Optimal Linear Rank Indexes for Latent Variables*

Yingyao Hu Ji-Liang Shiu Yi Xin Jiaxiong Yao

June 11, 2024

Abstract

Understanding the quality and performance differences among individuals, institutions, and countries is crucial for many economic applications. However, this is inherently challenging because analysts often only have access to proxy measures rather than direct observations of true quality and performance. In this paper, we develop a linear index model to rank the values of a latent variable based on at least three measurements. Monte Carlo simulations demonstrate that our proposed sieve-based optimal linear rank index estimator performs well with finite samples. We apply our method to evaluate true GDP growth across countries using various economic activity measurements and find that satellite-recorded luminosity and Google search volume indices can significantly complement official GDP statistics.

Keywords: Latent Variable, Optimal Linear Rank Index, Maximum Rank Correlation, Nonclassical Measurement Error.

*Hu: Department of Economics, Johns Hopkins University (email: yhu@jhu.edu); Shiu: Institute for Economic and Social Research, Jinan University (email: jishiu.econ@gmail.com); Xin: Division of the Humanities and Social Sciences, California Institute of Technology (email: yixin@caltech.edu); Yao: International Monetary Fund, 700 19th St NW, Washington, DC 20431 (email: jyao@imf.org). We greatly benefited from discussions with Bob Sherman. The usual disclaimer applies.

1. Introduction

A thorough grasp of the quality and performance variations among individuals (such as students and employees), institutions (like hospitals and universities), and countries is crucial for numerous economic applications. However, this is inherently challenging because analysts typically have access only to a set of proxy measures, rather than directly observing true quality and performance metrics. For instance, in educational research, although comparing latent cognitive skills among children would provide an accurate assessment of their cognitive development, researchers can only measure their performance through test scores. In finance, assessing borrowers' creditworthiness is crucial for setting appropriate interest rates and making lending decisions. Since true underlying creditworthiness cannot be directly observed, researchers must rely on data such as the borrower's past payment histories, outstanding debt balances, employment status, and other tangible variables. At a macro level, policymaking depends on understanding the true state of the economy and its global standing. However, true economic growth is imperfectly measured by official GDP statistics and alternative economic indicators, such as satellite-recorded luminosity, CO2 emissions, and Google search volume indices. To what extent can researchers rely on a set of imperfectly measured proxies to rank an unobservable latent variable?

In this paper, we introduce a novel and easy-to-implement method for ranking the values of a latent variable using a linear index model with multiple measurements. Our primary motivation is to develop a quantitative measure based on a series of observed data that can reveal relative positions, such as those of a country, in a specific area. We consider a case where there are K measurements of a scalar latent variable $X^* \in \mathscr{X}^*$. Let $\boldsymbol{X} = (X_1, X_2, \cdots, X_K) \in \mathscr{X}$ denote the vector of the K measurements observable to econometricians. A linear index is defined as a linear combination of these measurements, i.e.,

$$\boldsymbol{X}\boldsymbol{\beta} = \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_K X_K, \tag{1.1}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_K)^T$ with $\beta_k > 0$ for $k = 1, 2, \cdots, K$ and $\sum_{k=1}^{K} \beta_k = 1$. The linear index transforms the measurement vector into a single scalar, making comparisons across observations straightforward. The *optimal linear rank index* is denoted by $\boldsymbol{X}\boldsymbol{\beta}_0$, such that for any two random draws i and j,

$$\boldsymbol{\beta}_0 = \arg\max_{\boldsymbol{\beta}} \quad Pr\Big(\boldsymbol{X}_i\boldsymbol{\beta} > \boldsymbol{X}_j\boldsymbol{\beta} | X_i^* > X_j^*\Big). \tag{1.2}$$

Intuitively, the proposed optimal linear rank index aims to maximize the *true positive rate*, which is the probability that X_i^* is correctly ranked above X_j^* based on observable information when $X_i^* > X_j^*$ in the data generating process. This objective ensures that the ranking of the true latent variable of interest, based on the ordering labels assigned by our proposed estimator, is preserved as accurately as possible. We show in the paper that it is equivalent to constructing $\boldsymbol{\beta}_0$ by maximizing the *rank correlation* between the latent variable and the linear combination of the measurements. $\boldsymbol{\beta}_0$ effectively maximizes the degree of similarity between two rankings. This single index condenses multi-dimensional data, making it easier to interpret and aiding in the task of ranking entities on complex issues.

We provide a two-step nonparametric identification strategy for $\boldsymbol{\beta}_0$. We combine the identification results for general nonlinear models with measurement errors (Hu and Schennach, 2008) with the methods in generalized linear regression models (Han, 1987). In the first step, we demonstrate that the conditional distribution of the latent variable X^* given the measurement X_k (i.e., $f_{X^*|X_k}$) can be recovered from a dataset containing at least three measurements of the latent variable. This result relies on the assumption that observable measurements are independent conditional on the latent variable. We also consider a special case where the measurement error structure is specified through a nonparametric regression model, i.e. $X_k = g_k(X^*) + \varepsilon_k$. Under additivity, certain technical assumptions, such as injectivity, required for the general nonlinear model can be relaxed. In the second step, we show that the population parameter $\boldsymbol{\beta}_0$ attains a unique maximum for the objective function in Equation (1.2), which is closely related to the objective function of the maximum rank correlation estimator studied in Han (1987) and Sherman (1993), and can be constructed using the joint distribution of the latent variable and its measurements identified in the first step.

Our identification approach is constructive and provides a corresponding two-step sieve-based rank estimator for $\boldsymbol{\beta}_0$. We first estimate $f_{X_k|X^*}$ for $k = 1, \cdots, K$ using a sieve MLE and construct a consistent estimator for $f_{X^*|\boldsymbol{X}}$. An estimator for $\boldsymbol{\beta}_0$ can then be constructed as the unique maximizer of the sample analogue of the population objective function. Since the proposed estimator is a variant of a maximum rank correlation estimator, we follow Sherman (1993) to provide a set of conditions that allow us to establish consistency and asymptotic normality of the optimal linear rank index estimator.

In the numerical simulation exercises, we compare the performance of our proposed optimal linear rank estimator with that of the infeasible estimator (which assumes the joint distribution of X^* and its measurements is known), the principal component analysis (PCA) estimator, and the independent component analysis (ICA) estimator under different data-generating processes. Both PCA and ICA are commonly used dimensionality reduction algorithms. PCA transforms a set of correlated variables into a smaller number of uncorrelated variables while retaining as much variation as possible from the original dataset. This approach is useful for identifying latent variables (factors) underlying the observed data. ICA is an extension of the PCA technique. This algorithm assumes that each sample of data is a mixture of independent components and it extracts these components by maximizing the non-Gaussianity of these random variables. Although these estimators are not designed to maximize the probability of correctly ranking the latent factor based on observables, they are often used empirically to construct indexes for comparing latent variables, such as air and water quality (Mahapatra, Sahu, Patel, and Panda, 2012), socioeconomic status (Kolenikov and Angeles, 2009), stock market volatility (Li, Ma, Zhang, and Xiao, 2019), and eco-efficiency (Jollands, Lermit, and Patterson, 2004).[1] Policymakers have also been utilizing these tools to construct composite indicators to "compare and rank country performance in areas such as industrial competitiveness, sustainable development, globalisation and innovation" (Nardo, Saisana, Saltelli, Tarantola, Hoffman, and Giovannini, 2005).

We find that our optimal linear rank index estimator performs well in finite samples. The bias of the proposed linear rank index estimator is small and comparable to that of the infeasible estimator. The standard deviation of the linear rank index estimator decreases as the sample size increases.. In all simulation designs, the ranking of the latent variable across observations is largely preserved with the ordering labels assigned by our proposed

[1] Numerous empirical examples exist where PCA and ICA are used to construct rank indices. Relevant studies in environmental studies include Khatun (2009), Hao, Li, Li, Zhang, and Liu (2013), and Tripathi and Singal (2019). Filmer and Pritchett (2001), Vyas and Kumaranayake (2006), Krishnan (2010), and Friesen, Seliske, and Papadopoulos (2016) use PCA to create composite indexes of socioeconomic status. Agasisti and Pérez-Esparrells (2010), Muzamhindo, Kong, and Famba (2017), and Yi (2019) use PCA to rank educational institutions. Back and Weigend (1997) provides an example where they use ICA to extract features to rank stocks based on their underlying risk factors.

estimator, correctly predicting about 85% of the ranking orders of the latent variable. The percentage of correct predictions using ICA and PCA estimators is much lower in all simulations.

We then apply our method to real data to assess GDP growth across countries using multiple measurements including the official GDP growth rate, satellite-recorded luminosity, carbon dioxide emissions, and the Google search volume index. Our estimation results show that when accounting for the nonlinearity in the relationship between true economic growth and its measurements, luminosity growth is as informative about underlying economic growth as the official GDP data. Additionally, at a quarterly frequency, we find that the Google search volume index plays a more significant role in predicting economic growth than other measures, particularly in emerging markets and developing economies. The GDP growth index constructed using our estimator complements official data by revealing unrecorded economic activities, such as those during the demonetization period and the pandemic period in India.

This paper is closely related to the literature on identifying and estimating the latent variable distribution in nonlinear models with nonclassical measurement errors (Hu and Schennach, 2008; Li, 2002; Schennach, 2004; Chen, Hu, and Lewbel, 2009; Schennach and Hu, 2013; and Hu, Schennach, and Shiu, 2021). The construction of our optimal linear rank index estimator relies on recovering the joint distribution of the latent variable and its measurements. The linear rank index defined in Equation (1.2) is related to the maximum rank correlation (MRC) estimator proposed by Han (1987) for the generalized regression model, where the rank correlation between y_i and $x_i'\beta$ is maximized. Sherman (1993) shows that the MRC estimator is $\sqrt{n}$-consistent and asymptotically normal. Chen (2002) proposes a rank-based estimator for the transformation function in the generalized regression model, given a $\sqrt{n}$-consistent estimator for $\boldsymbol{\beta}_0$. Shin (2010) proposes a local rank correlation estimator for estimating transformation models with varying coefficients. Drawing insights from these various strands of literature, our paper develops a simple linear index model to rank the values of a latent variable using multiple imperfect measurements.

The rest of the paper is organized as follows. We present the main identification results in Section 2. The estimation procedure for the optimal linear rank index is described in Section 3. We discuss the Monte Carlo simulation exercises in Section 4 and the empirical illustration of our approach is presented in Section 5. Section 6 concludes. The technical details and asymptotic properties of the proposed estimator are provided in Appendix B. The proofs are provided in Appendix C.

2. Identification

In this section, we provide the identification results for the optimal linear rank index parameter $\boldsymbol{\beta}_0$ defined in Equation (1.2). The true positive rate can be represented as

$$Pr\left(\boldsymbol{X}_i\boldsymbol{\beta} > \boldsymbol{X}_j\boldsymbol{\beta} | X_i^* > X_j^*\right) = \frac{E[\mathbf{1}(\boldsymbol{X}_i\boldsymbol{\beta} > \boldsymbol{X}_j\boldsymbol{\beta})\mathbf{1}(X_i^* > X_j^*)]}{Pr\left(X_i^* > X_j^*\right)},$$

where $\mathbf{1}(\cdot)$ denotes an indicator function. Since the linear coefficient $\boldsymbol{\beta}$ only enters the numerator, it is equivalent to define $\boldsymbol{\beta}_0$ as

$$\boldsymbol{\beta}_0 = \arg\max_{\boldsymbol{\beta}} \; Q_0(\boldsymbol{\beta}), \tag{2.1}$$

where

$$\begin{aligned} Q_0(\boldsymbol{\beta}) &\equiv E[\mathbf{1}(\boldsymbol{X}_i\boldsymbol{\beta} > \boldsymbol{X}_j\boldsymbol{\beta})\mathbf{1}(X_i^* > X_j^*)] \\ &= E_{\boldsymbol{X}_i,\boldsymbol{X}_j}\left[\iint \mathbf{1}(\boldsymbol{X}_i\boldsymbol{\beta} > \boldsymbol{X}_j\boldsymbol{\beta})\mathbf{1}(x_i^* > x_j^*) f_{X^*|\boldsymbol{X}}(x_i^*|\boldsymbol{X}_i) f_{X^*|\boldsymbol{X}}(x_j^*|\boldsymbol{X}_j) dx_i^* dx_j^*\right]. \end{aligned} \tag{2.2}$$

The objective function in Equation (2.1) measures the rank correlation between the latent variable and the linear index. Equation (2.2) holds because i and j are randomly drawn from an i.i.d. sample. Evaluating $Q_0(\boldsymbol{\beta})$ requires the distribution of the latent variable conditional on the vector of observable measurements, i.e., $f_{X^*|\boldsymbol{X}}$.

