The Econometrics Journal

Econometrics Journal (2024), volume **00**, pp. 1–21.
https://doi.org/10.1093/ectj/utae016

Simple closed-form estimation of a binary latent variable model

YINGYAO HU[†], JINGRONG LI[‡], JI-LIANG SHIU[§] AND MATTHEW SHUM[‖]

[†]*Department of Economics, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218, USA.*
Email: yhu@jhu.edu

[‡]*College of Economics and Management, South China Agricultural University, 483 Wushan Street, Guangzhou, Guangdong 510642, China.*
Email: lijingrong@scau.edu.cn

[§]*Institute for Economic and Social Research, Jinan University, 601 Huangpu Road, Guangzhou, Guangdong 510632, China.*
Email: jishiu.econ@gmail.com

[‖]*Division of Humanities and Social Sciences, California Institute of Technology, 1200 East California Boulevard, MC 228-77 Pasadena, CA 91125, USA.*
Email: mshum@caltech.edu

First version received: 10 July 2023; final version accepted: 21 December 2023.

Summary: This paper develops a closed-form non-parametric estimator of the conditional distribution function for a binary outcome variable given an unobserved latent variable. This type of function is commonly used in models that involve measurement error and dynamic models with agent-specific unobserved heterogeneity. This paper presents a consistent extremum sieve estimator with the following advantages: (i) it has a closed-form expression for all the sieve coefficients, (ii) it is computationally straightforward, equivalent to computing eigenvalues and eigenvectors of matrices without the use of iterative optimization algorithms. While as flexible as the sieve maximum likelihood estimator previously proposed for this model, our estimator proves computationally simpler. The finite sample properties of the estimator are investigated through a Monte Carlo study, and the developed estimator is applied to a probit model to assess the targeting performance of a social welfare program.

Keywords: *Binary outcome variable*, *unobserved latent variable*, *closed-form expression*.

JEL codes: *C01*, *C14*.

1. INTRODUCTION

A sample includes three observed variables (y, x, z) that are associated with the latent scalar variable x^*. The variable y is binary, and the probability mass/density function $f_{xyz}(x, y, z)$ of (y, x, z) satisfies

$$f_{xyz}(x, y, z) = \int f_{x|x^*}(x|x^*) f_{y|x^*}(y|x^*) f_{z|x^*}(z|x^*) f_{x^*}(x^*) dx^*. \tag{1.1}$$

Hu and Schennach (2008)—henceforth HS (2008)—motivated model (1.1) and provided sufficient conditions for the identification of

$$\alpha_0 = (f_{x|x^*}(x|x^*),\ f_{z|x^*}(z|x^*),\ f_{x^*}(x^*)),$$

after normalizing $f_{y|x^*}(y|x^*)$ to be known. HS (2008) also proposed a sieve maximum likelihood estimator (MLE) for α_0. The primary contribution of this paper is to propose a novel closed-form non-parametric estimator for α_0, which exhibits the following advantages: (i) it is as flexible as the semi-parametric sieve MLE proposed in HS (2008), (ii) it provides a closed-form expression for all the sieve coefficients, and (iii) it can be computed using eigenvalues and eigenvectors of matrices without iterative optimization algorithms. The sieve MLE can be considered a benchmark estimator, but its computation can be challenging. Nonetheless, our estimator is as flexible as the sieve MLE, which makes it a valuable tool for empirical applications where computational efficiency is a priority. As a trade-off, our closed-form estimator provides a computationally convenient alternative at the cost of some loss in accuracy.

The observables (y, x, z) may include proxies of the latent variable x^* or dependent variables in some user-specific models with x^* being an explanatory variable. We provide a closed-form non-parametric estimator for the distribution of an observable conditional on the latent variable. Such a conditional distribution may be the model of the dependent variable itself or the conditional distribution associating a proxy with the latent variable.

HS (2008) demonstrated the non-parametric identification of all the distributions associated with the latent variable in measurement error models. Hence, we assume all identification assumptions in HS (2008) in this paper. Based on their identification results, we propose an alternative consistent extreme-value sieve estimator for their non-parametric framework, in addition to the sieve MLE provided in that paper. By utilizing linear sieve parameters, as proposed in HS (2008) for regularization, the problem of continuous data with finite sample sizes can be reduced to analogous matrices of discrete data. Thus, the proposed extremum sieve estimator can be considered as a particular case of the estimator in HS (2008), where continuous data can be discretized, and we apply the matrix manipulation estimator on the discretized data, with the fineness of the discretizing grid increasing with the sample size.

Our estimator is useful for many interesting empirical applications. In a dynamic panel data model, the observables y, x, z are the dependent variables in three different periods and the latent variable x^* may be the random effect; e.g., Shiu and Hu (2013) considered the identification and estimation of non-linear dynamic panel data models in this framework. In auction models, the observables are the bids from different bidders, and the latent variable is the unobserved heterogeneity. See, e.g., Krasnokutskaya (2011) and Hu et al. (2013) for detailed discussion. Such a framework may also be considered a continuous mixture model, where the latent variable may be the continuous type in the mixture model. One may also apply our estimator to the estimation of the production function using panel data, where the observables are outputs in different periods, and the latent variable is the productivity.

The latent variable model in (1.1) is relevant to the non-classical measurement error problem where the latent variable might be correlated with an observed variable. There is a vast literature studying measurement errors in non-linear models, including Hausman et al. (1991), Lewbel (2000), Bound et al. (2001), Newey (2001), Li (2002), Carroll et al. (2004), Schennach (2004; 2007), Mahajan (2006), Lewbel (2007), HS (2008), Chen et al. (2011), Hu and Ridder (2012), and Hu and Sasaki (2015).

We organize the paper as follows. Section 2 describes the proposed eigen-decomposition estimator and its consistency. Section 3 provides an empirical illustration. Section 4 briefly concludes the paper. All technical proofs are included in the Appendix of the main text and Online Appendix. The Monte Carlo study is provided in the Online Appendix.

2. AN EIGEN-DECOMPOSITION ESTIMATOR: A SIMPLE CLOSED-FORM APPROACH

The model considers a dichotomous 0–1 variable,[1] denoted y, and such models have been studied previously by HS (2008), who provided the identification result and proposed a sieve MLE estimator. In this paper, we assume that all identification assumptions hold and show how to construct a consistent non-iterative estimator.

ASSUMPTION 2.1 (NORMALIZATION). *Assume that the latent variable x^* is normalized to satisfy*

$$x^* \equiv \Pr(y = 1|x^*).$$

Assumption 2.1 implies that the conditional probability mass function $f_{y|x^*}(y|x^*) = (x^*)^{\mathbf{1}(y=1)}(1-x^*)^{\mathbf{1}(y=0)}$ is known. The normalization condition that $\Pr(y = 1|x^*)$ is monotonic in x^* can be found in several empirical applications. For example, let us consider estimating the impact of family earnings on college attendance, where people with higher family earnings are more likely to attend college. However, family earnings are susceptible to measurement error problems due to confidential information or privacy issues. The instrumental variable z can be the predicted earnings in the regression of reported income on demographic variables, such as parents' education, parents' occupations, race, age, and region. It is reasonable to assume that variables x and z do not affect y once the true family earnings x^* are known.

Another example is the estimation of the effect of health on labour force participation. The monotonicity condition holds in this case because healthier people are more likely to join the workforce. When a self-reported health status represents the health status, the variable is not an objective indicator of health, which leads to a measurement error problem. We may use smoking as an instrumental variable for the health measure, in which case x and z do not provide any more information about the dependent variable y than x already provides.

Based on the identification results in HS (2008), our model implies the following two equations that link probability mass/density functions observed in (or estimable from) the data to latent probability density functions, which are the objects of interest:

$$f_{xyz}(x, y = 1, z) = \int f_{x|x^*}(x|x^*)x^* f_{z|x^*}(z|x^*)f_{x^*}(x^*)dx^*, \tag{2.1}$$

$$f_{xz}(x, z) = \int f_{x|x^*}(x|x^*)f_{z|x^*}(z|x^*)f_{x^*}(x^*)dx^*. \tag{2.2}$$

In the preceding equations, we can observe or directly estimate the probability mass/density functions of the left-hand sides from the data. The goal is to recover estimates of the unknown densities on the right-hand sides of these equations, which we denote by $\alpha_0 = (f_{x|x^*}, f_{z|x^*}, f_{x^*})^T$. We propose a closed-form, non-iterative procedure for estimating the unknown densities in α_0.

When the parameter function spaces for densities are large, such as L^1 or L^2, an estimator based on the spaces could yield an inconsistent estimator. In this paper, our approximation analysis follows smoothness restrictions in Ai and Chen (2003) and Hu and Schennach (2008) and introduces weighted Hölder spaces. Given a $d \times 1$ vector of non-negative integers, $\kappa = (\kappa_1, \ldots, \kappa_d)'$, let $[\kappa] = \kappa_1 + \cdots + \kappa_d$. Let D^κ denote the differential operator defined by $D^\kappa = \partial^{[\kappa]}/(\partial\xi_1^{\kappa_1} \cdots \partial\xi_d^{\kappa_d})$. Set $\gamma = m + p$, where m denotes the largest integer satisfying $\gamma > m$. The

[1] Under certain additional assumptions, it is possible to transform the binary outcome y into a discrete variable with more than two values. For further details, we refer the reader to the Online Appendix.

