Bias	−0.0715	0	−0.1155	0	−0.1668	0

Table 3. Estimated and asymptotic theoretical bias for Bernoulli and gamma distributions with a gamma error.

		1.125		1.6		2.25	
	k_2	PN	CPN	PN	CPN	PN	CPN
$\hat{\beta}_0$	$\widehat{\text{BIAS}}$	−0.0878	0.0058	−0.1159	0.0051	−0.1445	0.0063
	Asy.Bias	−0.0942	0	−0.1215	0	−0.1515	0
$\hat{\beta}_1$	$\widehat{\text{BIAS}}$	−0.0002	−0.0002	0.0002	0.0002	−0.0005	−0.0005
	Asy.Bias	0	0	0	0	0	0
$\hat{\beta}_2$	$\widehat{\text{BIAS}}$	0.0000	0.0000	−0.0003	−0.0003	0.0000	0.0000
	Asy.Bias	0	0	0	0	0	0
$\hat{\beta}_3$	$\widehat{\text{BIAS}}$	−0.1298	−0.0027	−0.1684	−0.0025	−0.2125	−0.0029
	Asy.Bias	−0.1276	0	−0.1665	0	−0.2104	0

$$X_1 \sim \mathrm{Be}(p_1), \quad X_2 \sim \mathrm{Be}(p_2), \quad X_3 \sim \Gamma(k_1, \lambda), \quad U \sim \Gamma(k_2, \lambda),$$

where $0 < p_1 < 1, 0 < p_2 < 1, k_1 > 0, \lambda > 0, k_2 > 0$. We also assume X_1, X_2, and X_3 are mutually independent. Let

$$\boldsymbol{\beta} = (0.5, 0.1, 0.2, 0.6)^\top, p_1 = 0.4, p_2 = 0.55, k_1 = 2.5, \lambda = 1.5,$$

$n = 5000, \mathrm{MC} = 10{,}000$. We perform the simulation using three different values for k_2: 1.125, 1.6, 2.25. The true value of k_2 is assumed to be known. We estimate k_1 and λ in the formula of the CPN estimator using the method of moments in terms of W (the value of X_3 is not directly observable).

$$\hat{k}_1 = \left(\frac{1}{n}\sum_{i=1}^{n} w_i\right)\hat{\lambda} - k_2, \quad \hat{\lambda} = \frac{\frac{1}{n}\sum_{i=1}^{n} w_i}{\frac{1}{n}\sum_{i=1}^{n}(w_i - \bar{w})^2},$$

where w_i $(i = 1, \ldots, n)$ are samples of the distribution of W.

Table 3 lists the estimated and asymptotic biases of the estimators for the true $\boldsymbol{\beta}$. Bias correction of the PN estimator was performed using the CPN estimator.

5.4. Case 4

We assume

$$Y \mid \boldsymbol{X} \sim \mathrm{Po}\left(\exp\left(\boldsymbol{\beta}^\top \begin{pmatrix} 1 \\ \boldsymbol{X} \end{pmatrix}\right)\right), \quad \boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^\top, \quad \boldsymbol{X} = (X_1, X_2)^\top,$$

Table 4. Estimated and asymptotic theoretical bias for gamma and normal distributions with gamma and normal errors.

		(0.36,0.25)		(0.72,0.5)		(1.44,1)	
	(k_2, σ_u^2)	PN	CPN	PN	CPN	PN	CPN
$\hat{\beta}_0$	BIAS	0.0302	0.0003	0.0549	−0.0002	0.0956	0.0008
	Asy.Bias	0.0299	0	0.0551	0	0.0949	0
$\hat{\beta}_1$	BIAS	0.07772	−0.0002	0.1298	0.0001	0.1957	−0.0010
	Asy.Bias	0.0774	0	0.1297	0	0.1959	0
$\hat{\beta}_2$	BIAS	−0.0334	−0.0001	−0.0601	0.0000	−0.1002	−0.0001
	Asy.Bias	−0.0333	0	−0.0600	0	−0.1000	0

$$\begin{aligned}
&\boldsymbol{X}_e = (X_1, X_2)^\top = \boldsymbol{X}, \quad W_1 = X_1 + U_1, \quad W_2 = X_2 + U_2, \\
&X_1 \sim \Gamma(k_1, \lambda), \quad X_2 \sim N(\mu, \sigma_x^2), \quad U_1 \sim \Gamma(k_2, \lambda), \quad U_2 \sim N(0, \sigma_u^2), \\
&X_1 \perp X_2, \quad U_1 \perp U_2,
\end{aligned}$$

where $k_1 > 0, \lambda > 0, k_2 > 0, 0 < \sigma_x^2 < \infty, 0 < \sigma_u^2 < \infty$. Let

$$\boldsymbol{\beta} = \left(0.5, -0.4, 0.3\right)^\top, k_1 = 2, \lambda = 1.2, \mu = 1, \sigma_x^2 = 2,$$

$n = 5000, \mathrm{MC} = 10{,}000$. The simulation was performed in three ways.

$$(k_2, \sigma_u^2) = (0.36, 0.25), (0.72, 0.5), (1.44, 1).$$

However, we assume that the true values of k_2 and σ_u^2 are known. We estimate k_1, λ, μ, and σ_x^2 in the CPN estimator using the method of moments in terms of $\boldsymbol{W} = (W_1, W_2)^\top$ (the values of X_1 and X_2 are not directly observable).

$$\begin{aligned}
\hat{k}_1 &= \left(\frac{1}{n}\sum_{i=1}^{n} w_{1,i}\right)\hat{\lambda} - k_2, \quad \hat{\lambda} = \frac{\frac{1}{n}\sum_{i=1}^{n} w_{1,i}}{\frac{1}{n}\sum_{i=1}^{n}(w_{1,i} - \bar{w}_1)^2}, \\
\hat{\mu} &= \frac{1}{n}\sum_{i=1}^{n} w_{2,i}, \quad \hat{\sigma}_x^2 = \frac{1}{n}\sum_{i=1}^{n}(w_{2,i} - \bar{w}_2)^2 - \sigma_u^2,
\end{aligned}$$

where $\boldsymbol{w}_i = (w_{1,i}, w_{2,i})^\top$ $(i = 1, \ldots, n)$ are samples of the distribution of $\boldsymbol{W}$.