In the rest of this section, we first discuss the identification of the conditional density distribution $f_{X^*|\boldsymbol{X}}$. In the second step, we show that $\boldsymbol{\beta}_0$ attains a unique maximum of $Q_0(\boldsymbol{\beta})$ under certain normalization and regularity conditions. The two-step strategy yields the identification of the optimal linear rank index parameter $\boldsymbol{\beta}_0$.

2.1. Identification of $f_{X^*|\boldsymbol{X}}$

Identifying the joint distribution of X^* and $\boldsymbol{X}$ from the observed distribution of $\boldsymbol{X}$ is feasible only if the joint distribution admits a certain structure. We follow Hu and Schennach (2008) to provide a set of sufficient conditions to nonparametrically identify $f_{X^*|\boldsymbol{X}}$ for general nonlinear models given at least three measurements of the latent variable.

Assumption 2.1 (Conditional Independence)**.** *The measurements $X_1, X_2, \cdots, X_K$ with $K \geq 3$ are independent conditional on X^*.*

Assumption 2.2 (Bounded Densities)**.** *The joint distribution of $\boldsymbol{X}$ and X^* admits a bounded density on $\mathscr{X} \times \mathscr{X}^*$. All marginal and conditional densities are also bounded.*

Define an integral operator $L_{f_{U|V}}$ as

$$(L_{f_{U|V}} h)(u) \quad = \quad \int f_{U|V}(u|v)h(v)dv,$$

where $f_{U|V}$ is the density function of random variable U conditional on V.

Assumption 2.3 (Invertibility)**.** *The operators $L_{f_{X_1|X^*}}$ and $L_{f_{X_3|X_1}}$ are injective.*

Assumption 2.4 (Distinct Eigenvalues)**.** *The set $\left\{x_2 : f_{X_2|X^*}(x_2|x^*) \neq f_{X_2|X^*}(x_2|\tilde{x}^*)\right\}$ has positive probability (under the marginal of X_2) whenever $x^* \neq \tilde{x}^* \in \mathscr{X}^*$.*

Assumption 2.5 (Normalization)**.** *There exists a known functional M such that for any $x^* \in \mathscr{X}^*$, $M\left[f_{X_1|X^*}(\cdot|x^*)\right] = x^*$.*

Assumption 2.1 requires that the measurements are independent conditional on the latent variable. This restriction significantly reduces the number of unknown parameters in the joint distribution of X^* and $\boldsymbol{X}$. Consider the joint distribution of any three measurements, i.e., X_1, X_2, and X_3. By the law of total probability and Assumption 2.1,

$$f_{X_1,X_2,X_3}(X_1,X_2,X_3) = \int_{\mathscr{X}^*} f_{X_1|X^*}(X_1|x^*) f_{X_2|X^*}(X_2|x^*) f_{X_3|X^*}(X_3|x^*) f_{X^*}(x^*) dx^*. \tag{2.3}$$

Assumption 2.2 ensures that all the related densities are bounded. Based on Equation (2.3), we show that the identification of the conditional density $f_{X_1|X^*}$ and the latent variable density f_{X^*} relies on the spectral decomposition[2] of a linear operator generated from the observed density f_{X_1,X_2,X_3}. Notice that under the conditional independence assumption, identification of $f_{X_1|X^*}$ implies the identification of $f_{X_2,\cdots,X_k,X^*}$ because

$$f_{X_1,X_2,\cdots,X_k} = L_{f_{X_1|X^*}} f_{X_2,\cdots,X_k,X^*}, \tag{2.4}$$

[2]A spectral decomposition is the operator analog of the eigenvalue–eigenvector decomposition for matrices, in the finite-dimensional case.

and Assumption 2.3 ensures the invertibility of $L_{f_{X_1|X^*}}$. Therefore, we can identify the whole joint distribution $f_{X_1,X_2,\cdots,X_k,X^*} = f_{X_1|X^*} f_{X_2,\cdots,X_k,X^*}$. Once the joint distribution of all measurements and the latent variable is identified, it is straightforward to identify the conditional densities $f_{X_k|X^*}$ for all k and the density of the latent variable f_{X^*}.

Assumption 2.3 provides a high-level invertibility condition necessary for achieving the desired spectral decomposition representation. Intuitively, it requires that the measurement “transmits" information about the latent variable, meaning the distribution of the measurements changes uniquely with the value of X^*. Assumption 2.4 further ensures that the spectral decomposition has distinct eigenvalues. When the eigenvalues are the same for multiple values of X^* , the corresponding eigenfunctions are only determined up to an arbitrary linear combination, implying that they are not identified. Assumption 2.4 can be relaxed by replacing $f_{X_2|X^*}$ with $f_{X_2,X_4,X_5,\cdots,X_k|X^*}$. Finally, Assumption 2.5 imposes a location normalization to pin down the values of the unobserved X^* relative to the observed variables.

Theorem 2.1 (General Nonlinear Models)**.** *Under Assumptions 2.1, 2.2, 2.3, 2.4, and 2.5, the joint density of K observable measurements $f_{X_1,X_2,\cdots,X_K}(X_1,X_2,\cdots,X_K)$ uniquely determines the conditional densities $f_{X_1|X^*}(X_1|X^*)$, $f_{X_2|X^*}(X_2|X^*)$, $\cdots$, $f_{X_K|X^*}(X_K|X^*)$, and the density of the latent variable $f_{X^*}(X^*)$.*

Proof See Online Appendix C.1. *Q.E.D.*

The nonparametric identification results in Theorem 2.1 requires high-level assumptions that may not be empirically verified. Therefore, we also consider identifying a popular subclass of the model that imposes an additive structure on the relationship between the latent variable and its measurements. Under additivity, we are able to replace some technical assumptions (e.g., injectivity) with lower-level testable conditions, which are more appealing to applied researchers. From a practitioner's perspective, the estimation procedure for a flexibly specified additive model is also computationally much simpler (discussed in more detail in Section 3).

Assumption 2.6 (Additivity)**.** $X_k = g_k(X^*) + \varepsilon_k$ *for* $k = 1,\cdots,K$, *with* $\varepsilon_k \perp X^*$ *and* $E[\varepsilon_k] = 0$.

Assumption 2.7 (Invertibility)**.** *(i) the range of the functions g_1 and g_3 is the whole real line; (ii) the characteristic functions of the measurements X_1 and X_3 do not vanish on the real line; (iii) ε_1 and ε_3 have the support of the whole real line.*

Assumption 2.8 (Monotonicity)**.** *The functions g_1, g_2, and g_3 are strictly monotonic.*

Assumption 2.6 specifies the measurement error structure through a nonparametric regression model, where $g_k(\cdot)$ is an unknown function and ε_k represents an additive random shock which has zero mean and is independent of the latent variable. Under additivity, Assumptions 2.7 and 2.8 provide the sufficient conditions for the high-level invertibility assumption for $L_{f_{X_1|X^*}}$ and $L_{f_{X_3|X_1}}$ in Theorem 2.1 (i.e., Assumption 2.3). The condition that the characteristic functions of X_1 and X_3 do not equal to zero on the real line is testable from the data.[3] Examples of distributions satisfying this condition include normal, Chi-squared, Cauchy, Gamma, Exponential distributions, as well as any asymmetric distributions with bounded supports.

Assumption 2.8 requires that at least three of the g_k functions are strictly monotonic, which implies that the measurements are informative about the true latent variable. For example, in cognitive skill assessment, typical measurements include test scores in grammar, numeracy, reading, spelling, and writing. Assuming these tests reflect underlying cognitive abilities to some extent, the average test scores increase with cognitive skill level.[4] The monotonicity of g_1 implies that the inverse g_1^{-1} exists. Under this assumption, we can use g_1 to rescale the measurement error equations and obtain

$$X_k = g_k(g_1^{-1}(g_1(X^*))) + \varepsilon_k$$

for $k = 2, \cdots, K$. It is clear from this equation that the scale of the latent variable and the levels of the g_k functions cannot be jointly identified. We therefore, without loss of generality, normalize $g_1(X^*)$ to X^*, i.e., $X_1 = X^* + \varepsilon_1$. Under this normalization, the conditional mean function of X_1 given X^* is $E(X_1|x^*) = x^*$, which strictly increases in x^*. By defining M as the conditional mean function, we have $M\left[f_{X_1|X^*}(\cdot|x^*)\right] = E(X_1|x^*) = x^*$, as required by Assumption 2.5. Moreover, Assumption 2.4 is satisfied under the additive model because

[3] Hu and Shiu (2022) develops a test for the non-vanishing property of the characteristic functions. Their test applies to cases where the first zero point of the squared modulus of a characteristic function is finite and is an isolated point. For example, the characteristic function of a random variable $X \sim U(-1,1)$ falls into this category because $\phi_X(t) = \frac{\sin(t)}{t}$ equals zero at $t = n\pi$ for $n = \pm 1, \pm 2, \pm 3, \cdots$. More discussions on the non-vanishing property of characteristic functions can be found in Hu and Shiu (2022) and Ushakov (2011).

[4] While the monotonicity assumption is intuitive and likely to hold in many empirical contexts, it cannot be directly tested from the data because of the latent variable X^*. Empirically, we need to carefully consider the relationship between the observed measurements X_k and the latent variable X^*, and choose an appropriate measurement for normalization.

g_2 is strictly monotonic by Assumption 2.8.

Theorem 2.2 (Additive Models)**.** *Under Assumptions 2.1, 2.2, 2.6, 2.7, and 2.8, the observed joint density* $f_{X_1,X_2,\cdots,X_K}(X_1,X_2,\cdots,X_K)$ *uniquely determines* $g_k(X^*)$ *for* $k=2,\cdots,K$*, the density of* ε_k *for* $k=1,\cdots,K$*, and the density of the latent variable* $f_{X^*}(X^*)$*.*

Proof See Online Appendix C.2. *Q.E.D.*

Theorem 2.2 provides identification results for the nonparametric g_k functions, the error term distribution f_{ε_k}, and the density of the latent variable f_{X^*} under an additive measurement error structure. Although the additive structure might seem restrictive in some cases, the empirically testable conditions it provides are appealing to applied researchers. The additive structure is commonly used in many measurement error models. In classical measurement error models, $X_k = X^* + \varepsilon_k$, where $\varepsilon_k \perp X^*$. This implies that $g_k(X^*) = X^*$. In general, when $g_k(X^*) \neq X^*$ our additive structure represents a nonclassical measurement error model. To see this, note that $X_k = X^* + (g_k(X^*) - X^* + \varepsilon_k)$. The new error term $(g_k(X^*) - X^* + \varepsilon_k)$ is correlated with the true latent variable X^*. Since the variance of the error term depends on X^*, this specification allows for heteroskedasticity.[5]

To summarize, Theorem 2.1 or 2.2 shows that the joint distribution of $\boldsymbol{X}$ and X^* is uniquely determined by the observed distribution of $\boldsymbol{X}$. The conditional density of the latent variable given the measurements is also identified using the Bayes' rule:

$$f_{X^*|\boldsymbol{X}}(X^*|\boldsymbol{X}) = \frac{f_{\boldsymbol{X},X^*}(\boldsymbol{X},X^*)}{\int_{\mathcal{X}^*} f_{\boldsymbol{X},X^*}(\boldsymbol{X},x^*)dx^*}, \tag{2.5}$$

with which we can evaluate the objective function $Q_0(\boldsymbol{\beta})$ for a given $\boldsymbol{\beta}$ in Equation (2.2).