Hölder space $\Lambda^{\gamma}(\nu)$ of order $\gamma > 0$ is a collection of functions that are m times continuously differentiable on ν and the mth derivatives are Hölder continuous with exponent p. The Hölder space becomes a Banach space with the Hölder norm, i.e., for all $g \in \Lambda^{\gamma}(\nu)$,

$$\|g\|_{\Lambda^{\gamma}} = \sup_{\xi \in \nu} |g(\xi)| + \max_{\kappa_1 + \cdots + \kappa_d = m} \sup_{\xi \neq \xi' \in \nu} \frac{|D^{\kappa} g(\xi) - D^{\kappa} g(\xi')|}{\|\xi - \xi'\|_E^p}. \tag{2.3}$$

The weighted Hölder norm is defined as $\|g\|_{\Lambda^{\gamma,\omega}} \equiv \|\widetilde{g}\|_{\Lambda^{\gamma}}$ for $\widetilde{g}(\xi) \equiv g(\xi)\omega(\xi)$ with $\omega(\xi) = (1 + \|\xi\|_E^2)^{-\varsigma/2}$, $\varsigma > \gamma > 0$. The corresponding weighted Hölder space is $\Lambda^{\gamma,\omega}(\nu)$ and a weighted Hölder ball is defined as $\Lambda_c^{\gamma,\omega}(\nu) \equiv \{g \in \Lambda^{\gamma,\omega}(\nu) : \|g\|_{\Lambda^{\gamma,\omega}} \leq c < \infty\}$.

Let $\mathcal{Y}$, $\mathcal{X}$, $\mathcal{X}^*$, and $\mathcal{Z}$ denote the supports of the distributions of the random variables y, x, x^*, and z, respectively. In particular, $\mathcal{X}^* = [0, 1]$.[2] Set some positive constants γ_i for $i = 1, 2, 3$. The three unknown densities $f_{x|x^*}$, $f_{z|x^*}$, f_{x^*} are assumed to be in the spaces

$$\mathcal{F}_1 = \left\{ f_1(\cdot|\cdot) \in \Lambda_c^{\gamma_1,\omega}(\mathcal{X} \times \mathcal{X}^*) : f_1(\cdot) \geq 0 \text{ and } \int f_1(x|x^*)dx = 1 \text{ for all } x^* \in \mathcal{X}^* \right\},$$

$$\mathcal{F}_2 = \left\{ f_2(\cdot|\cdot) \in \Lambda_c^{\gamma_2,\omega}(\mathcal{Z} \times \mathcal{X}^*) : f_2(\cdot) \geq 0 \text{ and } \int f_2(z|x^*)dz = 1 \text{ for all } x^* \in \mathcal{X}^* \right\},$$

$$\mathcal{F}_3 = \left\{ f_3(\cdot) \in \Lambda_c^{\gamma_3,\omega}(\mathcal{X}^*) : f_3(\cdot) \geq 0 \text{ and } \int f_3(x^*)dx^* = 1 \right\}.$$

Let $\alpha = (f_1, f_2, f_3)^T \in \mathcal{A} = \mathcal{F}_1 \times \mathcal{F}_2 \times \mathcal{F}_3$. We introduce a strong norm $\|\cdot\|_s$ for α that will be used to show the consistency of the sieve estimator. For $\alpha \equiv (f_1, f_2, f_3)^T$,

$$\|\alpha\|_s = \sum_{i=1}^{3} \|f_i\|_{\Lambda^{\gamma_i,\omega}}.$$

First, we use a truncated serial expansion to approximate all the functions in (2.1) and (2.2). We specify the series used in this paper: an orthonormal Fourier series $\psi_0(w) = 1$ and $\psi_k(w) = \sqrt{2}\cos(k\pi w)$ for $k \geq 1$ and a polynomial series $\varphi_i(w) = w^i$ for $i \geq 0$. Let K be associated with the number of terms in the truncated series, and let $\mathcal{A}^K \equiv \mathcal{F}_1^K \times \mathcal{F}_2^K \times \mathcal{F}_3^K$ be a sequence of sieve spaces to approximate the function space $\mathcal{A}$ with

$$\mathcal{F}_1^K = \left\{ f^K(x|x^*) = \sum_{k=0}^{K} \sum_{j=0}^{K} e_{kj} \psi_k(x) \varphi_j(x^*) \in \mathcal{F}_1 \right\},$$

$$\mathcal{F}_2^K = \left\{ f^K(z|x^*) = \sum_{k=0}^{K} \sum_{j=0}^{K} h_{kj} \psi_k(z) \psi_j(x^*) \in \mathcal{F}_2 \right\},$$

$$\mathcal{F}_3^K = \left\{ f^K(x^*) = \sum_{k=0}^{K+1} q_k \psi_k(x^*) \in \mathcal{F}_3 \right\}.$$

While the orders of the sieve for $f(x|x^*)$ and $f(z|x^*)$ are the same, the basis functions for each of them are different. We use a tensor product of linear univariate sieves to approximate $f(x|x^*)$

[2] Under Assumption 2.1, the normalization condition $\Pr(y = 1|x^*) = x^*$ implies that the values of x^* must be between 0 and 1.

and $f(z|x^*)$. The univariate sieve for x^* in $f(x|x^*)$ is a power series, while the univariate sieve for x^* in $f(z|x^*)$ is a Fourier series.

Next, assume that the probability mass/density functions of observables $f_{xyz}(x, y = 1, z)$ and $f_{xz}(x, z)$ are in the space

$$\mathcal{H} = \left\{ f_4(\cdot,\cdot) \in \Lambda_c^{\gamma_4,\omega}(\mathcal{X} \times \mathcal{Z}) : f_4(\cdot) \geq 0 \text{ and } \int\int f_4(x,z)dxdz = 1 \right\},$$

where γ_4 is a positive constant. We approximate $f_{xyz}(x, y = 1, z)$ and $f_{xz}(x, z)$ by the truncated serial expansions

$$f_{xyz}(x, y = 1, z) \approx f^K(x, y = 1, z) = \sum_{k=0}^{K} \sum_{j=0}^{K} a_{kj} \psi_k(x) \psi_j(z), \tag{2.4}$$

$$f_{xz}(x, z) \approx f^K(x, z) \equiv \sum_{k=0}^{K} \sum_{j=0}^{K} b_{kj} \psi_k(x) \psi_j(z), \tag{2.5}$$

where $\psi_k(\cdot)$ is a known orthonormal basis function, and a_{kj} and b_{kj} are serial coefficients. The product of the univariate orthonormal bases $\psi_k(\cdot)\psi_j(\cdot)$ is an orthonormal basis function for multivariate functions. Thus, we can find the serial coefficients relative to this orthonormal basis by taking the inner product of the functions with each basis function.

Let $\mathcal{H}^K$ be a sequence of sieve spaces to approximate the function space $\mathcal{H}$ with

$$\mathcal{H}^K = \left\{ f^K(x, z) = \sum_{k=0}^{K} \sum_{j=0}^{K} c_{kj} \psi_k(x) \psi_j(z) \in \mathcal{H} \right\}.$$

For simplicity, we denote the norms of both spaces $\mathcal{A}$ and $\mathcal{H}$ as $\|\cdot\|_s$.

ASSUMPTION 2.2. *Suppose that*

(*i*) *$\alpha_0 \in \mathcal{A}$, and $\mathcal{A}$ is compact under $\|\cdot\|_s$;*
(*ii*) *for any $\alpha \in \mathcal{A}$, there exists $\Pi_K\alpha \in \mathcal{A}^K$ such that $\|\Pi_K\alpha - \alpha\|_s = o(1)$;*
(*iii*) *there exist $f^K(x, y = 1, z)$, $f^K(x, z) \in \mathcal{H}^K$ such that $\|f^K(\cdot, y = 1, \cdot) - f_{xyz}(\cdot, y = 1, \cdot)\|_s = o(1)$ and $\|f^K(\cdot,\cdot) - f_{xz}(\cdot,\cdot)\|_s = o(1)$;*
(*iv*) *the convergence of the sieve approximations in (ii) and (iii) depends on K, and $K \to \infty$ and $K/n \to 0$ as the sample size $n \to \infty$.*

Conditions (ii) and (iii) of Assumption 2.2 are satisfied for the parameter spaces $\mathcal{A}$ and $\mathcal{H}$ containing bounded and smooth functions and are commonly imposed conditions in series approximations.[3] Assumption 2.2(iv) restricts the number of terms used in the series approximations that is allowed to grow with the sample size at a controlled rate. Denote the corresponding coefficient matrices as $A = [a_{k-1,j-1}]_{k,j\in\{1,2,\ldots,K+1\}}$ and $B = [b_{k-1,j-1}]_{k,j\in\{1,2,\ldots,K+1\}}$.[4] Then, the matrix representations of $f^K(x, y = 1, z)$ and $f^K(x, z)$ in (2.4) and (2.5) can be expressed as $\overrightarrow{\psi}(x)^T A \overrightarrow{\psi}(z)$ and $\overrightarrow{\psi}(x)^T B \overrightarrow{\psi}(z)$, respectively. We assume that $f^K(x, z)$ converges to the true function $f_{xz}(x, z)$ as K goes to infinity concerning its corresponding Hölder norm. Since the joint distribution of (x, y, z) is observed in the sample, the sieve coefficients a_{kj} and b_{kj} may also be estimated directly from the sample. Similarly, we may define $f^K(x, z)$, $f^K(x|x^*)$, $f^K(z|x^*)$, and

[3] See Gallant and Nychka (1987) for detailed discussions.
[4] The definition of the sieve coefficients a_{kj} and b_{kj} can be found in (A.2) and (A.3) in the Appendix.

$f^{K+1}(x^*)$[5] as the truncated serial expansion of the functions $f_{xz}(x,z)$, $f_{x|x^*}(x|x^*)$, $f_{z|x^*}(z|x^*)$, and $f_{x^*}(x^*)$.