Table 4 lists the estimated and asymptotic biases of the estimators for the true $\boldsymbol{\beta}$. As before, the bias of the CPN estimator was smaller than that of the PN estimator in all cases.

6. Real data analysis

In this section, we apply the PN and CPN estimators to real data as discussed in Wada and Kurosawa (2023). We use financial data collected in the FinAccess survey conducted in 2019, provided by Kenya National Bureau of Statistics (2019). In this study, we focus on the values labelled as finhealthscore, Mobile Ownership, Formal Prudential, and Normalized Household weights. The sample size is $N = 8669$. Details of the features used in this section, such as their types and descriptions, are provided in Table 5. We use finhealthscore as an objective variable Y, Mobile Ownership as an explanatory variable X_1, Formal Prudential as

Table 5. Details of the variables.

Features	Type	Description
finhealthscore	Count	Financial health score for households.
Mobile Ownership	Binary	Indicator of whether the respondent household owns a mobile or not.
Formal Prudential	Binary	Indicator of whether the respondent household spends money prudently or not.
Normalized Household Weights	Continuous	Weighted and normalized households.

Table 6. Estimates of ϕ, R_{McF}, m_{pp}, and ECD.

$\hat{\phi}$	$\widehat{R}_{\text{McF}}$	$\hat{m}_{pp}$	$\widehat{ECD}$
1.1670	0.4829	0.3183	0.2415

an explanatory variable X_2, and normalized household weights as explanatory variables X_3. The true model is assumed to be as follows:

$$Y \mid \boldsymbol{X} \sim \text{Po}\left(\exp\left(\boldsymbol{\beta}^\top \begin{pmatrix} 1 \\ \boldsymbol{X} \end{pmatrix}\right)\right), \quad \boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^\top, \quad \boldsymbol{X} = (X_1, X_2, X_3)^\top.$$

We further assume that the true parameter $\boldsymbol{\beta}$ is obtained from the estimates of all N data points.

As a diagnostic technique, we calculate goodness-of-fit measures to verify that the dataset follows a Poisson regression model. Table 6 lists the estimates of ϕ, R_{McF} in McFadden (1974), m_{pp} in Kurosawa et al. (2020) and the ECD in Eshima and Tabata (2010). Overdispersion is present because the estimate of ϕ is greater than one. In Wada and Kurosawa (2023), for a univariate case, R_{McF} was reported to be 0.4478. By contrast, R_{McF} in this study is improved by the Poisson regression model with multivariate explanatory variables, which allows us to apply a more appropriate Poisson regression model to the dataset. The estimated value of $\boldsymbol{\beta}$ is $(0.6308, 0.3356, 0.4520, 0.0897)^\top$ and we regard the estimate as the true value.

According to Kenya National Bureau of Statistics (2019), the data from the FinAccess survey were weighted and adjusted for non-responses to obtain a representative dataset at the national and county levels. Thus, we may consider a situation in which X_3 exhibits stochastic error U as

$$\boldsymbol{X}_o = (X_1, X_2)^\top, \quad X_e = X_3, \quad W = X_3 + U.$$

We assume a positive error because the distribution of normalized household weights is positive. Thus, we assume

$$X_1 \sim \text{Be}(p_1), \quad X_2 \sim \text{Be}(p_2), \quad X_3 \sim \Gamma(k_1, \lambda), \quad U \sim \Gamma(k_2, \lambda).$$

We also assume X_1, X_2, and X_3 are mutually independent. This setting is the same as that of the application in Example 4.4. We obtain the estimates of k_1 and λ as $k_1 = 2.0746, \quad \lambda = 2.0746$ and estimate the true parameter with $k_2 = k_1/3, 2k_1/3, k_1$. We take 2000 random samples from all N samples to obtain the $\boldsymbol{\beta}$ estimates. We repeat the estimations for MC = 10,000

Table 7. Estimated and asymptotic theoretical bias in financial data.

		$k_1/3$		$2k_1/3$		k_1	
	k_2	PN	CPN	PN	CPN	PN	CPN
$\hat{\beta}_0$	$\widehat{\text{BIAS}}$	−0.0082	−0.0073	−0.0131	−0.0117	−0.0172	−0.0154
	Asy.Bias	−0.0008	0	−0.0013	0	−0.0016	0
$\hat{\beta}_1$	$\widehat{\text{BIAS}}$	0.0057	0.0057	0.0073	0.0073	0.0099	0.0099
	Asy.Bias	0	0	0	0	0	0
$\hat{\beta}_2$	$\widehat{\text{BIAS}}$	0.0119	0.0119	0.0184	0.0184	0.0248	0.0248
	Asy.Bias	0	0	0	0	0	0
$\hat{\beta}_3$	$\widehat{\text{BIAS}}$	−0.2364	0.0066	−0.3755	0.0215	−0.4753	0.0248
	Asy.Bias	−0.2418	0	−0.3894	0	−0.4890	0

iterations to obtain the Monte Carlo mean of $\boldsymbol{\beta}$s. The bias is calculated by the difference between the Monte Carlo mean and the true value. Furthermore, to clarify the magnitude of the bias, we divide the bias by the absolute value of the true parameter. Table 7 lists the estimated biases calculated from the MC simulations. The estimated biases of the PN and CPN estimators are equal for β_1 and β_2 because X_1 and X_2 are observable. The estimated biases of β_1 and β_2 are close to 0, which is a typical value for their asymptotic biases. In addition, the estimated biases of the CPN estimator for β_0 and β_3 are smaller than those of the PN estimator in all cases.

Funding

This work was supported by JSPS KAKENHI Grant Number JP25K07136.