2.2. Identification of $\boldsymbol{\beta}_0$

To show that $\boldsymbol{\beta}_0$ is the unique maximum of the objective function $Q_0(\boldsymbol{\beta})$, we invoke the following assumptions.

Assumption 2.9. *The optimal linear rank index parameter* $\boldsymbol{\beta}_0 = (\beta_{01}, \beta_{02}, \cdots, \beta_{0K})'$*, satisfies* $\beta_{0k} > 0$ *for* $k = 1,2,\cdots,K$*, and* $\sum_{k=1}^{K} \beta_{0k} = 1$.

[5] The identification of a general nonparametric additive model with heteroskedasticity, i.e., $X_k = g_k(X^*) + \delta_k(X^*)\varepsilon_k$, requires stronger assumptions. For this type of models to be identifiable, it is crucial that the associated linear operator, $L_{f_{X_k|X^*}}$, is injective (see Assumption 2.3 we impose for the general nonlinear models).

Assumption 2.10. *(i) The support $\mathscr{X}$ is not contained in a proper linear subspace of $\mathbb{R}^K$; (ii) for some k, the k-th component of $\boldsymbol{X}$ has an everywhere positive Lebesgue density conditional on almost every value of $\boldsymbol{X}_{-k} = (X_1, \cdots, X_{k-1}, X_{k+1}, \cdots, X_K)$.*

Assumption 2.9 restricts $\boldsymbol{\beta}_0$ to the unit interval for normalization purposes. Note that for any positive constant c, we have $E[\mathbf{1}(\boldsymbol{X}_i c\boldsymbol{\beta} > \boldsymbol{X}_j c\boldsymbol{\beta})\mathbf{1}(X_i^* > X_j^*)] = E[\mathbf{1}(\boldsymbol{X}_i\boldsymbol{\beta} > \boldsymbol{X}_j\boldsymbol{\beta})\mathbf{1}(X_i^* > X_j^*)]$. Thus for identification, we normalize the optimal linear rank index parameter by restricting $\sum_{k=1}^{K}\beta_{0k} = 1$. The assumption that β_{0k} is strictly positive implies that the k-th measurement is informative about the ranking of the latent variable. Assumption 2.10, the key full-rank assumption for identifying $\boldsymbol{\beta}_0$ is borrowed from Han (1987), which essentially requires that there should be no collinearity among the observable measurements. We state our main identification results for the optimal linear index parameter $\boldsymbol{\beta}_0$ in the following theorem.

Theorem 2.3 (Identification)**.** *Suppose all assumptions of Theorem 2.1 or Theorem 2.2 hold. Under Assumptions 2.9 and 2.10, the optimal linear rank index parameter $\boldsymbol{\beta}_0$ is identified.*

Showing that $\boldsymbol{\beta}_0$ uniquely maximizes $E[\mathbf{1}(\boldsymbol{X}_i\boldsymbol{\beta} > \boldsymbol{X}_j\boldsymbol{\beta})\mathbf{1}(X_i^* > X_j^*)]$ under Assumptions 2.9 and 2.10 is a direct application of Han (1987) (see part 1 of his proof for the main theorem). We therefore omit the proof of Theorem 2.3.

3. Estimation

Based on the identification results in Section 2, we propose a two-step sieve-based estimator for the optimal linear rank index parameter $\boldsymbol{\beta}_0$. We first estimate the conditional densities $f_{X_k|X^*}(X_k|X^*)$ for $k = 1, 2, \cdots, K$ and the latent variable density $f_{X^*}(X^*)$ using a sieve maximum likelihood estimator. Using the estimated sieve ML densities, we construct an estimator for the conditional density $f_{X^*|\boldsymbol{X}}(X^*|\boldsymbol{X})$. In the second step, we plug the estimates of $f_{X^*|\boldsymbol{X}}(X^*|\boldsymbol{X})$ into the objective function in Equation (2.2) and estimate $\boldsymbol{\beta}_0$ using a maximum rank correlation estimator.

We focus on describing the estimation procedure for the additive model (see Assumption 2.6) for the following reasons. First, the primary goal of our estimation is to recover the optimal linear rank index parameter $\boldsymbol{\beta}_0$, so the conditional density functions $f_{X_k|X^*}(X_k|X^*)$ are considered as nuisance parameters. With nonparametric functions g_k and a flexible distribution of the error terms, the additive model provides a good approximation for the relationship between the latent variable and its measurements. Second, with the additive structure, the two-dimensional conditional density function $f_{X_k|X^*}(X_k|X^*)$ can be replaced by $f_{\varepsilon_k}(X_k - g_k(X^*))$, which represents the one-dimensional density function of the error term evaluated at $X_k - g_k(X^*)$. This greatly simplifies the first-step sieve ML estimator, making our estimation procedure more computationally tractable and appealing to applied researchers.

Under the additive model, the likelihood function takes the following form

$$f_{\mathbf{X}}(X_1, X_2, \cdots, X_K) = \int_{\mathscr{X}^*} \prod_{k=1}^{K} f_{\varepsilon_k}(X_k - g_k(x^*)) f_{X^*}(x^*) dx^*,$$

where $g_1(X^*) = X^*$ and $(f_{\varepsilon_1}, \cdots, f_{\varepsilon_K}, f_{X^*}, g_2, \cdots, g_K)$ are identified under the conditions in Theorem 2.2. Our sieve maximum likelihood estimation relies on regularity restrictions on the function space that contains the true parameters of interest

$$\alpha_0 = (\sqrt{f^0_{\varepsilon_1}}, \cdots, \sqrt{f^0_{\varepsilon_K}}, \sqrt{f^0_{X^*}}, g^0_2, \cdots, g^0_K).$$

Under the identification assumptions, we need the unknown functions of interest to be smooth enough so that they can be well approximated by truncated sieve series, such as polynomials. We provide the technical details on constructing function spaces and sieve spaces that satisfy the density and monotonicity restrictions in Appendix B.1. The step-by-step estimation procedure is described below. To simplify notation, we use f_k to denote the density function of ε_k for $k = 1, 2, \cdots, K$ and f_{K+1} to denote the density function of X^* hereafter.

Step 1: Construct a sieve MLE of α_0:

$$\widehat{\alpha}_n \equiv \left(\widehat{\sqrt{f_1}}, \cdots, \widehat{\sqrt{f_K}}, \widehat{\sqrt{f_{K+1}}}, \widehat{g}_2, \cdots, \widehat{g}_K\right) = \arg\max_{\alpha \in \mathscr{A}_n} \widehat{\mathscr{L}_n}(\alpha), \tag{3.1}$$

where $\mathscr{A}_n$ is a finite dimensional sieve space that becomes dense in the function space covering the true unknown functions of interest. $\widehat{\mathscr{L}_n}(\alpha)$ is the sample analog of the log likelihood function using an i.i.d. sample $\left\{x_{1i}, \cdots, x_{Ki}\right\}_{i=1}^{n}$

and is given by

$$\widehat{\mathscr{L}}_n(\alpha) = \frac{1}{n}\sum_{i=1}^{n} \ln \int_{\mathscr{X}^*} \prod_{k=1}^{K} f_k(x_{ki} - g_k(x^*)) f_{K+1}(x^*) dx^*. \tag{3.2}$$

Step 2: With the estimated densities $\hat{f}_k = \left(\widehat{\sqrt{f_k}}\right)^2$ for $k = 1, \cdots, K, K+1$ and $\hat{g}_k$ for $k = 2, 3, \cdots, K$ (note that $\hat{g}_1(X^*) = X^*$), we construct:

$$\widehat{f}_{X^*|\boldsymbol{X}}(X^*|\boldsymbol{X}) = \frac{\prod_{k=1}^{K} \widehat{f}_k(X_k - \widehat{g}_k(X^*))\widehat{f}_{K+1}(X^*)}{\int_{\mathscr{X}^*} \prod_{k=1}^{K} \widehat{f}_k(X_k - \widehat{g}_k(x^*))\widehat{f}_{K+1}(x^*) dx^*}. \tag{3.3}$$

Step 3: Construct the sample analogue of the objective function $Q_0(\boldsymbol{\beta})$ in Equation (2.2) based on the analogy principle. Let $\boldsymbol{x_i} = (x_{1i}, x_{2i}, \cdots, x_{Ki})$.

$$\widehat{Q}_n(\boldsymbol{\beta}; \widehat{f}_{X^*|\boldsymbol{X}}) = \frac{\sum\limits_{i \neq j} \int \mathbf{1}(\boldsymbol{x_i\beta} > \boldsymbol{x_j\beta}) \mathbf{1}(x_i^* > x_j^*) \widehat{f}_{X^*|\boldsymbol{X}}(x_i^*|\boldsymbol{x_i}) \widehat{f}_{X^*|\boldsymbol{X}}(x_j^*|\boldsymbol{x_j}) dx_i^* dx_j^*}{n(n-1)}. \tag{3.4}$$

Step 4: The estimator of the optimal linear rank index parameter $\boldsymbol{\beta}_0$ is defined as

$$\widehat{\boldsymbol{\beta}}_n \equiv \arg\max_{\boldsymbol{\beta}} \widehat{Q}_n(\boldsymbol{\beta}; \widehat{f}_{X^*|\boldsymbol{X}}). \tag{3.5}$$

A few remarks about our estimation procedure are in order. First, the sieve MLE in Step 1 can be further simplified by imposing parametric assumptions on the g functions when appropriate. For example, consider the case where $g_k(X^*) = \gamma_k X^*$ for $k = 1, \cdots, K$. Under Assumptions 2.1 and 2.6, for any two different measurements X_k and X_l with $k \neq l$, we have

$$Cov(X_k, X_l) = \gamma_k \gamma_l Var(X^*). \tag{3.6}$$

By normalizing $\gamma_1 = 1$ (so that $g_1(X^*) = X^*$ as before), we can easily construct consistent estimators for γ's given an i.i.d. sample of measurements $\left\{x_{1i}, \cdots, x_{Ki}\right\}_{i=1}^{n}$. Let $\mu_k \equiv \frac{1}{n}\sum\limits_{i=1}^{n} x_{ki}$

be the sample mean of x_{ki} for $k = 1, 2, \cdots, K$.

$$\widehat{\gamma}_k = \frac{\sum_{i=1}^{n} \left(x_{ki} - \mu_k\right)\left(x_{li} - \mu_l\right)}{\sum_{i=1}^{n} \left(x_{1i} - \mu_1\right)\left(x_{li} - \mu_l\right)}, \text{ for } k = 2, \cdots, K, \text{ with } l \neq k. \tag{3.7}$$

Plugging the estimators $\widehat{\gamma}_k$ into the sample log-likelihood function in (3.2) significantly reduces the computational burden in the Step 1 sieve MLE.

Second, it is possible to estimate $(\boldsymbol{\beta}_0, \alpha_0)$ jointly using a sieve ML estimator. Instead of plugging the Step 2 estimated conditional density $\widehat{f}_{X^*|\boldsymbol{X}}(X^*|\boldsymbol{X})$ into the sample analogue of the objective function $Q_0(\boldsymbol{\beta})$, we can construct it directly as a function of $(\boldsymbol{\beta}, \alpha)$.

$$\widehat{Q}_n(\boldsymbol{\beta}, \alpha) = \frac{1}{n(n-1)} \sum_{i \neq j} \int \mathbf{1}(\boldsymbol{x_i\beta} > \boldsymbol{x_j\beta}) \mathbf{1}(x_i^* > x_j^*) f_{X^*|\boldsymbol{X}}(x_i^*|\boldsymbol{x_i}; \alpha) f_{X^*|\boldsymbol{X}}(x_j^*|\boldsymbol{x_j}; \alpha) dx_i^* dx_j^*.$$

The key difference is that $f_{X^*|\boldsymbol{X}}(x^*|\boldsymbol{X}; \alpha)$ is now a function of an infinite dimensional unknown parameter α. This sample objective function $\widehat{Q}_n(\boldsymbol{\beta}, \alpha)$ depends on the unknown parameter $(\boldsymbol{\beta}, \alpha)$ nonlinearly and pointwise nonsmoothly. This type of sieve estimators has been considered in Chen and Pouzo (2012) and Chen and Pouzo (2009). The desired theoretical properties of these sieve estimators are generally much more challenging to obtain. Chen and Liao (2014), Chen and Pouzo (2015), and Hahn and Liao (2018) provide examples of ill-posed inverse problems in which the finite dimensional functionals fail to be root-n estimable.