By replacing the functions in (2.1) and (2.2) with their serial approximations, we obtain

$$f^K(x, y=1, z) = \int f^K(x|x^*)x^* f^K(z|x^*) f^{K+1}(x^*)dx^*, \tag{2.6}$$

$$f^K(x, z) = \int f^K(x|x^*) f^K(z|x^*) f^{K+1}(x^*)dx^*. \tag{2.7}$$

Equations (2.6) and (2.7) hold at all values of x and z, and x and z are continuous variables. This implies an infinite number of restrictions on a finite number of unknown serial coefficients. Our key contribution is a proposed closed-form estimator for all the unknown densities on the right-hand sides (RHSs). Specifically, we provide a closed-form estimator for the unknown serial coefficients in the approximations of these unknown functions, which means expressing the serial coefficients in terms of statistics of the data sample without using any optimization algorithms.

For an estimator in the sieve space $\mathcal{A}^K$, each component of the estimator can be written in a matrix expression. Denote by $E = [e_{k-1,j-1}]_{k,j\in\{1,2,\ldots,K+1\}}$, $H = [h_{k-1,j-1}]_{k,j\in\{1,2,\ldots,K+1\}}$, and $\overrightarrow{q} = [q_1, q_2, \ldots, q_K, q_{K+1}]^T$ the coefficient matrices. It follows that

$$f^K(x|x^*) = \overrightarrow{\psi}_K(x)^T E \overrightarrow{\varphi}_K(x^*), \tag{2.8}$$

$$f^K(z|x^*) = \overrightarrow{\psi}_K(z)^T H \overrightarrow{\psi}_K(x^*), \tag{2.9}$$

$$f^K(x^*) = [1, \overrightarrow{q}^T] \overrightarrow{\psi}_{K+1}(x^*), \tag{2.10}$$

where $\overrightarrow{\psi}_K(x) = [\psi_0(x), \psi_1(x), \ldots, \psi_K(x)]^T$ and $\overrightarrow{\varphi}_K(x) = [\varphi_0(x), \varphi_1(x), \ldots, \varphi_K(x)]^T$. Our derivation for the estimator relies on the trick of using power series to absorb $x^* \equiv \Pr(y=1|x^*)$ into x^*. Thus, we use φ instead of ψ to specify the sieve approximation of $f^K(x|x^*)$ in x^*.

According to HS (2008), parameter $(f_{y|x^*}, f_{x|x^*}, f_{z|x^*}, f_{x^*})$ is identified and a sieve MLE is defined as

$$\widehat{\alpha}_{sv} = \underset{\alpha=(f_1,f_2,f_3,f_4)^T\in\mathcal{A}^n}{\arg\max} \frac{1}{n}\sum_{i=1}^{n} \ln f_0(x_i, y_i, z_i; \alpha),$$

where

$$f_0(x, y, z) = \int f_1(x|x^*) f_2(y|x^*) f_3(z|x^*) f_4(x^*)dx^*,$$

and $f_0(x_i, y_i, z_i; \alpha)$ is an empirical likelihood function with the densities in $(f_{y|x^*}, f_{x|x^*}, f_{z|x^*}, f_{x^*})$ replaced by their sieve approximations (f_1, f_2, f_3, f_4).

Ai and Chen (2003) and HS (2008) provided the consistency of the sieve MLE. Compared to the sieve MLE method, the most significant limitation of our approach is that we cannot handle the case where x^* is multivariate. However, it is essential to recognize that, when considering x^* as a multivariate variable, both z and x in (2.1) and (2.2) must also be multivariate, which imposes higher data requirements. Moreover, it is not that straightforward to extend the proposed method to a multivariate case of x^* because the proposed method combines the univariate power series sieve for x^* of $f(x|x^*)$ with the normalization condition, $\Pr(y=1|x^*) = x^*$. In a multivariate case of x^*, there are at least two sets of sieve coefficients and there may not exist a relation equation for

[5] In the Appendix, we start with a sieve of order M for $f_{x^*}(x^*)$. Later we take $M = K+1$ and require Assumption A.2 to solve for its corresponding sieve coefficients.

the closed-form solution. Therefore, this paper specifically focuses on the scenario where x^* is scalar, and the exploration of extending the method to handle multivariate x^* constitutes a future research direction.

Assumption 2.1 is an additional assumption compared to the sieve MLE method. This assumption is crucial because the trick in our paper relies on using the power series $\varphi(\cdot)$ for the basis functions of the x^* part of $f^K_{x|x^*}(x|x^*)$ and the normalization $x^* \equiv P(y = 1|x^*)$. These are the two key assumptions that enable us to obtain a closed-form estimation.

While our method imposes additional assumptions compared to the sieve MLE method, it is noteworthy that the estimation of sieve coefficients within the framework of sieve MLE presents a formidable challenge. Sieve MLE computational complexity is equivalent to parametric MLE, characterized by numerous parameters. In particular, most optimization algorithms require initial values of these sieve coefficients, which may be arbitrary. In contrast, here, we provide a simple closed-form estimator for these sieve coefficients that do not involve any optimization algorithm.

Substituting the sieve expressions (2.8), (2.9), and (2.10) into the sieve approximation of the identification equations (2.6) and (2.7), we can obtain the matrix representation of the coefficient matrices as[6]

$$\begin{aligned} A &= EC_q H^T, \\ B &= EC_{q0} H^T, \end{aligned} \tag{2.11}$$

where C_q and C_{q0} are functions of $\overrightarrow{q}$ and are defined as

$$C_q = \left[\int_0^1 \varphi_{k-1}(x^*) \cdot x^* \cdot \left[\sum_{l=0}^{K+1} q_l \psi_l(x^*) \right] \cdot \psi_{j-1}(x^*) dx^* \right]_{k,j \in \{1,2,\ldots,K+1\}},$$

$$C_{q0} = \left[\int_0^1 \varphi_{k-1}(x^*) \cdot \left[\sum_{l=0}^{K+1} q_l \psi_l(x^*) \right] \cdot \psi_{j-1}(x^*) dx^* \right]_{k,j \in \{1,2,\ldots,K+1\}}.$$

Thus, we have transformed the sieve population problem into a finite-dimensional sieve problem involving matrices. The estimated coefficient matrices $\widehat{A}$ and $\widehat{B}$ for A and B in (2.12) and (2.13) below are defined as follows: $\widehat{A} = [\widehat{a}_{k-1,j-1}]_{k,j \in \{1,2,\ldots,K+1\}}$ and $\widehat{B} = [\widehat{b}_{k-1,j-1}]_{k,j \in \{1,2,\ldots,K+1\}}$ with the estimated coefficients $\widehat{a}_{kj}$ and $\widehat{b}_{kj}$ satisfying

$$\widehat{a}_{kj} = \frac{1}{N} \sum_{i=1}^{N} [\mathbf{1}(y_i = 1)\psi_k(x_i)\psi_j(z_i)], \tag{2.12}$$

$$\widehat{b}_{kj} = \frac{1}{N} \sum_{i=1}^{N} [\psi_k(x_i)\psi_j(z_i)]. \tag{2.13}$$

Therefore, matrices A and B are available. For notational simplicity, we suppress $\widehat{A}$ and $\widehat{B}$ and still use A and B. The proposed eigen-decomposition estimator (EDE) $\widehat{\alpha}_{ede}$ of α_0 is to solve the coefficient matrices E, H, and $\overrightarrow{q}$ without any optimization algorithm.

Next, we present an informal step-by-step construction for the existence of the coefficient matrices E, H, and $\overrightarrow{q}$ from the estimated coefficients $\widehat{a}_{kj}$ and $\widehat{b}_{kj}$. A detailed and rigorous derivation with assumptions for the coefficient matrices is provided in the Appendix.

[6] The derivation is provided in the Appendix.

Recall that $\varphi_k(x^*) = (x^*)^{k-1}$; then the kth row of C_q is the same as the $(k+1)$th row of C_{q0} for $k = 1, \ldots, K$.[7] That is, we can write

$$C_q = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & 0 & 1 \\ -d_0 & -d_1 & -d_2 & \ldots & -d_K \end{bmatrix} C_{q0} = G_q C_{q0}. \tag{2.14}$$

The three matrices C_q, C_{q0}, and G_q are all unknown, and matrix G_q connects C_q and C_{q0}. Matrix G_q is unknown only up to $(d_0, d_1, \ldots, d_K)$, where the d's are implicitly defined in (2.14). We impose assumptions to solve for these unknown parameters.

Second, we show that $(d_0, d_1, \ldots, d_K)$ are functions of the eigenvalues of AB^{-1} and then that $\overrightarrow{q}$ can be solved from $(d_0, d_1, \ldots, d_K)$. If matrix B is invertible then we can obtain

$$AB^{-1} = E(C_q C_{q0}^{-1})E^{-1} = EG_q E^{-1}. \tag{2.15}$$

Equation (2.15) means that the observed matrix AB^{-1} is similar to matrix G_q, implying that they share the same eigenvalues and have the same characteristic polynomial.

Equation (2.15) implies that

$$\begin{aligned} \det(\lambda I - AB^{-1}) &= \det(\lambda I - G_q) \\ &= \lambda^{K+1} + d_K \lambda^K + \cdots + d_1 \lambda + d_0 \\ &\equiv (\lambda - \lambda_0)(\lambda - \lambda_1) \cdots (\lambda - \lambda_K). \end{aligned}$$

This means that the d_k are the coefficients of the characteristic polynomial $\det(\lambda I - AB^{-1})$ of the directly estimated AB^{-1}, so that the coefficients $(d_0, d_1, \ldots, d_K)$ can be recovered from the eigenvalues of AB^{-1}, $(\lambda_0, \lambda_1, \ldots, \lambda_K)$.[8]

At this point, we have recovered matrix G_q and regarded it as a known matrix so that we can use the RHS of (2.14) to solve for the serial coefficients q_l in $f^K(x^*)$.[9] After recovering $\overrightarrow{q}$ in this way, we have now estimated $f^K(x^*) = [1, \overrightarrow{q}^T]\overrightarrow{\psi}_{K+1}(x^*)$.