Disclosure statement

The authors declare no conflicts of interest.

References

Box, G. E. P. (1963). The effects of errors in the factor levels and experimental design. *Technometrics*, *5*(2), 247–262. https://doi.org/10.1080/00401706.1963.10490078

Burr, D. (1988). On errors-in-variables in binary regression-Berkson case. *Journal of the American Statistical Association*, *83*(403), 739–743.

Eshima, N., & Tabata, M. (2010). Entropy coefficient of determination for generalized linear models. *Computational Statistics & Data Analysis*, *54*(5), 1381–1389. https://doi.org/10.1016/j.csda.2009.12.003

Fuller, W. A. (1987). *Measurement Error Models*. Wiley.

Geary, R. C. (1953). Non-linear functional relationship between two variables when one variable is controlled. *Journal of the American Statistical Association*, *48*(261), 94–103. https://doi.org/10.1080/01621459.1953.10483458

Guo, J. Q., & Li, T. (2002). Poisson regression models with errors-in-variables: Implication and treatment. *Journal of Statistical Planning and Inference*, *104*(2), 391–401. https://doi.org/10.1016/S0378-3758(01)00250-6

Huwang, L., & Huang, Y. H. (2000). On errors-in-variables in polynomial regression-Berkson case. *Statistica Sinica*, *10*(3), 923–936.

Jiang, F., & Ma, Y. (2022). Poisson regression with error corrupted high dimensional features. *Statistica Sinica*, *32*(4), 2023–2046.

Kenya National Bureau of Statistics (2019). *FinAccess Survey 2019*. https://knbs.or.ke

Kukush, A., & Schneeweiss, H. (2000). A comparison of asymptotic covariance matrices of adjusted least squares and structural least squares in error ridden polynomial regression. *Sonderforschungsbereich*, *386*, 1–23.

Kukush, A., Schneeweiss, H., & Wolf, R. (2004). Three estimators for the Poisson regression model with measurement errors. *Statistical Papers*, *45*(3), 351–368. https://doi.org/10.1007/BF02777577

Kukush, A., & Shklyar, S. (2002). Comparison of three estimators in Poisson errors-in-variables model with one covariate. *Sonderforschungsbereich*, *386*, 1–23.

Kurosawa, T., Hui, F. K. C., Welsh, A. H., Shinmura, K., & Eshima, N. (2020). On goodness-of-fit measures for Poisson regression models. *Australian & New Zealand Journal of Statistics*, *62*(3), 340–366. https://doi.org/10.1111/anzs.v62.3

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics* (pp. 105–142). Academic Press.

Nakamura, T. (1990). Corrected score function for errors-in-variables models: Methodology and application to generalized linear models. *Biometrika*, *77*(1), 127–137. https://doi.org/10.1093/biomet/77.1.127

Rosenthal, J. S. (2025). *A First Look at Rigorous Probability Theory* (2nd ed.). World Scientific Publishing.

Rudin, W. (1976). *Principles of Mathematical Analysis* (3rd ed.). McGraw-Hill.

Shklyar, S., & Schneeweiss, H. (2005). A comparison of asymptotic covariance matrices of three consistent estimators in the Poisson regression model with measurement errors. *Journal of Multivariate Analysis*, *94*(2), 250–270. https://doi.org/10.1016/j.jmva.2004.05.002

Shklyar, S., Schneeweiss, H., & Kukush, A. (2007). Quasi score is more efficient than corrected score in a polynomial measurement error model. *Metrika*, *65*(3), 275–295. https://doi.org/10.1007/s00184-006-0076-5

van der Vaart, A. W. (2012). *Asymptotic Statistics*. Cambridge University Press.

Wada, K., & Kurosawa, T. (2023). The naive estimator of a Poisson regression model with a measurement error. *Journal of Risk and Financial Management*, *16*(3), 16–30. https://doi.org/10.3390/jrfm16030186

Appendices

Appendix 1. Proofs of Theorems

A.1 Proof of Theorem 3.1

Proof: First, we derive an expression for $\boldsymbol{b}$. From (5), the following equations are obtained.

$$\begin{aligned} \mathbf{E}[Y] &= \mathbf{E}[\exp(b_0 + \boldsymbol{b}_1^\top \boldsymbol{X}_o + \boldsymbol{b}_2^\top \boldsymbol{W})], \\ \mathbf{E}[Y\boldsymbol{X}_o] &= \mathbf{E}[\boldsymbol{X}_o \exp(b_0 + \boldsymbol{b}_1^\top \boldsymbol{X}_o + \boldsymbol{b}_2^\top \boldsymbol{W})], \\ \mathbf{E}[Y\boldsymbol{W}] &= \mathbf{E}[\boldsymbol{W} \exp(b_0 + \boldsymbol{b}_1^\top \boldsymbol{X}_o + \boldsymbol{b}_2^\top \boldsymbol{W})]. \end{aligned} \tag{A1}$$

From (6) and these equalities, we have

$$\begin{aligned} &\mathrm{e}\beta_0 M_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix} = \mathrm{e}^{b_0} M_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{pmatrix} M_{\boldsymbol{U}}(\boldsymbol{b}_2), \\ &\mathrm{e}^{\beta_0}\frac{\partial}{\partial\boldsymbol{\beta}_1} M_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix} = \mathrm{e}^{b_0}\frac{\partial}{\partial\boldsymbol{b}_1} M_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{pmatrix} M_{\boldsymbol{U}}(\boldsymbol{b}_2), \\ &\mathrm{e}^{\beta_0}\mathbf{E}[\boldsymbol{U}] M_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix} + \mathrm{e}^{\beta_0}\frac{\partial}{\partial\boldsymbol{\beta}_2} M_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix} = \mathrm{e}^{b_0}\frac{\partial}{\partial\boldsymbol{b}_2} M_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{pmatrix} M_{\boldsymbol{U}}(\boldsymbol{b}_2) \\ &\qquad + \mathrm{e}^{b_0} M_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{pmatrix}\frac{\partial}{\partial\boldsymbol{b}_2} M_{\boldsymbol{U}}(\boldsymbol{b}_2). \end{aligned}$$