3.1. Asymptotic Properties

We now briefly discuss the asymptotic properties of the proposed optimal linear rank index estimator in Equation (3.5). The technical assumptions and main theorems are presented in Appendices B.2–B.5. First, the estimator $\widehat{\alpha}_n$ is a direct application of the general sieve MLE and it is consistent following Shen (1997), Ai and Chen (2003), Hu and Schennach (2008), and Chen and Pouzo (2012). Theorem B.1 provides the convergence rate of the sieve MLE under additional assumptions following Ai and Chen (2003). Under the assumption that the likelihood function is continuous in a neighborhood of α_0, Corollary B.1 shows that the estimator $\widehat{f}_{X^*|\boldsymbol{X}}(X^*|\boldsymbol{X})$ in Equation (3.3) is consistent under a sup norm.

We then use the consistency of $\widehat{\alpha}_n$ as the foundation for achieving the consistency of

$\widehat{\boldsymbol{\beta}}_n$. To control the ill-posed inverse problem in the first step, we follow the treatment in Ai and Chen (2003) and Hu and Schennach (2008) to impose restrictions on the conditional density $\widehat{f}_{X^*|\boldsymbol{X}}$ and the relation between the conditional density and its sieve approximation. Note that $\boldsymbol{\beta}_0$ is assumed to be in the interior of a parameter space $\Theta = [0,1]^K$. Given the definition of the objective function, $Q_0(\boldsymbol{\beta})$ is continuous at $\boldsymbol{\beta}$, and therefore the consistency of $\widehat{\boldsymbol{\beta}}_n$ follows.

With the consistency of $\widehat{\boldsymbol{\beta}}_n$ in Theorem B.2, we consider estimators close to the population parameter $\boldsymbol{\beta}_0$, and follow the general method in Sherman (1993) to establish that the proposed optimal linear rank index estimator $\widehat{\boldsymbol{\beta}}_n$ is $\sqrt{n}$−consistent for $\boldsymbol{\beta}_0$ and asymptotically normally distributed. The key challenge we face is that the estimated conditional density $\widehat{f}_{X^*|\boldsymbol{X}}(X^*|\boldsymbol{X})$ enters the sample objective function. We describe our approach to decompose the sample objective function and represent it as a quadratic approximation in Section B.5. The asymptotic normality result is stated in Theorem B.3. To make inference, we suggest a bootstrap procedure for $\widehat{\boldsymbol{\beta}}_n$. More detailed discussions on the consistency of the bootstrap procedure can be found in Chen, Linton, and Van Keilegom (2003).

4. Monte Carlo Simulation

In this section, we use a Monte Carlo method to investigate the finite sample properties of the proposed optimal linear rank index estimator. We consider two scenarios within the additive model (see Assumption 2.6). In the first case, $g_k(\cdot)$ is a linear function and we estimate the distribution of ε_k nonparametrically. We consider an alternative scenario where $g_k(\cdot)$ is a nonlinear function with normally distributed ε_k. In this case, we estimate $g_k(\cdot)$ nonparametrically. For expository purposes, this section presents a simple model with $K = 3$.

4.1. Models with Linear g_k

We generate the random variable X^* using a normal distribution $N(0,1)$. To construct the contaminated variables X_1, X_2 and X_3, we set $X_k = \gamma_k X^* + \varepsilon_k$ for $k = 1,2,3$. We consider five specifications for the vector of coefficients $(\gamma_1, \gamma_2, \gamma_3)$ and the corresponding measurement errors $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$. The details are presented in Table 1.

For DGPs 1–5, we estimate the linear coefficients in g functions and estimate the distribution of ε_k and the latent variable X^* nonparametrically. We use Hermite orthogonal

Table 1: DGPs for the model with linear g_k

	DGP 1	DGP 2	DGP 3	DGP 4	DGP 5
$(\gamma_1,\gamma_2,\gamma_3)$	$(1,1,1)$	$(1,0.6,0.8)$	$(1,1,1)$	$(1,0.8,1)$	$(1,1,1)$
ε_1	$N(0,1)$	$N(0,0.2^2)$	$N(0,0.5^2)$	$N(0,1)$	$N(0,0.5^2)$
ε_2	$N(0,0.8^2)$	$\frac{1}{2}S(1.5,0,1,0)$	e_1+e_2, with $e_1 \sim N(0,1)$ $e_2 \sim U(0,5)$	$0.8N(-5,1)$ $+0.2N(5,5^2)$	$N(0,0.5^2)$
ε_3	$N(0,2^2)$	$N(0,1)$	$N(0,0.5^2)$	$N(0,0.5^2)$	$N(0,2^2)$

Note: In DGP II, $S(\grave{\alpha},\grave{\beta},c,\mu)$ represents a skewed centered stable distribution with a stability parameter $\grave{\alpha}$, a skewness parameter $\grave{\beta}$, a scale parameter c, and a location parameter μ. In DGP IV, $0.8N(-5,1)+0.2N(5,5^2)$ represents a bimodally distributed random variable that follows $N(-5,1)$ with probability 0.8 and $N(5,5^2)$ with probability 0.2.

series as the sieve basis functions for $\sqrt{f_1(\varepsilon_1)}$, $\sqrt{f_2(\varepsilon_2)}$, $\sqrt{f_3(\varepsilon_3)}$, and $\sqrt{f_4(x^*)}$.[6] For comparison, we also consider three alternative estimators. The first is the infeasible estimator of $\boldsymbol{\beta}_0$, which assumes that the joint distribution of the latent variable and its measurements is known to the econometrician. Therefore, by maximizing the rank correlation between $\boldsymbol{X\beta}$ and X^*, we obtain:

$$\widehat{\boldsymbol{\beta}}_{\text{inf.}} = \arg\max_{\boldsymbol{\beta}} \frac{1}{n(n-1)} \sum_{i \neq j} \mathbf{1}(\boldsymbol{X}_i\boldsymbol{\beta} > \boldsymbol{X}_j\boldsymbol{\beta})\mathbf{1}(X_i^* > X_j^*). \tag{4.1}$$

We also compare the performance of our estimator to the principal component analysis (PCA) estimator and the independent component analysis (ICA) estimator. Both PCA and ICA are widely used dimensionality reduction algorithms. PCA transforms a set of correlated variables into a smaller number of uncorrelated variables, retaining as much variation from the original dataset as possible. This approach is useful for identifying latent variables (factors) underlying the observed data. ICA extends the PCA technique. It is designed to recover a small number of independent linear components from a large number of noisy linear combinations.[7] Once the latent variables are recovered, it is possible to rank the values.

[6] We choose Hermite orthogonal series as the sieve basis functions mainly because it is easier to impose monotonicity conditions on the regression functions. The properties of the Hermite basis can be found in Walter (1977) and Gallant and Nychka (1987). The dimensions of sieve spaces are specified as $(J_{1,n}, J'_{1,n}) = (4,0)$, $(J_{2,n}, J'_{2,n}) = (4,1)$, $(J_{3,n}, J'_{3,n}) = (4,1)$, and $J_{4,n} = 4$. So $k_n = 18$ for linear g_k models.

[7] The ICA algorithm assumes that each sample of data is a mixture of independent components and it extracts these components by maximizing the non-Gaussianity of these random variables. More details of the

For all estimators, we consider a sample size of N=500 with each estimation involving 1000 simulation replications. In Table 3, we report the mean, median, bias, and standard deviation of the four estimators. We find that our optimal linear rank estimator performs well in finite samples. For all DGPs, only the infeasible estimator and the proposed linear rank index estimator are close to the population parameter of interest. The bias of our estimator is slightly larger than that of the infeasible estimator, as we would expect. As we increase the sample size, the standard deviation of the optimal linear rank index estimator decreases in all five simulation designs.[8]

Given the simulation data $\{\boldsymbol{x_i}, x_i^*\}_{i=1}^n$ and an estimator $\widehat{\boldsymbol{\beta}}$, we can assess how well the one dimensional representation of the measurements $\boldsymbol{x_i}\widehat{\boldsymbol{\beta}}$ preserves the ranking of the latent variable x_i^*. For each estimator, we report in Tables 4 the probability that the order of the latent variable is correctly predicted for two random draws, i and j, i.e.,

$$\frac{2}{n(n-1)} \sum_{i \neq j} \Big(\mathbf{1}(\boldsymbol{x_i}\widehat{\boldsymbol{\beta}} > \boldsymbol{x_j}\widehat{\boldsymbol{\beta}}) = \mathbf{1}(x_i^* > x_j^*) \Big). \tag{4.2}$$

In all simulation designs, the ranking of the latent variable across observations is largely preserved given ordering labels assigned by our proposed estimator, with approximately 85% of the ranking orders of the latent variable correctly predicted. As expected, the percentage of correct predictions using ICA and PCA estimators is much lower in all simulations. Although ICA and PCA estimators are often used empirically to construct indexes for comparing quality or performance across observations, they are not theoretically designed to rank the values of the latent variable.

4.2. Models with Non-linear g_k

We now consider a case where g_k is nonlinear and nonparametrically estimated, but we impose a normal parametric assumption on the distribution of ε_k and the latent variable X^*. To simulate the data, we use three types of nonlinear functions:

$$h_1(x; a_1, a_2) = a_1 x + a_2(x + x^2 + \frac{x^3}{3}),$$

ICA estimator can be found in Hyvärinen and Oja (1999), Hyvärinen and Oja (2000), and Chen and Bickel (2006). We greatly appreciate Professor Brian Moore for providing the codes for the Fast ICA algorithm.

[8] The estimation results for $N = 1000$ are reported in Online Appendix D.

$$h_2(x;a_1,a_2) = a_1 x + a_2 e^x,$$
$$h_3(x;a_1,a_2) = \frac{a_1 x^3}{3} - a_2 e^{-x},$$

where a_1 and a_2 are two parameters entering these nonlinear functions. Note that h_1, h_2, and h_3 are strictly monotone for positive values of a_1 and a_2. We consider four specifications for the g_k functions and the distribution of the errors $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$. The details are provided in Table 2.

Table 2: DGPs for the model with nonlinear g_k

	DGP 6	DGP 7	DGP 8	DGP 9
$g_1(X^*)$	X^*	X^*	X^*	X^*
$g_2(X^*)$	$h_2(X^*;0.2,0.2)$	$h_1(X^*;2,1)$	$h_1(X^*;1,0.2)$	$h_2(X^*;0.2,1)$
$g_3(X^*)$	$h_3(X^*;0.2,0.5)$	$h_2(X^*;1.2,0.2)$	$h_3(X^*;0.2,0.2)$	$h_2(X^*;0.2,0.5)$
ε_1	$N(0,1)$	$N(0,1)$	$N(0,1)$	$N(0,1)$
ε_2	$N(0,0.5^2)$	$N(0,0.8^2)$	$N(0,0.5^2)$	$N(0,0.5^2)$
ε_3	$N(0,2^2)$	$N(0,0.2^2)$	$N(0,2^2)$	$N(0,1.5^2)$

For DGPs 6–9, we estimate the nonlinear g_k functions nonparametrically. Specifically we use the polynomial series as the sieve approximations for $g_2(\cdot)$, and $g_3(\cdot)$.[9] The distributions of the errors and the latent variable are estimated parametrically. Again we compare our estimator with the infeasible estimator, the PCA and ICA estimators. The simulation results for a sample size of $N = 500$, based on 1000 replications, are presented in Table 5. We find that for models with nonlinear g_k functions, our proposed estimator performs well. The bias of our estimator is generally small and comparable to that of the infeasible estimator. Our estimator also achieves a much higher probability of correctly predicting the ranking of the latent variable compared to the PCA and ICA estimators (see Table 6).