Fourth, we turn to the unknown coefficient matrix E in $f^K(x|x^*)$. Let an eigen-decomposition of AB^{-1} be $AB^{-1} = VDV^{-1}$ with a diagonal matrix $D = \mathrm{diag}(\lambda_0, \lambda_1, \ldots, \lambda_K)$ and a corresponding eigenvector matrix V. Because AB^{-1} and G_q have the same set of eigenvalues, we can use the diagonal matrix D to obtain an eigen-decomposition of G_q as $G_q = S_q D S_q^{-1}$ with an eigenvector matrix S_q.[10]

We may then estimate E from

$$\underbrace{VDV^{-1}}_{\text{an eigen-dcomposition of } AB^{-1}} = AB^{-1} = EG_q E^{-1} = E \underbrace{S_q D S_q^{-1}}_{\text{an eigen-dcomposition of } G_q} E^{-1}.$$

[7] The trick of using power series to absorb x^* into a higher order of series φ works because $\varphi_{k-1}(x^*) \cdot x^* = \varphi_k(x^*)$ for all k. However, this trick likely does not extend to series other than the power series $\varphi_i(x^*) = x^{*i}$ for $i \geq 0$. Most other series involve scaling or location parameters when constructing higher-order terms. These modified series do not satisfy $\varphi_{k-1}(x^*) \cdot x^* = \varphi_k(x^*)$.

[8] The relations between the λ_k and d_k are given in (A.10)–(A.12) of the Appendix.

[9] See the proof of Proposition 2.1 in the Appendix for details that start at (A.13).

[10] The existence of the eigen-decomposition of AB^{-1} and G_q is based on Theorem 1.3.9 of Horn and Johnson (1985): ‘If an $n \times n$ matrix A has n distinct eigenvalues then the matrix A is diagonalizable, i.e. there exists an invertible matrix P such that $P^{-1}AP$ is a diagonal matrix.’ Because AB^{-1} and G_q have the same set of eigenvalues, Assumption A.1(ii) ensures that both AB^{-1} and G_q have distinct eigenvalues and are diagonalizable.

Given the similarity of the known matrices AB^{-1} and G_q, the two eigenvector matrices V and ES_q are equal to each other up to the scale of the eigenvectors.[11] From this, we can recover the E matrix up to scale.[12]

Finally, we estimate H in $f^K(z|x^*)$. Using (2.11), we have $H = (C_{q0}^{-1}E^{-1}B)^T$. At this point, all the matrices on the right-hand side are known; in particular, C_{q0} is known given knowledge of the serial coefficients $\overrightarrow{q}$. Then $f^K(z|x^*) = \overrightarrow{\psi}_K(z)^T H \overrightarrow{\psi}_K(x^*)$.

In short, these five steps constitute our closed-form semi-parametric estimator for α as follows.

PROPOSITION 2.1. *Suppose that Assumptions 2.1 and 2.2, and Assumptions A.1, A.2, A.3, and A.4 in the Appendix hold. Then the EDE $\widehat{\alpha}_{ede} = (\widehat{f}^K_{x|x^*}, \widehat{f}^K_{z|x^*}, \widehat{f}^K_{x^*})^T \in \mathcal{A}^K$ exists and is unique, i.e., there exists unique matrices $\widehat{E}$, $\widehat{H}$, and $\widehat{\overrightarrow{q}}$ that solve the equation*

$$\widehat{f}^K_{x,y,z}(x, y, z) = \int_0^1 \widehat{f}^K_{x|x^*}(x|x^*) \cdot f_{y|x^*}(y|x^*) \cdot \widehat{f}^K_{z|x^*}(z|x^*) \cdot \widehat{f}^K_{x^*}(x^*)dx^*$$

for any (x, y, z), where

$$f_{y|x^*}(y|x^*) = (x^*)^{\mathbf{1}(y=1)}(1 - x^*)^{\mathbf{1}(y=0)},$$

$$\widehat{f}^K_{x,y,z}(x, y, z) = \overrightarrow{\psi}_K(x)^T \left[A^{\mathbf{1}(y=1)}(B - A)^{\mathbf{1}(y=0)}\right] \overrightarrow{\psi}_K(z).$$

The closed-form representation for $\widehat{\alpha}_{ede}$ is

$$\widehat{f}^K_{x|x^*}(x|x^*) = \overrightarrow{\psi}_K(x)^T \widehat{E} \overrightarrow{\varphi}_K(x^*),$$

$$\widehat{f}^K_{z|x^*}(z|x^*) = \overrightarrow{\psi}_K(z)^T \widehat{H} \overrightarrow{\psi}_K(x^*),$$

$$\widehat{f}^K_{x^*}(x^*) = \left[1, \widehat{\overrightarrow{q}}^T\right] \overrightarrow{\psi}_{K+1}(x^*).$$

Proof. See the Appendix.

We provide a broad description of an algorithm that summarizes the five steps in the computation of the estimator and the order of the construction is $\overrightarrow{q}$ and E first, and then use $\overrightarrow{q}$ and E to construct H.

Algorithm for the EDE

(1) Given data $\{x_i, y_i, z_i\}$ of sample size n, estimate coefficient matrices A and B using (2.12) and (2.13) and calculate the square matrix AB^{-1}.
(2) Compute the characteristic polynomial of AB^{-1} and obtain the coefficients of the polynomial, $(d_0, d_1, \ldots, d_K)$.
(3) Use the coefficients $(d_0, d_1, \ldots, d_K)$ to construct

$$[c_{lj}]_{l,j\in\{1,2,\ldots,K+1\}} = \left[\int \left((x^*)^{K+1} + \sum_{k=0}^{K}(x^*)^k d_k\right) \cdot \psi_l(x^*) \cdot \psi_{j-1}(x^*)dx^*\right],$$

$$\overrightarrow{c}_0 = \left[\int \left((x^*)^{K+1} + \sum_{k=0}^{K}(x^*)^k d_k\right) \cdot \psi_0(x^*) \cdot \psi_{j-1}(x^*)dx^*\right].$$

[11] The relation between V and ES_q can be written as (A.14) in the Appendix.

[12] In the Appendix, we show how the scaling factors can be recovered from the condition that $\int f(x|x^*)dx = 1$ for any x^*.

Then, $\overrightarrow{q}^T = -\overrightarrow{c}_0^T([c_{lj}]_{l,j\in\{1,2,\dots,K+1\}})^{-1}$ and $f^K(x^*) = [1, \overrightarrow{q}^T]\overrightarrow{\psi}_{K+1}(x^*)$.

(4) Use the coefficients $(d_0, d_1, \dots, d_K)$ to construct

$$G_q \equiv \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 1 \\ -d_0 & -d_1 & -d_2 & \dots & -d_K \end{bmatrix}.$$

Compute eigenvalues and eigenvectors of AB^{-1} and G_q, and let $AB^{-1} = VDV^{-1}$ and $G_q = S_q D S_q^{-1}$ with a diagonal matrix $D = \text{diag}(\lambda_0, \lambda_1, \dots, \lambda_K)$. Let $[P]_{ij}$ stand for the (i, j)th element of matrix P. Calculate a diagonal matrix $\tilde{\tilde{D}}$ with $[\tilde{\tilde{D}}]_{ii} = [S_q]_{1,i}/[V]_{1,i}$. Then, $E = V\tilde{\tilde{D}}S_q^{-1}$ and $f^K(x|x^*) = \overrightarrow{\psi}_K(x)^T E \overrightarrow{\varphi}_K(x^*)$.

(5) Use $\overrightarrow{q}$ to compute

$$C_{q0} = \left[\int_0^1 x^{*k-1} \cdot \left[\sum_{l=0}^{M} q_l \psi_l(x^*)\right] \cdot \psi_{j-1}(x^*) dx^*\right]_{k,j\in\{1,2,\dots,K+1\}}.$$

Then, $H = (C_{q0}^{-1} E^{-1} B)^T$ and $f^K(z|x^*) = \overrightarrow{\psi}_K(z)^T H \overrightarrow{\psi}_K(x^*)$.

Here we summarize our consistent results for the EDE $\widehat{\alpha}_{ede}$.

PROPOSITION 2.2. *Suppose that Assumptions 2.1 and 2.2, and Assumptions A.1, A.2, A.3, and A.4 in the Appendix hold. Let $\widehat{\alpha}_{ede}$ be the sieve estimator for α_0 in Proposition 2.1, and suppose that Online Appendix Assumptions S2.1–S2.3 hold; then we have $\|\widehat{\alpha}_{ede} - \alpha_0\|_s = o_p(1)$.*

The Online Appendix presents a proof of the consistent result. The conditions required for the consistent result are general and are essentially the same as those imposed in Ai and Chen (2003), Newey and Powell (2003), and HS (2008). These conditions can be categorized into the following classes: (i) existence conditions for the EDE $\widehat{\alpha}_{ede}$, (ii) the existence of approximating subspaces $\mathcal{A}^K$ and $\mathcal{H}^K$ for function spaces $\mathcal{A}$ and $\mathcal{H}$, respectively, (iii) envelope conditions that secure a Hölder continuity property of the objective function, (iv) the compactness assumption in Assumption 2.2(i) limits the size of the space of functions, and it is satisfied when the infinite-dimensional parameter space $\mathcal{A}$ consists of bounded and smooth functions. We restrict the true structural function to the compact set, and this assumption essentially eliminates the ill-posed inverse problem.