Therefore, we apply a transformation to obtain the following system of equations:

$$\begin{aligned} b_0 &= \beta_0 + \log\left(\frac{M_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix}}{M_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{pmatrix}M_{\boldsymbol{U}}(\boldsymbol{b}_2)}\right),\\ \frac{\partial}{\partial\boldsymbol{\beta}_1}K_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix} &= \frac{\partial}{\partial\boldsymbol{b}_1}K_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{pmatrix},\\ \mathbf{E}[\boldsymbol{U}] + \frac{\partial}{\partial\boldsymbol{\beta}_2}K_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix} &= \frac{\partial}{\partial\boldsymbol{b}_2}K_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{pmatrix} + \frac{\partial}{\partial\boldsymbol{b}_2}K_{\boldsymbol{U}}(\boldsymbol{b}_2). \end{aligned} \tag{A2}$$

Thus, $\boldsymbol{b} = (b_0, \boldsymbol{b}_1^\top, \boldsymbol{b}_2^\top)^\top$ is determined by solving the following system of equations:

$$\begin{aligned} \frac{\partial}{\partial\boldsymbol{\beta}_1}K_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix} &= \frac{\partial}{\partial\boldsymbol{b}_1}K_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{pmatrix},\\ \mathbf{E}[\boldsymbol{U}] + \frac{\partial}{\partial\boldsymbol{\beta}_2}K_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix} &= \frac{\partial}{\partial\boldsymbol{b}_2}K_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{pmatrix} + \frac{\partial}{\partial\boldsymbol{b}_2}K_{\boldsymbol{U}}(\boldsymbol{b}_2). \end{aligned}$$

Here, we set

$$\boldsymbol{G}\left(\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix}, \begin{pmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{pmatrix}\right) := \begin{pmatrix} \frac{\partial}{\partial\boldsymbol{b}_1}K_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{pmatrix} - \frac{\partial}{\partial\boldsymbol{\beta}_1}K_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix}\\ \frac{\partial}{\partial\boldsymbol{b}_2}K_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{pmatrix} + \frac{\partial}{\partial\boldsymbol{b}_2}K_{\boldsymbol{U}}(\boldsymbol{b}_2) - \mathbf{E}[\boldsymbol{U}] - \frac{\partial}{\partial\boldsymbol{\beta}_2}K_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix} \end{pmatrix}.$$

From the definition of $\boldsymbol{b}$, $\boldsymbol{G}$ is always $\mathbf{0}$ in $\mathbb{R}^{2(p+q)}$. In addition, $\boldsymbol{G}$ is continuously differentiable, because we assume the existence of (7). We assume $\boldsymbol{G}$ satisfies

$$\det\frac{\partial\boldsymbol{G}}{\partial\begin{pmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{pmatrix}^\top} = \det\begin{pmatrix} \frac{\partial^2}{\partial\boldsymbol{b}_1^\top\partial\boldsymbol{b}_1}K_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{pmatrix} & \frac{\partial^2}{\partial\boldsymbol{b}_2^\top\partial\boldsymbol{b}_1}K_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{pmatrix}\\ \frac{\partial^2}{\partial\boldsymbol{b}_1^\top\partial\boldsymbol{b}_2}K_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{pmatrix} & \frac{\partial^2}{\partial\boldsymbol{b}_2^\top\partial\boldsymbol{b}_2}K_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{pmatrix} + \frac{\partial^2}{\partial\boldsymbol{b}_2^\top\partial\boldsymbol{b}_2}K_{\boldsymbol{U}}(\boldsymbol{b}_2) \end{pmatrix} \neq 0.$$

Then, according to the Implicit Function Theorem (see for example Rudin, 1976), there exists a unique C^1-class function $\boldsymbol{g}(\boldsymbol{x}) = (g_1(\boldsymbol{x}), \ldots, g_{p+q}(\boldsymbol{x}))^\top$ that satisfies

$$\begin{pmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{pmatrix} = \boldsymbol{g}\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix}$$

in the neighbourhood of the zeros of $\boldsymbol{G}$. Next, we describe the near-certain convergence of the PN estimator. Based on the strong law of large numbers, we obtain

$$\begin{aligned} S_n\left(\hat{\boldsymbol{\beta}}^{(\mathrm{PN})} \mid \mathcal{X}\right) &= \frac{1}{n}\sum_{i=1}^{n}\left\{Y_i - \exp\left(\hat{\beta}_0^{(\mathrm{PN})} + \hat{\boldsymbol{\beta}}_1^{(\mathrm{PN})^\top}\boldsymbol{X}_{o,i} + \hat{\boldsymbol{\beta}}_2^{(\mathrm{PN})^\top}\boldsymbol{W}_i\right)\right\}(1, \boldsymbol{X}_{o,i}^\top, \boldsymbol{W}_i^\top)^\top,\\ &\xrightarrow{\text{a.s.}} \mathbf{E}_{\boldsymbol{X},\boldsymbol{W}}\left[\mathbf{E}_{Y|(\boldsymbol{X},\boldsymbol{W})}\left[\left\{Y - \exp\left(\hat{\beta}_0^{(\mathrm{PN})} + \hat{\boldsymbol{\beta}}_1^{(\mathrm{PN})^\top}\boldsymbol{X}_o + \hat{\boldsymbol{\beta}}_2^{(\mathrm{PN})^\top}\boldsymbol{W}\right)\right\}\right.\right.\\ &\qquad \left.\left.\times(1, \boldsymbol{X}_o^\top, \boldsymbol{W}^\top)\right]\right]. \end{aligned} \tag{A3}$$

From (A3) and the same argument as in Kukush and Shklyar (2002), we obtain

$$\hat{\boldsymbol{\beta}}^{(\mathrm{PN})} \xrightarrow{\text{a.s.}} \boldsymbol{b}. \tag{A4}$$