5. Empirical Example

Timely and accurate information on the state of the economy is essential for effective macroeconomic policymaking. However, obtaining such information has been a long-standing challenge for many countries. While official GDP growth usually provides good guidance, its

[9] For nonlinear g_k models, the dimensions of sieve spaces are specified as $(J_{1,n}, J'_{1,n}) = (2,0)$, $(J_{2,n}, J'_{2,n}) = (2,4)$, $(J_{3,n}, J'_{3,n}) = (2,4)$, and $J_{4,n} = 2$. As a result, $k_n = 16$.

quality varies across countries. Additionally, the presence of the informal economy complicates the understanding of overall economic activity. Recently, the rapid rise of non-traditional data sources has offered alternative measurements of economic activity, potentially providing new insights into the state of the economy. In this section, we apply the proposed linear rank index estimator to assess GDP growth across countries using multiple measurements.

We consider three measurements of economic activity in addition to official GDP growth: satellite-recorded luminosity, carbon dioxide emissions, and the Google search volume index. Satellite-recorded luminosity, or nighttime lights, measures the brightness of a country at night and has been widely used in the literature as a proxy for economic activity.[10] Carbon dioxide emissions, a byproduct of burning fossil fuels, gauge the energy input of economic production. The relationship between carbon dioxide emissions and economic growth has long been recognized in the context of the environmental Kuznets curve.[11] Google Trends provides search volume indexes (SVIs) that measure search intensity of keywords and topics by location and over time. A growing body of literature has demonstrated that SVIs contain valuable information about the state of the economy.[12]

The identifying assumptions for the optimal linear rank index are likely to hold for these measurements. Satellite-recorded luminosity, carbon dioxide emissions, and the Google search volume index each reflect true economic activity from distinct perspectives, making them likely to be independent once conditioning on the true state of the economy.[13] Assumption 2.2 requires that all joint and marginal probability density functions of these measurements and true GDP growth are bounded on their supports. Regarding the monotonicity assumption, since official GDP growth and carbon dioxide emissions are informative

[10] Chen and Nordhaus (2011), Henderson, Storeygard, and Weil (2012) pioneered the use of luminosity as measures of economic activity. More recently, Hu and Yao (2021) estimate the relationship between luminosity and GDP in a measurement error model framework. Beyer, Franco-Bedoya, and Galdo (2021) examines the impact of COVID-19 in India using luminosity data. Gibson, Olivia, and Boe-Gibson (2020) provide an overview of sources and uses of luminosity data in economics.

[11] See, for example, Nordhaus (1977), Narayan and Narayan (2010), Fei, Dong, Xue, Liang, and Yang (2011).

[12] Since the seminal work of Choi and Varian (2012), there has been a burgeoning body of research using Google Trends data to nowcast GDP (e.g., Carrière-Swallow and Labbé (2013), Narita and Yin (2018), Woloszko (2020)) and understand unemployment (e.g., Baker and Fradkin (2017)).

[13] For example, measurement errors in official GDP data may not be relevant to the amount of nighttime light recorded by satellites or the search volume registered by Google. On the contrary, the conditional independence assumption may not be satisfied if traditional measures of economic activity such as manufacturing index are used. This is because manufacturing is an important active component of GDP and thus measurement errors in manufacturing and GDP can be correlated.

about true GDP growth, Assumption 2.8 is likely to be satisfied. The large range conditions in Assumption 2.7(i) are likely to hold given that we are using continuous measurements; Assumption 2.7(ii) is a technical assumption, but it is empirically testable since it involves only observable information.

We construct the measurements of economic activity following the existing literature. In particular, we follow Hu and Yao (2021) to construct luminosity growth by calculating the year-on-year growth of the sum of nighttime lights within country boundaries. We follow Narita and Yin (2018) to construct Google SVI growth using the names of countries as the search topic. For all of these proxy measurements, we compute the growth rates using log differences. We focus on three samples in this application. The first two consider annual data for a wide range of countries (Sample 1 covers the years 2000–2011, and Sample 2 covers the years 2014–2018), while Sample 3 considers quarterly data and focuses on emerging markets and developing economies. In this application, we assume that the relationship between the latent variable and its measurements remains stable over time and across countries.[14] The details of the data sources for each measurement in all samples are provided in Table D.5 in Online Appendix.

Table 7 shows the estimated coefficients for ICA, PCA, and the optimal linear rank index estimator with linear and nonlinear g_k functions, respectively.[15] The coefficients describe how the ranking order of the latent variable, which in this example is economic growth, can be explained by its measurements. Interestingly, the results with nonlinear g_k indicate that luminosity has a larger coefficient than official GDP growth, implying that luminosity growth is informative about underlying economic growth when the nonlinearity of the relationship between true economic growth and its measurements is taken into account. In contrast to Gibson, Olivia, Boe-Gibson, and Li (2021), who demonstrate limited predictive power of luminosity data for economic activity in low-density rural areas, our results suggest that country-level luminosity data can be at least as useful as official data.

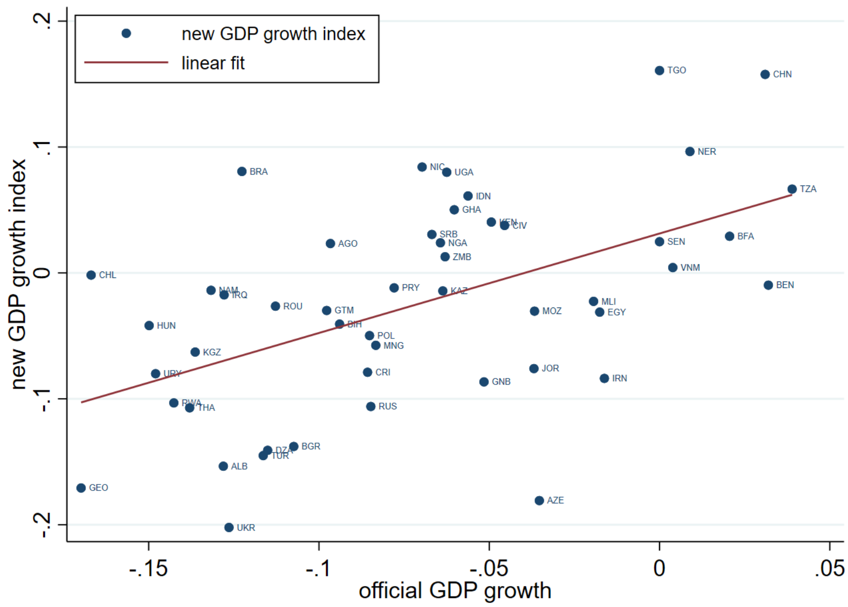
Based on our estimates of the optimal linear rank index (with nonlinear g functions), we create a *GDP Growth Index*.[16] In Figure 1, we plot the relationship between our index and

[14]It is possible to empirically test this assumption by estimating the linear rank index model using subsets of time periods or countries.

[15]In our empirical application, the dimensions of the sieve space are specified as $(J_{1,n}, J'_{1,n}) = (2,0)$, $(J_{2,n}, J'_{2,n}) = (2,4)$, $(J_{3,n}, J'_{3,n}) = (2,4)$, and $J_{4,n} = 2$. So $k_n = 16$.

[16]One caveat is that *GDP growth index* may not be directly comparable to official GDP growth rate. This is because our proposed index is a linear combination of multiple measurements, but different measurements of

Figure 1: GDP Growth Index vs. Official GDP Growth Rate: 2020Q2



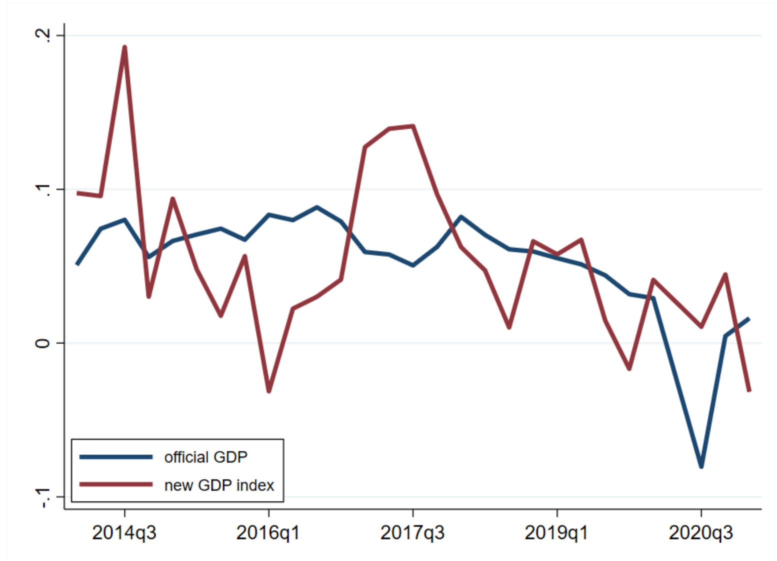
the true GDP growth rate in the second quarter of 2020 (the peak of the pandemic period) across different countries. This figure suggests that while there is a positive relationship between our GDP growth index and the official GDP growth rate, with a correlation of 0.525, they are clearly different. Our GDP growth index leverages various new measurements of economic activity, potentially containing information not captured in official GDP growth data. From this figure, we observe that during the pandemic period, for some western African countries, such as Senegal, Burkina Faso, and Benin, the official data might have underestimated the detrimental effects of the pandemic. Conversely, for Brazil and Chile, their economy might have weathered the shocks better than the official data suggest.

Our GDP growth index may also offer important insights into patterns of economic growth over time. In Figure 2, we plot the dynamic patterns of the official GDP growth rate and our proposed index for India the years 2014 to 2020. Our GDP growth index suggests that India may have experienced higher economic activity in 2017 and 2020 than what was reported in the official data. Higher-than-official GDP growth in 2017 could be related to the effects of India's demonetization in late 2016, which official GDP growth was not able to capture.[17] Similarly, during the pandemic, economic activity might have shifted to the

economic activity may have different elasticities with respect to the true GDP growth.

[17] In a related study, Chanda and Cook (2022) use nighttime light data to show that demonetization in India had a positive effect on India's poorest districts in 2017.

Figure 2: New GDP Growth Index vs. Official GDP Growth for India: 2014–2020



informal sector, which is not recorded in official data.[18]

In sum, the linear rank index offers a new perspective on the state of the economy. Notably, where it differs from official data, it raises questions about economic activity that may be overlooked by official sources and provides new insights.

6. Conclusion

This paper develops an econometric method to rank the values of a latent variable using a set of imperfect measurements. We provide a statistical framework to describe the ordering relationship between the latent variable and a linear combination of observable measurements. We first leverage of the variation of at least three observed measurements and provide sufficient conditions to identify the joint distribution of the latent variable and its measurements. We then construct an estimator to maximize the rank correlation between the latent variable and the linear combination of observables. The asymptotic properties of the sieve-based linear rank index estimator are discussed. Our flexible model allows us to reduce the dimensions of the observables and provides a simple but informative way to compare the values of the latent variable across observations in the sample.

Applying our novel method to assess true GDP growth across counties, we show that

[18]In a recent paper, Schneider (2022) shows that the size of the shadow economy increased from 2019 to 2020 for 27 European Union countries and the United Kingdom.

alternative measurements, such as luminosity and Google search volume indexes, can be important complements to official GDP statistics. Our proposed method can also be easily applied to other empirical settings, where we aim to rank cognitive skills among children, the creditworthiness of borrowers, and the quality or performance of schools and hospitals.

In this paper, we focus on the scenario where the number of the measurements K is finite. When high-dimensional measurements are available, our framework could potentially be extended to estimate the linear rank index by combining a penalized sieve MLE with a penalized maximum correlation method (see Dong, Gao, and Linton, 2023; Cheng, Dong, Gao, and Linton, 2022; Lin and Peng, 2013; Han, Ji, Ji, and Wang, 2015; Dai, Zhang, and Sun, 2014).[19] Our paper focuses on a scalar latent variable. In some applications, the latent variable may be multi-dimensional, such as cognitive and non-cognitive skills, both of which play an important role in education and labor market outcomes. Our framework may be extended to rank multi-dimensional latent variables based on observable measurements. We leave a thorough investigation of these directions for future research.