3. A MONTE CARLO STUDY

In this section, we evaluate the performance of the proposed EDE estimator. The simulation design for the discrete choice model is similar to the experiments considered by Hu and Schennach (2008). We have included the data-generation process and specific results of the simulation in the Online Appendix. For comparison, we also computed the sieve MLE proposed by Hu and Schennach (2008).

The estimation results for the sieve MLE in Hu and Schennach (2008) and the EDEs of degree 3 are presented in Online Appendix Figures S1–S4. Both the sieve MLE and the EDEs of degree 3 perform well, except for the boundary points. In Online Appendix Figures S1–S4, the values of the EDEs exceed their actual values near the boundary points. The figures indicate that the EDEs exhibit overshoot and cause boundary bias. Thus, the EDEs perform poorly near the boundary points due to their boundary bias, which may restrict the EDEs to an interior subset of the support of the variable of interest. Nevertheless, when excluding points near the boundaries, the shapes of the EDEs closely resemble the true functions.

We define the following integrated mean squared errors (IMSEs) to comprehensively evaluate our estimator $\widehat{\alpha}_{ede}$, with each IMSE representing the performance of individual components of $\widehat{\alpha}_{ede}$:

$$IMSE(x^*) = \begin{pmatrix} E[\int [\widehat{f}^K(x|x^*) - f_{x|x^*}(x|x^*)]^2 dx] \\ E[\int [\widehat{f}^K(z|x^*) - f_{z|x^*}(z|x^*)]^2 dx] \\ E[\int [\widehat{f}^K(x^*) - f_{x^*}(x^*)]^2 dx] \end{pmatrix}.$$

The IMSEs and the sum of IMSEs for the sieve MLE and the EDEs of degrees 3 are reported in Online Appendix Tables S1–S2. The sieve MLE outperforms the EDEs of degrees 3 in terms of the IMSE. However, while the sieve MLE serves as a benchmark, it is difficult to compute due to a large number of nuisance parameters, making it harder to find a consistent estimator as an initial value for numerical optimization. As a trade-off, the EDEs offer a computationally convenient option, but sacrifice some accuracy. Therefore, the EDEs can be used as a first-step estimator for a more efficient estimator, complementing the sieve MLE. The Monte Carlo exercises presented here illustrate the extent of this trade-off.

4. EMPIRICAL ILLUSTRATION

China's Targeted Poverty Alleviation (TPA) is the most extensive social safety-net program in the world. It is essential to evaluate it so that policymakers know the extent to which the program meets its intended objective of reducing poverty. An essential criterion for selecting TPA beneficiaries is that the average disposable income of households is lower than the national poverty line (about \$442 in 2015). As a result, TPA candidates are incentivized to report lower income than their actual income to increase their chances of being selected as beneficiaries. In this case, there exists a likelihood that the actual beneficiaries may not belong to the eligible impoverished segments, thereby leading to the possibility of TPA displaying misdirected performance. Therefore, this section applies the developed estimator to a probit model to evaluate the targeting performance of the TPA program based on households' income:

$$y_i = 1(\beta_0 + \beta_1 x_i^* + \beta_2 w_i + u_i > 0). \tag{4.1}$$

Here y_i is a binary variable that equals 1 if the ith household is selected as a TPA beneficiary and 0 if not, x_i^* is the ith household's actual income, w_i is a covariate that equals 1 if there are village cadres among the ith household members and 0 if not, and u_i is normally distributed. Then the conditional density function is

$$f(y_i|x_i^*, w_i; \beta) = \Phi(\beta_0 + \beta_1 x_i^* + \beta_2 w_i)^{y_i}[1 - \Phi(\beta_0 + \beta_1 x_i^* + \beta_2 w_i)]^{1-y_i}.$$

However, households' actual income x_i^* is latent and may be measured with error, precluding direct estimation of the probit model. To address this measurement error problem and apply our estimator, we utilize households' reported income x_i and an instrumental variable (IV) z. The IV z used here is the predicted income in the regression of reported income on demographic variables, i.e., household head's age, gender, education, household labour, family members over sixty, and children under ten.

We use a two-step estimation to obtain the estimates of $(\beta_0, \beta_1, \beta_2)$ in (4.1). In the first step, we impose the normalization assumption $x^* \equiv \Pr(y = 1|x^*)$ and use the EDE method to obtain $\widehat{f}_{x|x^*}(x|x^*)$ and $\widehat{f}_{x^*}(x^*)$. In the second step, to estimate the parameters in the outcome equation, consider the expression

$$\begin{aligned} f(y, x, w) &= \int f(y|x^*, x, w) f(x|x^*, w) f(w|x^*) f(x^*) dx^* \\ &= \int f(y|x^*, w) f(x|x^*) f(w|x^*) f(x^*) dx^*, \end{aligned}$$

where we have used $f(y|x^*, x, w) = f(y|x^*, w)$ and $f(x|x^*, w) = f(x|x^*, w)$. While the dichotomous model components $f(y|x^*, w)$ and $f(w|x^*)$ are parameterized, the other components $f(x|x^*)$ and $f(x^*)$ are treated as non-parametric nuisance functions. We additionally parameterize $f(w|x^*)$ as $f(w|x^*; \beta_3, \beta_4) = \phi([w - \beta_3 x^*]/\beta_4)$. The density functions $f(x|x^*)$ and $f(x^*)$ inside the integration can be obtained from the proposed EDE procedure. We can use these estimated EDE estimators for $f(x|x^*)$ and $f(x^*)$ along with the parametric choice model $f(y|x^*, w; \beta_0, \beta_1, \beta_2)$, $f(w|x^*; \beta_3, \beta_4)$ to construct an MLE estimator for $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$. That is,

$$\begin{aligned} (\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2, \widehat{\beta}_3, \widehat{\beta}_4) = \underset{\beta_0,\beta_1,\beta_2,\beta_3,\beta_4}{\arg\max} \frac{1}{n} \sum_{i=1}^{n} \ln \int & f(y_i|x^*, w_i; \beta_0, \beta_1, \beta_2) \widehat{f}_{x_i|x^*}(x_i|x^*) \\ & \times f(w_i|x^*; \beta_3, \beta_4) \widehat{f}_{x^*}(x^*) dx^*, \end{aligned}$$

where $\widehat{f}_{x_i|x^*}(x_i|x^*)$ and $\widehat{f}_{x^*}(x^*)dx^*$ are obtained from the proposed EDE procedure.

Our EDE estimator builds upon the identification assumptions proposed in HS (2008) and articulated in the Online Appendix. Online Appendix Assumption S1.1 requires that all joint and marginal probability mass/density functions are bounded on their support. Online Appendix Assumption S1.2(i) implies that the variables x and z do not offer any additional information about the outcome variable y beyond what is already captured by variable x^*, while Online Appendix Assumption S1.2(ii) states that z does not provide any additional information about x than x^* already provides. These assumptions serve as standard exclusion restrictions. Since the z we use here is a predicted value from a regression, it is reasonable to assume that the least-squares projection has purged the instrument from components that would affect the probability of participating in the TPA program directly. Hence, Online Appendix Assumption S1.2 is plausibly satisfied. Online Appendix Assumption S1.3 is a technical assumption to allow some algebra manipulations. Online Appendix Assumption S1.4 requires that the distribution of y conditional on x^* is not identical at two values of x^*. The location restriction in Online Appendix Assumption S1.5 is related to Online Appendix Assumption S1.4. Online Appendix Assumption S1.5 keeps the ordering of the eigenvalues and eigenfunctions of AB^{-1} to identify $f^K(x|x^*)$ that converges to the true population function $f(x|x^*)$ as $K \to \infty$.

The data used in this application come from the 2015 China Household Finance Survey (CHFS).[13] The sample size of the 2015 survey is 40,000 households, covering 29 provinces (municipalities, districts), 363 counties, and 1,439 village or neighbourhood committees in China. The data cover various information, including household assets, liabilities and credit constraints,

[13] See the third round of the China Household Finance Survey (2015).

Table 1. Estimates of participation parameters.

	Probit	IV probit	EDE estimator		
			$K=2$	$K=3$	$K=4$
	(1)	(2)	(3)	(4)	(5)
Coefficient					
β_0	−0.525	1.933	−1.147	−1.213	−1.221
	(0.037)	(0.076)	(0.144)	(0.229)	(0.194)
β_1	−0.071	−0.296	0.118	0.077	0.060
	(0.003)	(0.006)	(0.384)	(0.145)	(0.223)
β_2	0.087	0.085	0.111	0.084	0.131
	(0.071)	(0.062)	(0.367)	(0.144)	(0.171)
PAE					
	−0.013	−0.120	0.025	0.015	0.011
	(0.000)	(0.006)	(0.036)	(0.023)	(0.046)

Note: The data come from the 2015 CHFS. The standard errors are reported in parentheses, which are calculated using the bootstrap method with 500 replications. PAE is a partial effect at the mean. PAEs of probit and IV probit are calculated using $\widehat{\beta}_1\phi(\widehat{\beta}_0+\widehat{\beta}_1\bar{x}+\widehat{\beta}_2\bar{w})$, where $\bar{x}=\frac{1}{N}\sum_{i=1}^{N}x_i$ and $\bar{w}=\frac{1}{N}\sum_{i=1}^{N}w_i$. The PAE of the EDE estimator is calculated using $\widehat{\beta}_1\phi(\widehat{\beta}_0+\widehat{\beta}_1\bar{x}^*+\widehat{\beta}_2\bar{w})$, where $\bar{x}^*=\int_0^1 x^*\widehat{f}(x^*)dx^*$.

income, consumption, social security and insurance, demographic characteristics, employment, payment habits, etc.[14]

We compare the results of our EDE estimator to those of a probit model using reported income alone (column (1) in Table 1), as well as a conventional IV probit estimator[15] (column (2) in Table 1) and the EDE estimator (columns (3)–(7) in Table 1).