From (A4) and the uniform integrability using the Vitali convergence theorem (Rosenthal, 2025), the asymptotic bias of the PN estimator is as follows:

$$\lim_{n\to\infty}\mathbf{E}\left[\hat{\beta}_0^{(\mathrm{PN})} - \beta_0\right] = \mathbf{E}\left[\lim_{n\to\infty}\hat{\beta}_0^{(\mathrm{PN})} - \beta_0\right] = b_0 - \beta_0$$

$$= \log \left(\frac{M_{\boldsymbol{X}}\left(\begin{smallmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{smallmatrix}\right)}{M_{\boldsymbol{X}}\left(\boldsymbol{g}\left(\begin{smallmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{smallmatrix}\right)\right) M_{\boldsymbol{U}}\left(\boldsymbol{g}_2\left(\begin{smallmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{smallmatrix}\right)\right)} \right),$$

$$\lim_{n\to\infty} \mathbf{E}\left[\begin{pmatrix}\hat{\boldsymbol{\beta}}_1^{(\mathrm{PN})} - \boldsymbol{\beta}_1\\ \hat{\boldsymbol{\beta}}_2^{(\mathrm{PN})} - \boldsymbol{\beta}_2\end{pmatrix}\right] = \mathbf{E}\left[\lim_{n\to\infty}\begin{pmatrix}\hat{\boldsymbol{\beta}}_1^{(\mathrm{PN})} - \boldsymbol{\beta}_1\\ \hat{\boldsymbol{\beta}}_2^{(\mathrm{PN})} - \boldsymbol{\beta}_2\end{pmatrix}\right] = \begin{pmatrix}\boldsymbol{b}_1 - \boldsymbol{\beta}_1\\ \boldsymbol{b}_2 - \boldsymbol{\beta}_2\end{pmatrix}$$

$$= \boldsymbol{g}\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix} - \begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix},$$

where $\boldsymbol{g}_2(\boldsymbol{x}) = (g_{p+1}(\boldsymbol{x}), \ldots, g_{p+q}(\boldsymbol{x}))^\top$ denotes a subvector of $\boldsymbol{g}$. We also derive the asymptotic MSE of the PN estimator. The asymptotic MSE of $\hat{\beta}_0^{(\mathrm{PN})}$ is obtained by uniform integrability using the Vitali convergence theorem as follows:

$$\begin{aligned}
\lim_{n\to\infty} \mathbf{E}\left[\left(\hat{\beta}_0^{(\mathrm{PN})} - \beta_0\right)^2\right] &= \lim_{n\to\infty} \mathbf{E}\left[\left(\hat{\beta}_0^{(\mathrm{PN})} - b_0 + b_0 - \beta_0\right)^2\right] \\
&= (b_0 - \beta_0)^2 + 2(b_0 - \beta_0) \lim_{n\to\infty} \mathbf{E}\left[\hat{\beta}_0^{(\mathrm{PN})} - b_0\right] \\
&\quad + \lim_{n\to\infty} \mathbf{E}\left[\left(\hat{\beta}_0^{(\mathrm{PN})} - b_0\right)^2\right] \\
&= (b_0 - \beta_0)^2.
\end{aligned}$$

For $(\hat{\boldsymbol{\beta}}_1^{(\mathrm{PN})}, \hat{\boldsymbol{\beta}}_2^{(\mathrm{PN})})^\top$, we obtain the following by the same argument.

$$\begin{aligned}
&\lim_{n\to\infty} \mathbf{E}\left[\begin{pmatrix}\hat{\boldsymbol{\beta}}_1^{(\mathrm{PN})} - \boldsymbol{\beta}_1\\ \hat{\boldsymbol{\beta}}_2^{(\mathrm{PN})} - \boldsymbol{\beta}_2\end{pmatrix}\begin{pmatrix}\hat{\boldsymbol{\beta}}_1^{(\mathrm{PN})} - \boldsymbol{\beta}_1\\ \hat{\boldsymbol{\beta}}_2^{(\mathrm{PN})} - \boldsymbol{\beta}_2\end{pmatrix}^\top\right] \\
&= \begin{pmatrix}\boldsymbol{b}_1 - \boldsymbol{\beta}_1\\ \boldsymbol{b}_2 - \boldsymbol{\beta}_2\end{pmatrix}\begin{pmatrix}\boldsymbol{b}_1 - \boldsymbol{\beta}_1\\ \boldsymbol{b}_2 - \boldsymbol{\beta}_2\end{pmatrix}^\top \\
&= \left(\boldsymbol{g}\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix} - \begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix}\right)\left(\boldsymbol{g}\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix} - \begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix}\right)^\top.
\end{aligned}$$

■

A.2 Proof of Theorem 4.1

Proof: From (A2), we obtain the following system of equations:

$$\begin{aligned}
\beta_0 &= b_0 + \log\left(\frac{M_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{pmatrix} M_{\boldsymbol{U}}(\boldsymbol{b}_2)}{M_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix}}\right), \\
\boldsymbol{G} &= \begin{pmatrix} \dfrac{\partial}{\partial \boldsymbol{b}_1} K_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{pmatrix} - \dfrac{\partial}{\partial \boldsymbol{\beta}_1} K_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix} \\ \dfrac{\partial}{\partial \boldsymbol{b}_2} K_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{pmatrix} + \dfrac{\partial}{\partial \boldsymbol{b}_2} K_{\boldsymbol{U}}(\boldsymbol{b}_2) - \mathbf{E}[\boldsymbol{U}] - \dfrac{\partial}{\partial \boldsymbol{\beta}_2} K_{\boldsymbol{X}}\begin{pmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{pmatrix}\end{pmatrix} = \mathbf{0}_{p+q}.
\end{aligned} \tag{A5}$$