[19] Relatedly, Feng (2021) develops a general causal inference method for treatment effect models with a large set of noisy measurements linked with the underlying latent confounders. The author combines K-nearest neighbors matching and local principal component analysis to extract information from the latent confounders, and then use the information to construct estimators of causal parameters of interest based on doubly-robust score functions.

References

AGASISTI, T., AND C. PÉREZ-ESPARRELLS (2010): "Comparing efficiency in a cross-country perspective: the case of Italian and Spanish state universities," *Higher Education*, 59, 85–103.

AI, C., AND X. CHEN (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71(6), 1795–1843.

BACK, A. D., AND A. S. WEIGEND (1997): "A first application of independent component analysis to extracting structure from stock returns," *International journal of neural systems*, 8(04), 473–484.

BAKER, S. R., AND A. FRADKIN (2017): "The Impact of Unemployment Insurance on Job Search: Evidence from Google Search Data," *Review of Economics and Statistics*, 99(5), 756–768.

BEYER, R. C., S. FRANCO-BEDOYA, AND V. GALDO (2021): "Examining the Economic Impact of COVID-19 in India through Daily Electricity Consumption and Nighttime Light Intensity," *World Development*, 140, 105287.

CARRIÈRE-SWALLOW, Y., AND F. LABBÉ (2013): "Nowcasting with Google Trends in an Emerging Market," *Journal of Forecasting*, 32(4), 289–298.

CHANDA, A., AND C. J. COOK (2022): "Was India Demonetization Redistributive? Insights from Satellites and Surveys," *Journal of Macroeconomics*, 73, 103438.

CHEN, A., AND P. J. BICKEL (2006): "Efficient Independent Component Analysis," *The Annals of Statistics*, 34(6), 2825–2855.

CHEN, S. (2002): "Rank Estimation of Transformation Models," *Econometrica*, 70(4), 1683–1697.

CHEN, X., Y. HU, AND A. LEWBEL (2009): "Nonparametric Identification and Estimation of Nonclassical Errors-In-Variables Models without Additional Information," *Statistica Sinica*, 19, 949–968.

CHEN, X., AND Z. LIAO (2014): “Sieve M Inference on Irregular Parameters,” *Journal of Econometrics*, 182(1), 70–86.

CHEN, X., O. LINTON, AND I. VAN KEILEGOM (2003): “Estimation of Semiparametric Models When the Criterion Function is not Smooth,” *Econometrica*, 71(5), 1591–1608.

CHEN, X., AND W. D. NORDHAUS (2011): “Using Luminosity Data as a Proxy for Economic Statistics,” *Proceedings of the National Academy of Sciences*, 108(21), 8589–8594.

CHEN, X., AND D. POUZO (2009): “Efficient Estimation of Semiparametric Conditional Moment Models with Possibly Nonsmooth Residuals,” *Journal of Econometrics*, 152(1), 46–60.

——— (2012): “Estimation of Nonparametric Conditional Moment Models with Possibly Nonsmooth Generalized Residuals,” *Econometrica*, 80(1), 277–321.

——— (2015): “Sieve Wald and QLR inferences on Semi/nonparametric Conditional Moment Models,” *Econometrica*, 83(3), 1013–1079.

CHENG, T., C. DONG, J. GAO, AND O. B. LINTON (2022): “GMM Estimation for High–Dimensional Panel Data Models,” *Available at SSRN 4107164*.

CHOI, H., AND H. VARIAN (2012): “Predicting the Present with Google Trends,” *Economic record*, 88, 2–9.

DAI, P.-J., Q.-Z. ZHANG, AND Z.-H. SUN (2014): “Variable Selection of Generalized Regression Models Based on Maximum Rank Correlation,” *Acta Mathematicae Applicatae Sinica, English Series*, 30(3), 833–844.

DONG, C., J. GAO, AND O. LINTON (2023): “High Dimensional Semiparametric Moment Restriction Models,” *Journal of Econometrics*, 232, 320–345.

FEI, L., S. DONG, L. XUE, Q. LIANG, AND W. YANG (2011): “Energy Consumption-economic Growth Relationship and Carbon Dioxide Emissions in China,” *Energy Policy*, 39(2), 568–574.

FENG, Y. (2021): “Causal Inference in Possibly Nonlinear Factor Models,” *arXiv preprint arXiv:2008.13651v3*.

FILMER, D., AND L. PRITCHETT (2001): "Estimating Wealth Effects Without Expenditure Data–Or Tears: An Application to Educational Enrollments in States of India," *Demography*, 38(1), 115–132.

FRIESEN, C. E., P. SELISKE, AND A. PAPADOPOULOS (2016): "Using Principal Component Analysis to Identify Priority Neighbourhoods for Health Services Delivery by Ranking Socioeconomic Status," *Online Journal of Public Health Informatics*, 8(2).

GALLANT, A. R., AND D. W. NYCHKA (1987): "Semi-nonparametric Maximum Likelihood Estimation," *Econometrica*, pp. 363–390.

GIBSON, J., S. OLIVIA, AND G. BOE-GIBSON (2020): "Night Lights in Economics: Sources and Uses," *Journal of Economic Surveys*, 34(5), 955–980.

GIBSON, J., S. OLIVIA, G. BOE-GIBSON, AND C. LI (2021): "Which Night Lights Data should we Use in Economics, and Where?," *Journal of Development Economics*, 149, 102602.

HAHN, J., AND Z. LIAO (2018): "Nonparametric Instrumental Variables And Regular Estimation," *Econometric Theory*, 34(3), 574–597.

HAN, A. K. (1987): "Non-Parametric Analysis of a Generalized Regression Model: The Maximum Rank Correlation Estimator," *Journal of Econometrics*, 35(2-3), 303–316.

HAN, F., H. JI, Z. JI, AND H. WANG (2015): "A Provable Smoothing Approach for High Dimensional Generalized Regression with Applications in Genomics," *arXiv preprint arXiv:1509.07158*.

HAO, R., S. LI, J. LI, Q. ZHANG, AND F. LIU (2013): "Water Quality Assessment for Wastewater Reclamation Using Principal Component Analysis.," *Journal of Environmental Informatics*, 21(1).

HENDERSON, J. V., A. STOREYGARD, AND D. N. WEIL (2012): "Measuring Economic Growth from Outer Space," *American Economic Review*, 102(2), 994–1028.

HU, Y., AND S. SCHENNACH (2008): "Instrumental Variable Treatment of Nonclassical Measurement Error Models," *Econometrica*, 76(1), 195–216.

HU, Y., S. SCHENNACH, AND J. L. SHIU (2021): "Identification of Nonparametric Monotonic Rregression Models with Continuous Nonclassical Measurement Errors," *Journal of Econometrics, forthcoming*.

HU, Y., AND J.-L. SHIU (2022): "A Simple Test of Completeness in a Class of Nonparametric Specification," *Econometric Reviews*, 41(4), 373–399.

HU, Y., AND J. YAO (2021): "Illuminating Economic Growth," *Journal of Econometrics*.

HYVÄRINEN, A., AND E. OJA (1999): "Independent component analysis: A Tutorial," *Working Paper*.

——— (2000): "Independent Component Analysis: Algorithms and Applications," *Neural Networks*, 13(4-5), 411–430.

JOLLANDS, N., J. LERMIT, AND M. PATTERSON (2004): "Aggregate eco-efficiency indices for New Zealand—a principal components analysis," *Journal of environmental Management*, 73(4), 293–305.

KHATUN, T. (2009): "Measuring Environmental Degradation by Using Principal Component Analysis," *Environment, Development and Sustainability*, 11, 439–457.

KOLENIKOV, S., AND G. ANGELES (2009): "Socioeconomic status measurement with discrete proxy variables: Is principal component analysis a reliable answer?," *Review of Income and Wealth*, 55(1), 128–165.

KRISHNAN, V. (2010): "Constructing an Area-based Socioeconomic Index: A Principal Components Analysis Approach," *Edmonton, Alberta: Early Child Development Mapping Project*.

LI, T. (2002): "Robust and Consistent Estimation of Nonlinear Errors-in-variables Models," *Journal of Econometrics*, 110(1), 1–26.

LI, Y., F. MA, Y. ZHANG, AND Z. XIAO (2019): "Economic policy uncertainty and the Chinese stock market volatility: new evidence," *Applied Economics*, 51(49), 5398–5410.

LIN, H., AND H. PENG (2013): "Smoothed Rank Correlation of the Linear Transformation Regression Model," *Computational Statistics & Data Analysis*, 57(1), 615–630.

MAHAPATRA, S., M. SAHU, R. PATEL, AND B. N. PANDA (2012): "Prediction of water quality using principal component analysis," *Water Quality, Exposure and Health*, 4(2), 93–104.

MUZAMHINDO, S., Y. KONG, AND T. FAMBA (2017): "Principal component analysis as a ranking tool-A case of world universities," *International Journal of Advanced Research*, 5(6), 2114–2135.

NARAYAN, P. K., AND S. NARAYAN (2010): "Carbon Dioxide Emissions and Economic Growth: Panel Data Evidence from Developing Countries," *Energy Policy*, 38(1), 661–666.

NARDO, M., M. SAISANA, A. SALTELLI, S. TARANTOLA, A. HOFFMAN, AND E. GIOVANNINI (2005): "Handbook on constructing composite indicators: methodology and user guide," *OECD Statistics Working Papers, 2005/3, OECD Publishing.*

NARITA, M. F., AND R. YIN (2018): "In Search of Information: Use of Google Trends? Data to Narrow Information Gaps for Low-income Developing Countries," IMF Working Paper No. 18/286.

NEWEY, W., AND D. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," *Handbook of Econometrics*, 4, 2111–2245.

NORDHAUS, W. D. (1977): "Economic Growth and Climate: the Carbon Dioxide Problem," *The American Economic Review*, pp. 341–346.

PAKES, A., AND D. POLLARD (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57(5), 1027–1057.

SCHENNACH, S. (2004): "Nonparametric Estimation in the Presence of Measurement Error," *Econometric Theory*, 20, 1046–1093.

SCHENNACH, S., AND Y. HU (2013): "Nonparametric Identification and Semiparametric Estimation of Classical Measurement Error Models without Side Information," *Journal of the American Statistical Association*, 108(501), 177–186.

SCHNEIDER, F. (2022): "New COVID-related Results for Estimating the Shadow Economy in the Global Economy in 2021 and 2022," *International Economics and Economic Policy*, 19(2), 299–313.

SHEN, X. (1997): "On Methods of Sieves and Penalization," *Annals of Statistics*, 25, 2555–2591.

SHERMAN, R. (1994): "Maximal Inequalities for Degenerate U-processes with Applications to Optimization Estimators," *The Annals of Statistics*, 22(1), 439–459.

SHERMAN, R. P. (1993): "The Limiting Distribution of the Maximum Rank Correlation Estimator," *Econometrica*, 61, 123–137.

SHIN, Y. (2010): "Local Rank Estimation of Transformation Models with Varying Coefficients," *Econometric Theory*, 26, 1807–1819.

TRIPATHI, M., AND S. K. SINGAL (2019): "Use of Principal Component Analysis for Parameter Selection for Development of a Novel Water Quality Index: a Case Study of River Ganga India," *Ecological indicators*, 96, 430–436.

USHAKOV, N. G. (2011): *Selected Topics in Characteristic Functions*. De Gruyter, Berlin.

VYAS, S., AND L. KUMARANAYAKE (2006): "Constructing Socio-economic Status Indices: How to Use Principal Components Analysis," *Health policy and planning*, 21(6), 459–468.

WALTER, G. G. (1977): "Properties of Hermite Series Estimation of Probability Density," *The Annals of Statistics*, pp. 1258–1264.

WOLOSZKO, N. (2020): "Tracking Activity in Real Time with Google Trends," OECD Working Paper No. 1634.

YI, M. (2019): "Application of principal component analysis in teaching evaluation," *Frontiers in Sport Research*, 1(1), 34–42.