The results from the probit model without using IV indicate a negative and statistically significant relationship between households' income and the probability of being selected as TPA beneficiaries. When using predicted income as an IV to address the endogeneity issue, the conventional IV estimator shows a stronger negative relationship. However, our EDE estimator yields substantially different results, finding that there is a positive correlation between households' actual income and the probability of TPA beneficiary classification, but the impact is not significant. Combining the estimates from the probit model, it becomes apparent that the probability of being selected as TPA beneficiaries is related to the observed income of households rather than their actual income. Our results corroborate previous research uniformly indicating poor targeting of resources in China's TPA program (Golan et al., 2017; Han and Gao, 2019; Kakwani et al., 2019). This result highlights the need for further improvements in the TPA program to ensure its effective targeting of intended beneficiaries. Policymakers must prioritize accuracy in measuring income to effectively target those in need. This can be achieved through the integration of multiple data sources and advanced statistical techniques.

[14] The specific screening process for TPA beneficiaries and descriptive statistics of the variables can be found in the Online Appendix materials.

[15] The conventional IV estimator for probit models makes some more parametric assumptions. First, it assumes a linear relationship between the endogenous variable x and the instrument variable z. Second, it assumes that the joint distribution of the disturbance terms is bivariate normal to construct the likelihood. Therefore, the conventional IV approach for probit models parametrically specifies the densities $f_{x|x^*}(x|x^*)$ and $f_{z|x^*}(z|x^*)$. In contrast, the proposed EDE estimator leaves $f_{x|x^*}(x|x^*)$ and $f_{z|x^*}(z|x^*)$ unspecified.

The empirical results could also be influenced by other factors. For example, it is possible that linearity does not hold in the outcome equation. Because of technical limitations, we are currently unable to include x^{*2} in the model. Besides, boundary effects could potentially lead to bias in our parameter estimates. It is important to note, however, that boundary effects are common challenges in non-parametric estimation, and they can still be a concern even when using the sieve method.

5. CONCLUSION

We develop a closed-form semi-parametric estimator for non-linear models with unobserved explanatory variables. Building upon existing identification results, we propose a consistent extreme-value sieve estimator that serves as a simpler alternative to the sieve maximum likelihood estimator for such models. We demonstrate how to construct a consistent non-iterative estimator defined as an optimal solution of some objective function. Our estimator is helpful for empirical work because it has a closed form and does not require a standard iterative optimization algorithm. Empirical applications demonstrate that our estimator performs well in practice.

ACKNOWLEDGEMENTS

We are grateful to the Editor, Dennis Kristensen, for a highly efficient reviewing process, as well as three anonymous referees for very useful comments and suggestions. Ji-Liang Shiu gratefully acknowledges the National Natural Science Foundation of China (Project No. 72073050).

REFERENCES

Ai, C. and X. Chen (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica 71*, 1795–843.

Bound, J., C. Brown and N. Mathiowetz (2001). Measurement error in survey data. In J. J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, *Volume 5*, 3705–843. Amsterdam: Elsevier.

Carroll, R., D. Ruppert, C. Crainiceanu, T. Tosteson and M. Karagas (2004). Nonlinear and non-parametric regression and instrumental variables. *Journal of the American Statistical Association 99*, 736–50.

Chen, X., H. Hong and D. Nekipelov (2011). Nonlinear models of measurement errors. *Journal of Economic Literature 49*, 901–37.

China Household Finance Survey (2015). Survey and Research Center for China Household Finance, Southwestern University of Finance and Economics, Chengdu Province, China. https://chfser.swufe.edu.cn/datas/Products/Datas/DataList (last accessed: 5 January 2023).

Gallant, A. R. and D. Nychka (1987). Semi-non-parametric maximum likelihood estimation. *Econometrica 55*, 363–90.

Golan, J., T. Sicular and N. Umapathi (2017). Unconditional cash transfers in China: Who benefits from the rural minimum living standard guarantee (dibao) program? *World Development 93*, 316–36.

Han, H. and Q. Gao (2019). Community-based welfare targeting and political elite capture: Evidence from rural China. *World Development 115*, 145–59.

Hausman, J., H. Ichimura, W. Newey and J. Powell (1991). Measurement errors in polynomial regression models. *Journal of Econometrics 50*, 271–95.

Horn, R. and C. Johnson (1985). *Matrix Analysis*. Cambridge: Cambridge University Press.
Hu, Y., D. McAdams and M. Shum (2013). Identification of first-price auctions with non-separable unobserved heterogeneity. *Journal of Econometrics 174*, 186–93.
Hu, Y. and G. Ridder (2012). Estimation of non-linear models with mismeasured regressors using marginal information. *Journal of Applied Econometrics 27*, 347–85.
Hu, Y. and Y. Sasaki (2015). Closed-form estimation of non-parametric models with non-classical measurement errors. *Journal of Econometrics 185*, 392–408.
Hu, Y. and S. Schennach (2008). Instrumental variable treatment of non-classical measurement error models. *Econometrica 76*, 195–216.
Kakwani, N., S. Li, X. Wang and M. Zhu (2019). Evaluating the effectiveness of the rural minimum living standard guarantee (dibao) program in China. *China Economic Review 53*, 1–14.
Krasnokutskaya, E. (2011). Identification and estimation of auction models with unobserved heterogeneity. *The Review of Economic Studies 78*, 293–327.
Lewbel, A. (2000). Identification of the binary choice model with misclassification. *Econometric Theory 16*, 603–9.
Lewbel, A. (2007). Estimation of average treatment effects with misclassification. *Econometrica 75*, 537–51.
Li, T. (2002). Robust and consistent estimation of non-linear errors-in-variables models. *Journal of Econometrics 110*, 1–26.
Mahajan, A. (2006). Identification and estimation of regression models with misclassification. *Econometrica 74*, 631–65.
Newey, W. (2001). Flexible simulated moment estimation of non-linear errors-in-variables models. *Review of Economics and Statistics 83*, 616–27.
Newey, W. and J. Powell (2003). Instrumental variable estimation of non-parametric models. *Econometrica 71*, 1565–78.
Schennach, S. (2004). Estimation of non-linear models with measurement error. *Econometrica 72*, 33–75.
Schennach, S. (2007). Instrumental variable estimation of non-linear errors-in-variables models. *Econometrica 75*, 201–39.
Shiu, J. and Y. Hu (2013). Identification and estimation of non-linear dynamic panel data models with unobserved covariates. *Journal of Econometrics 175*, 116–31.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Online Appendix
Replication Package

Co-editor Dennis Kristensen handled this manuscript.

APPENDIX: PROOF PROPOSITION 2.1

Consider

$$\sum_y \mathbf{1}(y=1) f_{xyz}(x,y,z) = \int f_{x|x^*} \cdot \left(\sum_y \mathbf{1}(y=1) f_{y|x^*} \right) \cdot f_{z|x^*} \cdot f_{x^*}\, dx^*$$
$$= \int f_{x|x^*} \cdot \Pr(y=1|x^*) \cdot f_{x^*} \cdot f_{z|x^*}\, dx^*. \quad \text{(A.1)}$$

For the EDE, we assume a normalization condition such that

$$\Pr(y=1|x^*) = x^*.$$

The EDE is based on sieve approximations of all the densities on the RHS of (1.1). We consider the sieve approximations for these probability mass/density functions and probability densities in (2.4), (2.5), (2.8), (2.9), and (2.10).

We first use a different number of terms M for $f_{x^*}(x^*)$ and later take $M = K+1$ to solve for its corresponding sieve coefficients. In $f_{x|x^*}(x|x^*) \approx \sum_{k=0}^{K} \sum_{j=0}^{K} e_{kj} \psi_k(x) \varphi_j(x^*)$, we use a polynomial series $\varphi_k(x^*) = (x^*)^k$ for $k = 0, 1, 2, \ldots$. Note that, for simplicity, we use the linear combination of basis functions to approximate a density itself instead of the square root of the density. This greatly simplifies the derivation.

From the data, we can recover the orthonormal coefficients a_{kj} and b_{kj} corresponding to the probability mass/density functions as

$$a_{kj} = \iint (\mathbf{1}(y=1) f_{xyz}(x,y,z)) \psi_k(x) \psi_j(z) dx dz = E[\mathbf{1}(y=1) \psi_k(x) \psi_j(z)], \quad \text{(A.2)}$$

$$b_{kj} = \iint \psi_k(x) \psi_j(z) f_{xz}(x,z) dx dz = E[\psi_k(x) \psi_j(z)]. \quad \text{(A.3)}$$

According to the law of large numbers, the average of the random events obtained from a large number of trials should be close to the expected value. Thus, coefficients a_{ij} and b_{ij} can be estimated using sample averages to approximate the expectations above. Thus, we have expressions (2.12) and (2.13). The approximation results are summarised as follows.

LEMMA A.1. *Define*

$$\widehat{f}^K_{x,y=1,z}(x, y=1, z) = \sum_{k=0}^{K} \sum_{j=0}^{K} \widehat{a}_{kj} \psi_k(x) \psi_j(z), \qquad \widehat{f}^K_{x,z}(x,z) = \sum_{k=0}^{K} \sum_{j=0}^{K} \widehat{b}_{kj} \psi_k(x) \psi_j(z).$$

We have $\|\widehat{f}^K_{x,y=1,z} - f_{x,y=1,z}\| = o_p(1)$ *and* $\|\widehat{f}^K_{x,z} - f_{x,z}\| = o_p(1)$.