Following the same argument utilized in the proof of Theorem 3.1, $\boldsymbol{G}$ is always $\mathbf{0}$ and is continuously differentiable. We assume the following:

$$\det \frac{\partial \boldsymbol{G}}{\partial \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}^\top} = (-1)^{p+q} \det \begin{pmatrix} \frac{\partial^2}{\partial \boldsymbol{\beta}_1^\top \partial \boldsymbol{\beta}_1} K_{\boldsymbol{X}} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} & \frac{\partial^2}{\partial \boldsymbol{\beta}_2^\top \partial \boldsymbol{\beta}_1} K_{\boldsymbol{X}} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} \\ \frac{\partial^2}{\partial \boldsymbol{\beta}_1^\top \partial \boldsymbol{\beta}_2} K_{\boldsymbol{X}} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} & \frac{\partial^2}{\partial \boldsymbol{\beta}_2^\top \partial \boldsymbol{\beta}_2} K_{\boldsymbol{X}} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} \end{pmatrix} \neq 0.$$

Then, according to the Implicit Function Theorem, there exists a unique C^1-class function $\boldsymbol{h}$ that satisfies

$$\begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} = \boldsymbol{h} \begin{pmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \end{pmatrix}$$

in the neighbourhood of the zeros of $\boldsymbol{G}$. By replacing $\boldsymbol{b} = (b_0, \boldsymbol{b}_1^\top, \boldsymbol{b}_2^\top)^\top$ with the PN estimator $\hat{\boldsymbol{\beta}}^{(\mathrm{PN})} = (\hat{\beta}_0^{(\mathrm{PN})}, (\hat{\boldsymbol{\beta}}_1^{(\mathrm{PN})})^\top, (\hat{\boldsymbol{\beta}}_2^{(\mathrm{PN})})^\top)^\top$ in the solution for $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$, we obtain the CPN estimator as

$$\hat{\beta}_0^{(\mathrm{CPN})} = \hat{\beta}_0^{(\mathrm{PN})} + \log \left(\frac{M_{\boldsymbol{X}} \begin{pmatrix} \hat{\boldsymbol{\beta}}_1^{(\mathrm{PN})} \\ \hat{\boldsymbol{\beta}}_2^{(\mathrm{PN})} \end{pmatrix} M_{\boldsymbol{U}}(\hat{\boldsymbol{\beta}}_2^{(\mathrm{PN})})}{M_{\boldsymbol{X}} \begin{pmatrix} \hat{\boldsymbol{\beta}}_1^{(\mathrm{CPN})} \\ \hat{\boldsymbol{\beta}}_2^{(\mathrm{CPN})} \end{pmatrix}} \right),$$

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}_1^{(\mathrm{CPN})} \\ \hat{\boldsymbol{\beta}}_2^{(\mathrm{CPN})} \end{pmatrix} = \boldsymbol{h} \begin{pmatrix} \hat{\boldsymbol{\beta}}_1^{(\mathrm{PN})} \\ \hat{\boldsymbol{\beta}}_2^{(\mathrm{PN})} \end{pmatrix}.$$

Here, we have the almost sure convergence of $\hat{\boldsymbol{\beta}}^{(\mathrm{PN})}$ in (A4). Therefore, by using the continuous mapping theorem (see for example van der Vaart, 2012), we obtain $\hat{\boldsymbol{\beta}}^{(\mathrm{CPN})} \xrightarrow{\text{a.s.}} \boldsymbol{\beta}$. ■

A.3 Proof of Theorem 4.2

Proof: When the components of $\boldsymbol{X} = (X_1, \ldots, X_{p+q})^\top$ and $\boldsymbol{U} = (U_1, \ldots, U_q)^\top$ are independent, we obtain

$$K_{\boldsymbol{X}} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} = \log M_{\boldsymbol{X}} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} = \log \prod_{j=1}^{p+q} M_{X_j}(\beta_j) = \sum_{j=1}^{p+q} K_{X_j}(\beta_j).$$

The same property holds for $K_{\boldsymbol{U}}$. Thus, the derivatives of $K_{\boldsymbol{X}}$ and $K_{\boldsymbol{U}}$ are

$$\frac{\partial}{\partial \boldsymbol{\beta}_1} K_{\boldsymbol{X}} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} = \begin{pmatrix} K'_{X_1}(\beta_1) \\ \vdots \\ K'_{X_p}(\beta_p) \end{pmatrix},$$

$$\frac{\partial}{\partial \boldsymbol{\beta}_2} K_{\boldsymbol{X}} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} = \begin{pmatrix} K'_{X_{p+1}}(\beta_{p+1}) \\ \vdots \\ K'_{X_{p+q}}(\beta_{p+q}) \end{pmatrix},$$

$$\frac{\partial}{\partial \boldsymbol{b}_2} K_{\boldsymbol{U}}(\boldsymbol{b}_2) = \begin{pmatrix} K'_{U_1}(b_{p+1}) \\ \vdots \\ K'_{U_q}(b_{p+q}) \end{pmatrix}.$$

Therefore, we can write the system of Equations (A5) as follows:

$$\beta_0 = b_0 + \sum_{i=1}^{p+q} K_{X_i}(b_i) + \sum_{i=1}^{q} K_{U_i}(b_{p+i}) - \sum_{i=1}^{p+q} K_{X_i}(\beta_i), \tag{A6}$$

$$\boldsymbol{G} = \begin{pmatrix} G_1(\beta_1, b_1) \\ \vdots \\ G_p(\beta_p, b_p) \\ G_{p+1}(\beta_{p+1}, b_{p+1}) \\ \vdots \\ G_{p+q}(\beta_{p+q}, b_{p+q}) \end{pmatrix} \tag{A7}$$

$$= \begin{pmatrix} K'_{X_1}(b_1) - K'_{X_1}(\beta_1) \\ \vdots \\ K'_{X_p}(b_p) - K'_{X_p}(\beta_p) \\ K'_{X_{p+1}}(b_{p+1}) + K'_{U_1}(b_{p+1}) - \mathbf{E}[U_1] - K'_{X_{p+1}}(\beta_{p+1}) \\ \vdots \\ K'_{X_{p+q}}(b_{p+q}) + K'_{U_q}(b_{p+q}) - \mathbf{E}[U_q] - K'_{X_{p+q}}(\beta_{p+q}) \end{pmatrix} = \mathbf{0}_{p+q}. \tag{A8}$$