Appendix

A. Tables

Table 3: Simulation Results for Models with Linear g_k (n=500)

	Infeasible			ICA			PCA			Linear Rank		
DGP 1	$(\beta_{01},\beta_{02},\beta_{03}) = (0.356, 0.555, 0.089)$											
	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
Mean	0.357	0.555	0.089	0.105	0.120	0.108	0.213	0.197	0.590	0.344	0.581	0.075
Median	0.357	0.556	0.089	0.152	0.215	0.156	0.213	0.198	0.589	0.341	0.579	0.075
Bias	0.001	0.000	0.000	-0.251	-0.435	0.019	-0.143	-0.358	0.501	-0.012	0.026	-0.014
Std. Dev.	0.028	0.031	0.019	0.570	0.570	0.560	0.018	0.015	0.028	0.116	0.125	0.018
DGP 2	$(\beta_{01},\beta_{02},\beta_{03}) = (0.953, 0.017, 0.030)$											
	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
Mean	0.946	0.023	0.031	-0.165	0.539	-0.036	0.155	0.688	0.156	0.894	0.042	0.065
Median	0.948	0.024	0.029	-0.282	0.933	-0.050	0.172	0.659	0.166	0.905	0.036	0.051
Bias	-0.007	0.006	0.001	-1.118	0.522	-0.066	-0.798	0.671	0.126	-0.059	0.025	0.035
Std. Dev.	0.021	0.020	0.015	0.289	0.765	0.112	0.065	0.140	0.076	0.079	0.028	0.073
DGP 3	$(\beta_{01},\beta_{02},\beta_{03}) = (0.481, 0.039, 0.480)$											
	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
Mean	0.484	0.037	0.480	0.069	0.015	0.066	0.218	0.565	0.217	0.491	0.031	0.478
Median	0.483	0.039	0.479	0.072	0.081	0.063	0.218	0.565	0.218	0.490	0.031	0.475
Bias	0.003	-0.002	0.000	-0.412	-0.024	-0.414	-0.263	0.526	-0.263	0.010	-0.008	-0.002
Std. Dev.	0.032	0.018	0.033	0.446	0.772	0.443	0.011	0.020	0.011	0.148	0.006	0.148
DGP 4	$(\beta_{01},\beta_{02},\beta_{03}) = (0.236, 0.003, 0.762)$											
	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
Mean	0.234	0.003	0.763	-0.023	0.602	-0.038	0.046	0.919	0.035	0.310	0.007	0.684
Median	0.237	0.006	0.760	-0.060	0.983	-0.094	0.046	0.919	0.035	0.282	0.007	0.711
Bias	-0.002	0.000	0.001	-0.259	0.599	-0.800	-0.190	0.916	-0.727	0.074	0.004	0.078
Std. Dev.	0.045	0.012	0.048	0.213	0.729	0.246	0.014	0.020	0.009	0.202	0.003	0.203
DGP 5	$(\beta_{01},\beta_{02},\beta_{03}) = (0.485, 0.484, 0.030)$											
	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
Mean	0.485	0.485	0.030	0.113	0.114	0.102	0.187	0.187	0.627	0.496	0.479	0.025
Median	0.485	0.484	0.030	0.164	0.191	0.167	0.187	0.187	0.626	0.495	0.476	0.025
Bias	0.000	0.001	0.000	-0.372	-0.370	0.072	-0.298	-0.297	0.597	0.011	-0.005	0.005
Std. Dev.	0.038	0.039	0.011	0.577	0.558	0.566	0.013	0.012	0.023	0.164	0.163	0.005

Note: The population quantity $\boldsymbol{\beta}_0$ for each dgp is approximated by computing the mean of the infeasible estimator with a sample size $N = 5000$ in 1000 repetitions.

Table 4: Percentage of Rankings Correctly Predicted for Models with Linear g_k (n=500)

	Infeasible	ICA	PCA	Linear Rank
DGP 1	0.832	0.565	0.722	0.826
	(0.008)	(0.210)	(0.016)	(0.010)
DGP 2	0.926	0.628	0.807	0.873
	(0.004)	(0.179)	(0.018)	(0.099)
DGP 3	0.896	0.539	0.748	0.892
	(0.005)	(0.203)	(0.015)	(0.007)
DGP 4	0.847	0.601	0.677	0.806
	(0.009)	(0.125)	(0.016)	(0.089)
DGP 5	0.896	0.573	0.716	0.891
	(0.006)	(0.239)	(0.016)	(0.008)

Note: Standard deviations are reported in parenthesis, which are computed using estimates across 1000 simulations.

Table 5: Simulation Results for Models with Non-linear g_k (n=500)

	Infeasible			ICA			PCA			Linear Rank		
DGP 6	$(\beta_{01},\beta_{02},\beta_{03})=(0.265,0.639,0.096)$											
	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
Mean	0.258	0.646	0.096	-0.062	0.750	0.554	0.153	0.231	0.616	0.224	0.641	0.135
Median	0.257	0.647	0.096	-0.066	0.799	0.590	0.154	0.231	0.615	0.223	0.635	0.134
Bias	-0.007	0.007	0.000	-0.327	0.111	0.458	-0.112	-0.408	0.520	-0.041	0.002	0.039
Std. Dev.	0.031	0.034	0.016	0.088	0.278	0.204	0.012	0.015	0.012	0.062	0.079	0.040
DGP 7	$(\beta_{01},\beta_{02},\beta_{03})=(0.268,0.596,0.137)$											
	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
Mean	0.259	0.603	0.138	-0.034	0.236	0.739	0.149	0.251	0.600	0.222	0.621	0.157
Median	0.258	0.604	0.137	-0.060	0.271	0.944	0.150	0.251	0.598	0.215	0.626	0.151
Bias	-0.007	0.007	0.000	-0.327	0.111	0.458	-0.112	-0.408	0.520	-0.041	0.002	0.039
Std. Dev.	0.032	0.033	0.020	0.194	0.234	0.553	0.013	0.010	0.019	0.053	0.074	0.042
DGP 8	$(\beta_{01},\beta_{02},\beta_{03})=(0.083,0.784,0.133)$											
	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
Mean	0.083	0.785	0.132	0.124	0.550	0.502	0.577	0.213	0.210	0.130	0.692	0.178
Median	0.083	0.786	0.132	0.121	0.963	0.237	0.578	0.212	0.208	0.123	0.695	0.177
Bias	0.000	0.001	-0.001	0.041	-0.234	0.369	0.494	-0.571	0.077	0.047	-0.092	0.045
Std. Dev.	0.013	0.025	0.023	0.090	0.497	0.419	0.030	0.017	0.029	0.045	0.082	0.054
DGP 9	$(\beta_{01},\beta_{02},\beta_{03})=(0.245,0.650,0.105)$											
	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3	β_1	β_2	β_3
Mean	0.243	0.652	0.105	-0.031	0.829	0.428	0.234	0.265	0.501	0.200	0.661	0.139
Median	0.243	0.653	0.105	-0.037	0.883	0.456	0.237	0.261	0.501	0.195	0.657	0.134
Bias	-0.002	0.002	0.000	-0.276	0.179	0.323	-0.011	-0.385	0.396	-0.045	0.011	0.034
Std. Dev.	0.025	0.030	0.019	0.106	0.301	0.164	0.025	0.027	0.016	0.076	0.094	0.053

Note: The population quantity $\boldsymbol{\beta}_0$ for each dgp is approximated by computing the mean of the infeasible estimator with a sample size $N = 5000$ in 1000 repetitions.

Table 6: Percentage of Rankings Correctly Predicted for Models with Non-linear g_k (n=500)

	Infeasible	ICA	PCA	Linear Rank
DGP 6	0.853	0.772	0.777	0.848
	(0.009)	(0.097)	(0.012)	(0.010)
DGP 7	0.875	0.733	0.812	0.872
	(0.007)	(0.167)	(0.010)	(0.007)
DGP 8	0.881	0.814	0.794	0.874
	(0.007)	(0.086)	(0.011)	(0.010)
DGP 9	0.850	0.781	0.785	0.843
	(0.008)	(0.098)	(0.012)	(0.011)

Note: Standard deviations are reported in parenthesis, which are computed using estimates across 1000 simulations.

Table 7: Estimation Results of the Linear Rank Index for GDP Growth Rate

	Official GDP	Luminosity	CO2 emissions	Google SVI
Sample 1 ($N = 1,927$)				
ICA	0.240	0.000	-0.971	
	(0.258)	(0.406)	(0.667)	
PCA	0.013	0.950	0.037	
	(0.004)	(0.017)	(0.015)	
Linear Rank Index with linear g_k	0.683	-0.029	0.346	
	(0.189)	(0.019)	(0.189)	
Linear Rank Index with nonlinear g_k	0.243	0.250	0.508	
	(0.096)	(0.061)	(0.106)	
Sample 2 ($N = 804$)				
ICA	0.981	-0.124	0.148	
	(0.448)	(0.519)	(0.458)	
PCA	0.005	0.980	0.015	
	(0.004)	(0.008)	(0.006)	
Linear Rank Index with linear g_k	0.611	0.007	0.382	
	(0.251)	(4.702)	(4.674)	
Linear Rank Index with nonlinear g_k	0.205	0.317	0.478	
	(0.107)	(0.104)	(0.123)	
Sample 3 ($N = 1865$)				
ICA	0.992	0.002		0.127
	(0.500)	(0.459)		(0.366)
PCA	0.007	0.981		0.012
	(0.003)	(0.014)		(0.013)
Linear Rank Index with linear g_k	0.921	0.010		0.069
	(0.156)	(0.028)		(0.149)
Linear Rank Index with nonlinear g_k	0.321	0.283		0.396
	(0.128)	(0.092)		(0.119)

Note: (1). Linear Rank Index 1, Linear Rank Index 2, and Linear Rank Index 3 are referred to the estimation results using the data on the period between 2000 and 2011, the period between 2014 and 2018 and the period between 2014 and 2021, respectively. (2). Standard errors (calculated by 1000 repetitions) are in parentheses. (3). Symbols ** and *** indicate that the test is significant at a level of 5% and 1% respectively.

B. Estimation: Technical Details

B.1. Regularity Conditions for the Function Spaces

Our sieve maximum likelihood estimation described in Section 3 relies on regularity restrictions on the function spaces containing the true parameters of interest α_0. We first introduce notations for Hölder spaces. For a $d \times 1$ vector of nonnegative integers, $\boldsymbol{a} = (a_1, \cdots, a_d)'$, denote $[\boldsymbol{a}] = a_1 + \cdots + a_d$. Let $D^{\boldsymbol{a}}$ denote the differential operator given by $D^{\boldsymbol{a}} = \frac{\partial^{[\boldsymbol{a}]}}{\partial \xi_1^{a_1} \cdots \partial \xi_d^{a_d}}$. Let m denote the largest integer satisfying $\gamma > m$ and set $\gamma = m + p$. The Hölder space $\Lambda^{\gamma}(\nu)$ of order $\gamma > 0$ is a collection of functions which are m times continuously differentiable on ν and the $m-$th derivative are Hölder continuous with the exponent p. For all $g \in \Lambda^{\gamma}(\nu)$, the Hölder norm is defined as

$$\|g\|_{\Lambda^{\gamma}} = \sup_{\xi \in \nu} |g(\xi)| + \max_{a_1 + \cdots + a_d = m} \sup_{\xi \neq \xi' \in \nu} \frac{|D^{\boldsymbol{a}} g(\xi) - D^{\boldsymbol{a}} g(\xi')|}{\|\xi - \xi'\|_E^p}.$$

The weighted Hölder norm is defined as $\|g\|_{\Lambda^{\gamma,\omega}} \equiv \|\widetilde{g}\|_{\Lambda^{\gamma}}$ with $\widetilde{g}(\xi) \equiv g(\xi)\omega(\xi)$ and the corresponding weighted Hölder space is $\Lambda^{\gamma,\omega}(\nu)$. Define a weighted Hölder ball as $\Lambda_c^{\gamma,\omega}(\nu) \equiv \{g \in \Lambda^{\gamma,\omega}(\nu) : \|g\|_{\Lambda^{\gamma,\omega}} \leq c < \infty\}$. We define the following function spaces that satisfy the density and monotonicity restrictions:

$$\begin{aligned}
\mathscr{F} = \Big\{ \sqrt{f_\varepsilon(\cdot - g(\cdot))} : \sqrt{f_\varepsilon(\cdot)} \in \Lambda_c^{\kappa,\omega}(\mathbb{R}), f_\varepsilon(\cdot) \geq 0 \text{ and } \int_{\mathbb{R}} f_\varepsilon(\varepsilon) d\varepsilon = 1, \\
g(\cdot) \in \Lambda_c^{\kappa,\omega}(\mathbb{R}), g(\cdot) \text{ is strictly monotonic} \Big\}, \\
\mathscr{G} = \Big\{ g(\cdot) \in \Lambda_c^{\kappa,\omega}(\mathbb{R}) : g(\cdot) \text{ is strictly monotonic} \Big\}.
\end{aligned}$$

We impose the following regularity conditions for the square roots of the densities and nonparametric functions in α_0.