The left-hand side of (A.1) can be estimated as

$$\begin{aligned} LHS &= \sum_y \mathbf{1}(y=1) f_{xyz}(x,y,z) \\ &\approx \sum_{k=0}^{K} \sum_{j=0}^{K} a_{kj} \psi_k(x) \psi_j(z) \\ &= \overrightarrow{\psi}(x)^T A \overrightarrow{\psi}(z) \\ &\equiv \widehat{LHS}, \end{aligned}$$

where $A = [a_{k-1,j-1}]_{k,j}$ and $\overrightarrow{\psi}(x) = [\psi_0(x), \ldots, \psi_K(x)]^T$. The right-hand side of (A.1) is

$$
\begin{aligned}
RHS &= \int f_{x|x^*} \cdot \Pr(y=1|x^*) \cdot f_{x^*} \cdot f_{z|x^*}\, dx^* \\
&\approx \int \left[\sum_{k=0}^{K}\sum_{j=0}^{K} e_{kj}\psi_k(x)\varphi_j(x^*)\right] \cdot x^* \cdot \left[\sum_{l=0}^{M} q_l\psi_l(x^*)\right] \cdot \left[\sum_{k=0}^{K}\sum_{j=0}^{K} h_{kj}\psi_k(z)\psi_j(x^*)\right] dx^* \\
&= \int [\overrightarrow{\psi}(x)^T E \overrightarrow{\varphi}(x^*)] \cdot x^* \cdot \left[\sum_{l=0}^{M} q_l\psi_l(x^*)\right] \cdot [\overrightarrow{\psi}(x^*)^T H^T \overrightarrow{\psi}(z)] dx^* \\
&= \overrightarrow{\psi}(x)^T E \left[\int \overrightarrow{\varphi}(x^*) \cdot x^* \cdot \left[\sum_{l=0}^{M} q_l\psi_l(x^*)\right] \cdot \overrightarrow{\psi}(x^*)^T dx^*\right] H^T \overrightarrow{\psi}(z) \\
&\equiv \overrightarrow{\psi}(x)^T E C_q H^T \overrightarrow{\psi}(z) \\
&\equiv \widehat{RHS},
\end{aligned}
$$

where

$$C_q = \left[\int_0^1 x^{*k} \cdot \left[\sum_{l=0}^{M} q_l\psi_l(x^*)\right] \cdot \psi_{j-1}(x^*) dx^*\right]_{k,j\in\{1,2,\ldots,K+1\}}$$

is a matrix that depends on the q and the approximating basis functions.

Equating

$$\widehat{LHS} = \widehat{RHS},$$

we obtain, for any x and z,

$$\overrightarrow{\psi}(x)^T A \overrightarrow{\psi}(z) = \overrightarrow{\psi}(x)^T E C_q H^T \overrightarrow{\psi}(z).$$

Therefore, we have

$$A = E C_q H^T. \tag{A.4}$$

Similarly, we have

$$f_{xz}(x,z) = \int f_{x|x^*} \cdot f_{x^*} \cdot f_{z|x^*}\, dx^*.$$

Using a similar argument as above with matrix C_q replaced by

$$C_{q0} = \left[\int_0^1 x^{*k-1} \cdot \left[\sum_{l=0}^{M} q_l\psi_l(x^*)\right] \cdot \psi_{j-1}(x^*) dx^*\right]_{k,j\in\{1,2,\ldots,K+1\}},$$

we obtain the matrix equation

$$B = E C_{q0} H^T. \tag{A.5}$$

Note that the kth row of C_q is the same as the $(k+1)$th row of C_{q0} for $k = 1, \ldots, K$. Therefore, the relation between matrices C_q and C_{q0} satisfies $C_q = G_q C_{q0}$ because the d_k, $k = 0, \ldots, K$, satisfy

$$\begin{aligned}
&(-d_0, -d_1, -d_2, \ldots, -d_K)C_{q0} \\
&\quad = \text{the } (K+1)\text{th row of } C_q \\
&\quad = \left(\int (x^*)^{K+1} \cdot \left[\sum_{l=0}^{M} q_l \psi_l(x^*) \right] \cdot \psi_0(x^*) dx^*, \ldots, \int (x^*)^{K+1} \cdot \left[\sum_{l=0}^{M} q_l \psi_l(x^*) \right] \cdot \psi_K(x^*) dx^* \right).
\end{aligned} \tag{A.6}$$

Because we can replace A and B with the expressions in (A.4) and (A.5), respectively, these equations contain sample information and we can utilize their structures to acquire the sieve coefficient matrices E, H, and $\overrightarrow{q}$.

Assumption A.1. *Suppose that*

(*i*) *the matrix of sieve coefficients of $f_{xz}(x, z)$, B, is invertible;*
(*ii*) *AB^{-1} has distinctive eigenvalues $\lambda_0 > \lambda_1 > \cdots > \lambda_K$.*

Intuitively, Assumption A.1(i) indicates that x contains enough information on z and vice versa. Assumption A.1(i) implies that we can use (A.4) and (A.5) to solve for the sieve coefficients e_{kj} in E and h_{kj} in H from the observed $\widehat{a}_{kj}$ in A, and $\widehat{b}_{kj}$ in B through

$$AB^{-1} = E(C_q C_{q0}^{-1})E^{-1} = EG_q E^{-1}. \tag{A.7}$$

The above equation implies that the observed matrix AB^{-1} is similar to matrix G_q, which is associated with the unknown $\overrightarrow{q}$.

Because AB^{-1} and G_q have the same set of eigenvalues, Assumption A.1(ii) ensures that both AB^{-1} and G_q have distinct eigenvalues and are diagonalizable. Thus, the eigen-decompositions of AB^{-1} and G_q exist. The assumption rules out eigenvalues with multiplicity greater than 1. Let AB^{-1} and G_q have the eigen-decompositions

$$AB^{-1} = VDV^{-1}, \tag{A.8}$$

$$G_q = S_q D S_q^{-1}. \tag{A.9}$$

Next, we show that q_l may be uniquely determined by the eigenvalues of AB^{-1}. Note that

$$\det(\lambda I - AB^{-1}) = \det(\lambda I - G_q) = \lambda^{K+1} + d_K \lambda^K + \cdots + d_1 \lambda + d_0.$$

This is in fact the characteristic polynomial of directly estimable AB^{-1}. Therefore, coefficients d_k are all directly estimable. In other words, we have

$$d_0 = (-1)^{K+1} \lambda_0 \lambda_1 \lambda_2 \cdots \lambda_K, \tag{A.10}$$

$$d_1 = (-1)^K \sum_{k=0}^{K} \lambda_0 \lambda_1 \cdots \lambda_{k-1} \lambda_{k+1} \cdots \lambda_K, \tag{A.11}$$

$$\vdots$$

$$d_K = (-1) \sum_{k=0}^{K} \lambda_k, \tag{A.12}$$

where the λ_k are eigenvalues of the observed matrix AB^{-1}. Therefore, we may treat d_k as known.

We then solve for q_l from d_k as follows. The definition of d_k in (A.6) implies that

$$\begin{aligned} LHS &= (-d_0, -d_1, -d_2, \ldots, -d_K)C_{q0} \\ &= (-d_0, -d_1, -d_2, \ldots, -d_K)\left[\int (x^*)^{k-1} \cdot \left[\sum_{l=0}^{M} q_l \psi_l(x^*)\right] \cdot \psi_{j-1}(x^*)dx^*\right]_{k,j\in\{1,2,\ldots,K+1\}} \\ &= \left[\int \left[-\sum_{k=0}^{K}(x^*)^k d_k\right] \cdot \left[\sum_{l=0}^{M} q_l \psi_l(x^*)\right] \cdot \psi_{j-1}(x^*)dx^*\right]_{\cdot,j\in\{1,2,\ldots,K+1\}} \\ &= \sum_{l=0}^{M} q_l \left[\int \left(-\sum_{k=0}^{K}(x^*)^k d_k\right) \cdot \psi_l(x^*) \cdot \psi_{j-1}(x^*)dx^*\right]_{\cdot,j\in\{1,2,\ldots,K+1\}}. \end{aligned} \tag{A.13}$$

The $(K+1)$th row of C_q is

$$\begin{aligned} RHS &= \left[\int (x^*)^{K+1} \cdot \left[\sum_{l=0}^{M} q_l \psi_l(x^*)\right] \cdot \psi_{j-1}(x^*)dx^*\right]_{\cdot,j\in\{1,2,\ldots,K+1\}} \\ &= \sum_{l=0}^{M} q_l \left[\int (x^*)^{K+1} \cdot \psi_l(x^*) \cdot \psi_{j-1}(x^*)dx^*\right]_{\cdot,j\in\{1,2,\ldots,K+1\}}. \end{aligned}$$