Following the same argument utilized in the proof of Theorem 3.1, G_j ($j = 1, \dots, p+q$) is always $\mathbf{0}$ in $\mathbb{R}^2$ and is continuously differentiable. We assume the following:

$$\frac{\partial G_j(\beta_j, b_j)}{\partial \beta_j} = -K''_{X_j}(\beta_j) \neq 0, \quad j = 1, \dots, p+q.$$

Then, according to the Implicit Function Theorem, there exists a unique C^1-class function h_j that satisfies $\beta_j = h_j(b_j)$ in the neighbourhood of the zeros of G_j. Furthermore, we obtain $h_j(x) = x$ ($j = 1, \dots, p$) because K_{X_j} increases monotonically in the neighbourhood of β_j. Thus, the CPN estimator $\hat{\boldsymbol{\beta}}^{(\mathrm{CPN})} = (\hat{\beta}_0^{(CPN)}, \dots, \hat{\beta}_{p+q}^{(\mathrm{CPN})})^\top$ is represented as

$$\begin{aligned} \hat{\beta}_0^{(\mathrm{CPN})} &= \hat{\beta}_0^{(\mathrm{PN})} + \sum_{i=p+1}^{p+q} K_{X_i}\left(\hat{\beta}_i^{(\mathrm{PN})}\right) + \sum_{i=1}^{q} K_{U_i}\left(\hat{\beta}_{p+i}^{(\mathrm{PN})}\right) - \sum_{i=p+1}^{p+q} K_{X_i}\left(\hat{\beta}_i^{(\mathrm{CPN})}\right), \\ \hat{\beta}_j^{(\mathrm{CPN})} &= \hat{\beta}_j^{(\mathrm{PN})}, \quad j = 1, \dots, p, \\ \hat{\beta}_j^{(\mathrm{CPN})} &= h_j\left(\hat{\beta}_j^{(\mathrm{PN})}\right), \quad j = p+1, \dots, p+q. \end{aligned}$$

■

Appendix 2. Supplementary Lemmas

Proof: The joint distribution of $(\boldsymbol{X}, Y, \boldsymbol{U})$ is transformed as follows:

$$\begin{aligned} F_{\boldsymbol{X},Y,\boldsymbol{U}}(\boldsymbol{x}, y, \boldsymbol{u}) &= \int_{-\infty}^{x_{p+q}} \cdots \int_{-\infty}^{x_1} \int_{-\infty}^{y} \int_{-\infty}^{u_q} \cdots \int_{-\infty}^{u_1} f_{\boldsymbol{X},Y,\boldsymbol{U}}(\boldsymbol{s}, t, \boldsymbol{v})\, \mathrm{d}\boldsymbol{v}\mathrm{d}t\mathrm{d}\boldsymbol{s} \\ &= \int_{-\infty}^{x_{p+q}} \cdots \int_{-\infty}^{x_1} \int_{-\infty}^{y} \int_{-\infty}^{u_q} \cdots \int_{-\infty}^{u_1} f_{Y,\boldsymbol{U}|\boldsymbol{X}}(t, \boldsymbol{v} \mid \boldsymbol{s})\, \mathrm{d}\boldsymbol{v}\, \mathrm{d}t\, f_{\boldsymbol{X}}(\boldsymbol{s})\, \mathrm{d}\boldsymbol{s} \\ &= \int_{-\infty}^{x_{p+q}} \cdots \int_{-\infty}^{x_1} P(Y \leq y, \boldsymbol{U} \leq \boldsymbol{u} \mid \boldsymbol{X} = \boldsymbol{s}) f_{\boldsymbol{X}}(\boldsymbol{s})\, \mathrm{d}\boldsymbol{s} \\ &= \int_{-\infty}^{x_{p+q}} \cdots \int_{-\infty}^{x_1} P(Y \mid (\boldsymbol{X} = \boldsymbol{s}) \leq y, \boldsymbol{U} \leq \boldsymbol{u}) f_{\boldsymbol{X}}(\boldsymbol{s})\, \mathrm{d}\boldsymbol{s} \\ &= P(\boldsymbol{X} \leq x, (Y \mid \boldsymbol{X}) \leq y, \boldsymbol{U} \leq \boldsymbol{u}) = F_{\boldsymbol{X},Y|\boldsymbol{X},U}(\boldsymbol{x}, y, \boldsymbol{u}). \end{aligned}$$

■

Proof: As shown, the independence of $(\boldsymbol{X}, Y \mid \boldsymbol{X})$ from $\boldsymbol{U}$ or that of $(\boldsymbol{X}, Y)$ from $\boldsymbol{U}$ satisfies $\boldsymbol{X} \perp \boldsymbol{U}$. Thus, from Lemma A.1, we have

$$f_{\boldsymbol{X},Y,\boldsymbol{U}}(\boldsymbol{x}, y, \boldsymbol{u}) = f_{\boldsymbol{X},Y|\boldsymbol{X},\boldsymbol{U}}(\boldsymbol{x}, y, \boldsymbol{u}). \tag{A9}$$

By integrating Equation (A9) with respect to $\boldsymbol{U}$, we obtain

$$f_{\boldsymbol{X},Y}(\boldsymbol{x}, y) = f_{\boldsymbol{X},Y|\boldsymbol{X}}(\boldsymbol{x}, y).$$

Thus, the following necessary conditions are obtained:

$$f_{\boldsymbol{X},Y,\boldsymbol{U}}(\boldsymbol{x}, y, \boldsymbol{u}) = f_{\boldsymbol{X},Y|\boldsymbol{X},\boldsymbol{U}}(\boldsymbol{x}, y, \boldsymbol{u}) = f_{\boldsymbol{X},Y|\boldsymbol{X}}(\boldsymbol{x}, y) f_{\boldsymbol{U}}(\boldsymbol{u}) = f_{\boldsymbol{X},Y}(\boldsymbol{x}, y) f_{\boldsymbol{U}}(\boldsymbol{u}).$$

Similarly, for the sufficient condition, we have

$$f_{\boldsymbol{X},Y|\boldsymbol{X},\boldsymbol{U}}(\boldsymbol{x}, y, \boldsymbol{u}) = f_{\boldsymbol{X},Y,\boldsymbol{U}}(\boldsymbol{x}, y, \boldsymbol{u}) = f_{\boldsymbol{X},Y}(\boldsymbol{x}, y) f_{\boldsymbol{U}}(\boldsymbol{u}) = f_{\boldsymbol{X},Y|\boldsymbol{X}}(\boldsymbol{x}, y) f_{\boldsymbol{U}}(\boldsymbol{u}).$$

■

Lemma A.1: *Under the assumptions in Section 2.1, Y and* $\boldsymbol{W}$ *are independent for a given* $\boldsymbol{X}$.

Proof: The conditional distribution of $(Y, \boldsymbol{W}) \mid \boldsymbol{X}$ is transformed as follows:

$$\begin{aligned} f_{(Y,\boldsymbol{W})|\boldsymbol{X}}(y, \boldsymbol{w} \mid \boldsymbol{x}) &= \frac{f_{Y,\boldsymbol{W},\boldsymbol{X}}(y, \boldsymbol{w}, \boldsymbol{x})}{f_{\boldsymbol{X}}(\boldsymbol{x})} = \frac{f_{Y,\boldsymbol{W},\boldsymbol{X}_o,\boldsymbol{X}_e}(y, \boldsymbol{w}, \boldsymbol{x}_o, \boldsymbol{x}_e)}{f_{\boldsymbol{X}}(\boldsymbol{x})} \\ &= \frac{f_{Y,\boldsymbol{W},\boldsymbol{X}_o,\boldsymbol{U}}(y, \boldsymbol{w}, \boldsymbol{x}_o, \boldsymbol{w}-\boldsymbol{x}_e)}{f_{\boldsymbol{X}}(\boldsymbol{x})} = \frac{f_{Y,\boldsymbol{X}_o,\boldsymbol{X}_e,\boldsymbol{U}}(y, \boldsymbol{x}_e, \boldsymbol{x}_o, \boldsymbol{w}-\boldsymbol{x}_e)}{f_{\boldsymbol{X}}(\boldsymbol{x})} \\ &= \frac{f_{Y,\boldsymbol{X}_o,\boldsymbol{X}_e,\boldsymbol{U}}(y, \boldsymbol{x}_o, \boldsymbol{x}_e, \boldsymbol{w}-\boldsymbol{x}_e)}{f_{\boldsymbol{X}}(\boldsymbol{x})} = \frac{f_{Y,\boldsymbol{X},\boldsymbol{U}}(y, \boldsymbol{x}, \boldsymbol{w}-\boldsymbol{x}_e)}{f_{\boldsymbol{X}}(\boldsymbol{x})} \\ &= \frac{f_{Y,\boldsymbol{X}}(y, \boldsymbol{x}) f_{\boldsymbol{U}}(\boldsymbol{w}-\boldsymbol{x}_e)}{f_{\boldsymbol{X}}(\boldsymbol{x})} = f_{Y|\boldsymbol{X}}(y \mid \boldsymbol{x}) f_{\boldsymbol{W}|\boldsymbol{X}_e}(\boldsymbol{w} \mid \boldsymbol{x}_e). \end{aligned}$$

We use Corollary A.2 in the transformation process. ■

Wada and Kurosawa (2023) used Lemma A.3 for the univariate case without supplying detailed proof. Thus, we provide an explicit proof in this paper.

Appendix 3. Simulation results of small samples

Table A1 shows the estimated biases of the estimators in the case of a multivariate normal distribution with a normal error for $n = 30, 100, 500, 1000$. Overall, the bias correction appears to be effective for small samples. Table A2 lists the estimated biases for Bernoulli and gamma distributions with a gamma error for small samples. Similarly, the bias of the PN estimator was corrected well by the CPN estimator.

Table A1. Estimated bias for a multivariate normal distribution with a normal error.

	30		100		500		1000	
n	PN	CPN	PN	CPN	PN	CPN	PN	CPN
$\widehat{BIAS}(\hat{\beta}_0)$	0.0429	−0.0510	0.0724	−0.0152	0.0829	−0.0033	0.0847	−0.0014
$\widehat{BIAS}(\hat{\beta}_1)$	−0.0668	0.0001	−0.0658	0.0007	−0.0659	0.0006	−0.0662	0.0003
$\widehat{BIAS}(\hat{\beta}_2)$	0.0472	0.0033	0.0448	0.0012	0.0434	−0.0003	0.0431	−0.0005
$\widehat{BIAS}(\hat{\beta}_3)$	−0.1142	0.0020	−0.1154	0.0001	−0.1152	0.0004	−0.1151	0.0006

Table A2. Estimated bias for Bernoulli and gamma distributions with a gamma error.

	30		100		500		1000	
n	PN	CPN	PN	CPN	PN	CPN	PN	CPN
$\widehat{BIAS}(\hat{\beta}_0)$	−0.0726	0.0601	−0.0764	0.0426	−0.0985	0.0211	−0.1060	0.0138
$\widehat{BIAS}(\hat{\beta}_1)$	−0.0043	−0.0043	−0.0006	−0.0006	−0.0002	−0.0002	0.0005	0.0005
$\widehat{BIAS}(\hat{\beta}_2)$	0.0067	0.0067	−0.0011	−0.0011	−0.0004	−0.0004	0.0003	0.0003
$\widehat{BIAS}(\hat{\beta}_3)$	−0.1956	−0.0494	−0.1852	−0.0281	−0.1749	−0.0115	−0.1722	−0.0077