Therefore, we have

$$\begin{aligned} RHS - LHS &= \sum_{l=0}^{M} q_l \left[\int (x^*)^{K+1} \cdot \psi_l(x^*) \cdot \psi_{j-1}(x^*)dx^*\right]_{\cdot,j\in\{1,2,\ldots,K+1\}} \\ &\quad - \sum_{l=0}^{M} q_l \left[\int \left(-\sum_{k=0}^{K}(x^*)^k d_k\right) \cdot \psi_l(x^*) \cdot \psi_{j-1}(x^*)dx^*\right]_{\cdot,j\in\{1,2,\ldots,K+1\}} \\ &= \sum_{l=0}^{M} q_l \left[\int \left((x^*)^{K+1} + \sum_{k=0}^{K}(x^*)^k d_k\right) \cdot \psi_l(x^*) \cdot \psi_{j-1}(x^*)dx^*\right]_{\cdot,j\in\{1,2,\ldots,K+1\}} \\ &= \sum_{l=1}^{M} q_l \left[\int \left((x^*)^{K+1} + \sum_{k=0}^{K}(x^*)^k d_k\right) \cdot \psi_l(x^*) \cdot \psi_{j-1}(x^*)dx^*\right]_{\cdot,j\in\{1,2,\ldots,K+1\}} \\ &\quad + q_0 \left[\int \left((x^*)^{K+1} + \sum_{k=0}^{K}(x^*)^k d_k\right) \cdot \psi_l(x^*) \cdot \psi_{j-1}(x^*)dx^*\right]_{\cdot,j\in\{1,2,\ldots,K+1\}} \\ &= 0. \end{aligned}$$

Because the density restriction of $f(x^*)$, $\int(\sum_{l=0}^{M} q_l \psi_l(x^*))dx^* = 1$, with an orthogonal Fourier series ψ_l and $\psi_0(x^*) = 1$, we have $q_0 = 1$. Therefore,

$$\begin{aligned} &\sum_{l=1}^{M} q_l \underbrace{\left[\int \left((x^*)^{K+1} + \sum_{k=0}^{K}(x^*)^k d_k\right) \cdot \psi_l(x^*) \cdot \psi_{j-1}(x^*)dx^*\right]_{\cdot,j\in\{1,2,\ldots,K+1\}}}_{\text{defined as } c_{lj}} \\ &= -\underbrace{\left[\int \left((x^*)^{K+1} + \sum_{k=0}^{K}(x^*)^k d_k\right) \cdot \psi_0(x^*) \cdot \psi_{j-1}(x^*)dx^*\right]_{\cdot,j\in\{1,2,\ldots,K+1\}}}_{\text{defined as } c_{0j}}. \end{aligned}$$

We take $M = K + 1$. The equation above may be written as

$$\overrightarrow{q}^T [c_{lj}]_{l,j\in\{1,2,\dots,K+1\}} = -\overrightarrow{c}_0^T,$$

where, for $l \in \{0, 1, 2, \dots, K+1\}$ and $j \in \{1, 2, \dots, K+1\}$,

$$\begin{aligned}
[c_{lj}]_{l,j\in\{1,2,\dots,K+1\}} &= \left[\int \left((x^*)^{K+1} + \sum_{k=0}^{K}(x^*)^k d_k\right)\cdot \psi_l(x^*)\cdot \psi_{j-1}(x^*)dx^*\right] \\
&= \left[\int [\det(x^* I - AB^{-1})]\cdot \psi_l(x^*)\cdot \psi_{j-1}(x^*)dx^*\right], \\
\overrightarrow{q} &= [q_1, q_2, \dots, q_K, q_{K+1}]^T, \\
\overrightarrow{c}_0 &= [c_{01}, c_{02}, \dots, c_{0K}, c_{0K+1}]^T.
\end{aligned}$$

Note that $[c_{lj}]_{l,j\in\{1,2,\dots,K+1\}}$ and $\overrightarrow{c}_0$ are known because the estimators for d_k are available.

ASSUMPTION A.2. *Matrix $[c_{lj}]_{l,j\in\{1,2,\dots,K+1\}}$ is invertible.*

This implies that, under Assumption A.2, we may solve for q_l as

$$\overrightarrow{q}^T = -\overrightarrow{c}_0^T \left([c_{lj}]_{l,j\in\{1,2,\dots,K+1\}}\right)^{-1}.$$

With the estimated sieve coefficients $\overrightarrow{q}^T$, we obtain the sieve estimator $f^K(x^*)$. Because the matrix $[c_{lj}]_{l,j\in\{1,2,\dots,K+1\}}$ is constructed through $f_{xyz}(x, y = 1, z)$, $f_{xz}(x, z)$, and $\Pr(y = 1|x^*) = x^*$, the assumption implies that these probability mass/density functions provide enough information to recover $f^K(x^*)$, following appropriate manipulations. Next, matrices C_q and C_{q0} can also be obtained by the estimated sieve coefficients $\overrightarrow{q}^T$.

ASSUMPTION A.3. *Matrix C_{q0} is invertible.*

Then, we may treat $G_q = C_q C_{q0}^{-1}$ as known, as well as its eigen-decomposition (A.9). Combining (A.7) with eigen-decompositions (A.8) and (A.9) yields

$$\underbrace{VDV^{-1}}_{\text{an eigen-dcomposition of } AB^{-1}} = AB^{-1} = EG_qE^{-1} = E\underbrace{S_qDS_q^{-1}}_{\text{an eigen-dcomposition of } G_q}E^{-1}.$$

Note that V and S_q are the eigenvector matrices of the known matrices AB^{-1} and G_q, respectively.[16] This implies that the two eigenvector matrices V and ES_q equal each other up to the scale of the eigenvectors and we may then have

$$ES_q = V\widetilde{\widetilde{D}} \tag{A.14}$$

for some unknown diagonal matrix $\widetilde{\widetilde{D}}$. In order to pin down the unknown $\widetilde{\widetilde{D}}$, we need to use the density property that $\int f^K(x|x^*)dx = 1$ for any x^*. This is so that the elements of the coefficient matrix E for $f^K(x|x^*)$ satisfy

$$f^K(x|x^*) = \sum_{k=0}^{K}\sum_{j=0}^{K} e_{kj}\psi_k(x)\varphi_j(x^*)$$

with $\int f^K(x|x^*)dx = 1$ for any x^*. This restriction implies that, for any x^*,

$$\sum_{k=0}^{K}\left(\sum_{j=0}^{K} e_{kj}\int \psi_k(x)dx\right)\varphi_j(x^*) = 1,$$

[16] The relations are given in (A.8) and (A.9).

but $\varphi_0(\cdot) = 1$ and $\varphi_k(\cdot) \neq 1$ for $k \geq 1$ with the polynomial series above. Therefore, we have

$$1 = \left(\sum_{k=0}^{K} e_{k0} \int \psi_k(x)dx\right) + \left(\sum_{k=0}^{K} e_{k1} \int \psi_k(x)dx\right)\varphi_1(x^*) + \cdots + \left(\sum_{k=0}^{K} e_{kK} \int \psi_k(x)dx\right)\varphi_K(x^*).$$

The above equation holds for any x^* so that the coefficients on $\varphi_k(x^*)$ satisfy

$$\sum_{k=0}^{K} e_{kj} \int \psi_k(x)dx = \begin{cases} 1 & \text{for } j = 0, \\ 0 & \text{for } j = 1, \ldots, K, \end{cases}$$

or

$$\overrightarrow{d}_{\psi}^{T} E = \overrightarrow{e}_{1}^{T},$$

where $\overrightarrow{d}_{\psi} = [\int \psi_0(x)dx, \ldots, \int \psi_K(x)dx]^T = [1, 0, 0, \ldots, 0]^T$ and $\overrightarrow{e}_1 = [1, 0, 0, \ldots, 0]^T$. Therefore, by (A.14) we have

$$\overrightarrow{e}_{1}^{T} = \overrightarrow{d}_{\psi}^{T} E = \overrightarrow{d}_{\psi}^{T} V \widetilde{\widetilde{D}} S_q^{-1}.$$

This implies that

$$\underbrace{\overrightarrow{e}_{1}^{T} S_q}_{\text{first row of } S_q} = \underbrace{\overrightarrow{d}_{\psi}^{T} V}_{\text{first row of } V} \widetilde{\widetilde{D}}.$$

Let $[P]_{ij}$ stand for the (i, j)th element of matrix P.

Assumption A.4. *Suppose that* $\left[S_q\right]_{1,i} \neq 0$ *and* $[V]_{1,i} \neq 0$ *for all* $i \in \{1, 2, \ldots, K+1\}$.

To solve for E, we need to establish uniqueness of representation (A.7). The assumption is to help us resolve the ordering or indexing ambiguity of the eigenvalues and eigenfunctions of AB^{-1}. This assumption allows us to generate the ith diagonal element of the unknown diagonal matrix $\widetilde{\widetilde{D}}$ as

$$[\widetilde{\widetilde{D}}]_{ii} = \frac{[S_q]_{1,i}}{[V]_{1,i}}.$$

Assumption A.4 also guarantees that $\widetilde{\widetilde{D}}$ is invertible. We then have $E = V\widetilde{\widetilde{D}}S_q^{-1}$. This means that we may estimate

$$f^K(x|x^*) = \sum_{k=0}^{K}\sum_{j=0}^{K} e_{kj}\psi_k(x)\varphi_j(x^*).$$

As for H in $f^K(z|x^*)$, we also need the invertibility of C_{q0} in Assumption A.3. Matrix C_{q0} is related to $f^K(x^*)$ and its invertibility helps us recover the information of $f(z|x^*)$ through (A.5) with the known matrices B and E. Note that matrix E is also invertible because the diagonal matrix $\widetilde{\widetilde{D}}$ is also invertible by Assumption A.4. We then have $H = (C_{q0}^{-1}E^{-1}B)^T$ for $f^K(z|x^*) = \overrightarrow{\psi}_K(z)^T H \overrightarrow{\psi}_K(x^*)$. Therefore, we have identified matrices E, H, and $\overrightarrow{q}$. Applying these matrix relations to the estimated coefficients $\widehat{a}_{kj}$ and $\widehat{b}_{kj}$ yields the unique matrices $\widehat{E}$, $\widehat{H}$, and $\widehat{\overrightarrow{q}}$, and these matrices give the closed-form representation for $\widehat{\alpha}_{ede}$.