Assumption B.1. *Suppose $\kappa > 1$. (i) all the assumptions of Theorem 2.2 hold; (ii) $\sqrt{f^0_{\varepsilon_k}(\cdot)} \in \mathscr{F}$ for $k = 1, \cdots, K$; (iii) $\sqrt{f^0_{X^*}(\cdot)} \in \mathscr{F}$; (iv) $g^0_k(\cdot) \in \mathscr{G}$ for $k = 2, \cdots, K$.*

Assumption B.1 imposes stronger monotonicity conditions on g functions as required in Theorem 2.2. This assumption is invoked primarily to ease the implementation of sieve MLE, and is not required for achieving desired asymptotic properties of our estimator. Assumption B.1 implies that $\alpha_0 \in \mathscr{A} \equiv \mathscr{F}^{K+1} \times \mathscr{G}^{K-1}$. When the function space $\mathscr{A}$ is large, the direct ML estimation method based on the sample analog of the log likelihood function in Equation (3.2) could yield an inconsistent estimator or an estimator which converges very slowly. Thus, we replace $\mathscr{A}$ with a finite dimensional sieve space $\mathscr{A}_n \equiv \mathscr{F}_n^{K+1} \times \mathscr{G}_n^{K-1}$ that becomes dense in $\mathscr{A}$ as the sample size n increases. Let $h^{J_n}(\cdot) = (h_{1_n}(\cdot), \cdots, h_{J_n}(\cdot))^T$ denote a vector of the Hermite orthogonal polynomial basis functions. p_j represents polynomial basis functions satisfying monotonicity. We define the seive spaces satisfying the density and monotonicity conditions as follows.

$$
\begin{aligned}
\mathscr{F}_n = \Big\{ \sqrt{f(\cdot - g(\cdot))} \in \mathscr{F} :\ & \text{there exists } \beta_f \in \mathbb{R}^{J_n} \text{ and } \left(\mu_0, \beta_{g0}, \beta_g\right) \in \mathbb{R}^{J'_n} \text{ such that} \\
& \sqrt{f(\varepsilon)} = h^{J_n}(\varepsilon) \equiv (h_1(\varepsilon), \cdots, h_{J_n}(\varepsilon))\beta_f^T, \\
& g^{J'_n}(x^*) = \mu_0 + \beta_{g0} x^* + \int_a^{x^*} \left(\textstyle\sum_{j=1}^{J'_n-2} \beta_{gj} p_j(x^*)\right)^2 dt \Big\}, \\
\mathscr{G}_n = \Big\{ g(\cdot) \in \mathscr{G} :\ & \text{there exists } \left(\mu_0, \beta_{g0}, \beta_g\right) \in \mathbb{R}^{J'_n} \text{ such that} \\
& g'(\cdot) = \beta_{g0} + \left(\textstyle\sum_{j=1}^{J'_n-2} \beta_{gj} p_j(\cdot)\right)^2, \beta_{g0} > 0 \Big\}.
\end{aligned}
$$

In the rest of the analysis, we use $J_{k,n}$ to denote the dimension of the sieve space $\mathscr{F}_n$ for $\sqrt{f^0_{\varepsilon_k}}$ for $k = 1, \cdots, K$. Let $J_{K+1,n}$ denote the dimension of $\mathscr{F}_n$ for $\sqrt{f^0_{X^*}}$, and $J'_{k,n}$ denote the dimension of the sieve space $\mathscr{G}_n$ for g^0_k for $k = 2, \cdots, K$. The total number of sieve coefficients in the sieve estimator $\hat{\alpha}_n$ is $k_n = \sum_{k=1}^{K+1} J_{k,n} + \sum_{k=2}^{K} J'_{k,n}$. Importantly, k_n represents the number of constraints imposed during estimation, and it grows with the sample size n to approximate the population parameter.[20]

With the seive spaces $\mathscr{F}_n$ and $\mathscr{G}_n$, the corresponding sieve space for $\widehat{f}_{X^*|\boldsymbol{X}}(X^*|\boldsymbol{X})$ is

$$
\begin{aligned}
\mathscr{H}_n = \Big\{ f(x^*|x) = \frac{\prod_{k=1}^K f_k(x_k - g_k(x^*)) f_{K+1}(x^*)}{\int_{\mathscr{X}^*} \prod_{k=1}^K f_k(x_k - g_k(x^*)) f_{K+1}(x^*) dx^*} : \sqrt{f_k(\cdot)} \in \mathscr{F}_n, \text{ for } k = 1, \cdots, K+1, \\
\text{and } g_k(\cdot) \in \mathscr{G}_n \text{ for } k = 2, 3, \cdots, K \Big\}.
\end{aligned}
$$

[20]While it would be desirable to have a formal selection rule for k_n, providing a general guideline is challenging. A practical approach for determining the tuning parameter k_n is to begin with a sieve approximation of order one or two and then gradually increase the order. The aim is to select the term so that the estimates remain relatively stable and insensitive to minor variations in k_n. This heuristic method can help balance computational complexity with the precision of the estimates.

B.2. Consistency of $\widehat{\alpha}_n$

For each observation i and $\alpha = (\sqrt{f_1}, \cdots, \sqrt{f_K}, \sqrt{f_{K+1}}, g_2, \cdots, g_K)$, the log likelihood function is given by

$$\ell(\boldsymbol{x_i};\alpha) = \ln f(\boldsymbol{x_i};\alpha) = \ln\left(\int_{\mathscr{X}^*} \prod_{k=1}^{K} f_k(x_{ki} - g_k(x^*)) f_{K+1}(x^*) dx^*\right),$$

where $\boldsymbol{x_i}$ is a realization of a random variable $\boldsymbol{X}$ in the sample. We define a strong norm $\|\cdot\|_s$ as

$$\|\alpha\|_s = \sum_{k=1}^{K+1} \|\sqrt{f_k}\|_{\infty,\omega} + \sum_{k=2}^{K} \|g_k\|_{\infty,\omega},$$

where $\|g\|_{\infty,\omega} \equiv \sup_\xi |g(\xi)\omega(\xi)|$ with $\omega(\xi) = (1 + \|\xi\|_E^2)^{-\varsigma/2}$ for some $\varsigma > 0$, and $\|\cdot\|_E$ is the Euclidean norm. We invoke the following assumptions to achieve the consistency of $\widehat{\alpha}_n$.

Assumption B.2. (i) *The data* $\{\boldsymbol{x}_i\}_{i=1}^n$ *are i.i.d.;* (ii) *The density function* $f_{\boldsymbol{X},X^*}(\boldsymbol{X}, X^*)$ *is bounded and bounded away from zero.*

Assumption B.3. *Assumption B.1 holds for a neighborhood of* α_0 *under the norm* $\|\cdot\|_s$.

Assumption B.4. (i) *For any* $\alpha \in \mathscr{A}$, *there exists* $\Pi_n\alpha \in \mathscr{A}_n$ *such that* $\|\Pi_n\alpha - \alpha\|_s = o(1)$*;* (ii) $J_{k,n} \to +\infty$ *and* $J_{k,n}/n \to 0$ *for* $k = 1, \cdots, K+1$ *as* $n \to +\infty$, *and* $J'_{k,n} \to +\infty$ *and* $J'_{k,n}/n \to 0$ *for* $k = 2, \cdots, K$ *as* $n \to +\infty$.

Assumption B.5. (i) $E\{|\ell(\boldsymbol{x_i};\alpha)|^2\}$ *is bounded for all* α*;* (ii) *there exits a positive measurable function* $\widetilde{h}$ *with* $E\{\widetilde{h}(\boldsymbol{X})^2\} < \infty$ *such that, for any*

$$\bar{\alpha} = (\sqrt{\bar{f}_1(\cdot - \cdot)}, \sqrt{\bar{f}_2(\cdot - \bar{g}_2(\cdot))}, \cdots, \sqrt{\bar{f}_K(\cdot - \bar{g}_K(\cdot))}, \sqrt{\bar{f}_{K+1}(\cdot)})$$

and $\bar{\omega}(\boldsymbol{x}, x^*) \equiv \left[\omega^{-1}(x_1, x^*), \cdots, \omega^{-1}(x_K, x^*), \omega^{-1}(x^*)\right]^T$, *we have* $|h_1(\boldsymbol{X}, \bar{\alpha}, \bar{\omega})| < \widetilde{h}(\boldsymbol{X})$, *where*

$$\begin{aligned} h_1(\boldsymbol{X}, \bar{\alpha}, \bar{\omega}) = \frac{1}{|f(\boldsymbol{x_i};\bar{\alpha})|}\Bigg(&\sum_{l=1}^{K} \int_{\mathscr{X}^*} \left|2\sqrt{\bar{f}_l(\bar{\varepsilon}_l)}\omega^{-1}(x_l, x^*) \prod_{\substack{k=1\\k\neq l}}^{K} \bar{f}_k(\bar{\varepsilon}_k)\bar{f}_{K+1}(x^*)\right| dx^* \\ &+ \int_{\mathscr{X}^*} \left|\prod_{k=1}^{K} \bar{f}_k(\bar{\varepsilon}_k) 2\sqrt{\bar{f}_{K+1}(x^*)}\omega^{-1}(x^*)\right| dx^*\Bigg) \quad \text{with } \bar{\varepsilon}_k = x_k - \bar{g}_k(x^*). \end{aligned}$$

Assumption B.2(i) rules out serially dependent observations. Assumptions B.2(ii) and B.3 are standard regularity conditions imposed for a sieve estimation. Assumption B.4(i)

states that there exists a finite dimensional sieve approximation space $\mathscr{A}_n$ to $\mathscr{A}$ and Assumption B.4(ii) imposes that the number of sieve coefficients grows with the sample size n to approximate the population parameter. It is also important to assume that the number of terms in the sieve grows slower than the sample size, so that the degrees of freedom also grow with the sample size. The function $h_1(\boldsymbol{X}, \bar{\alpha}, \bar{\omega})$ in Assumption B.5 is constructed by the path derivatives of $\ell(\boldsymbol{x_i}; \alpha)$. Assumption B.5 ensures that the log-likelihood function for observation i is Hölder continuous with respect to $\alpha \in \mathscr{A}$.

Lemma B.1. *Under Assumptions B.1–B.5,* $\|\widehat{\alpha}_n - \alpha_0\|_s = o_p(1)$.

Proof See Online Appendix C.3. $Q.E.D.$

B.3. Convergence Rate of $\widehat{\alpha}_n$

Next, we consider $n^{-1/4}$ convergence rates of $\widehat{\alpha}_n$ under the weaker Fisher metric $\|\cdot\|_F$ introduced by Ai and Chen (2003). Suppose the function space $\mathscr{A}$ is convex. For any $v \in \bar{V}$, define the pathwise derivative as:

$$\frac{d\ell(\boldsymbol{x_i}; \alpha)}{d\alpha}[v] \equiv \left.\frac{d\ell(\boldsymbol{x_i}; \alpha + \tau v)}{d\tau}\right|_{\tau=0} \quad \text{a.s. } X.$$

For any $\alpha_1, \alpha_2 \in \mathscr{A}$, the Fisher norm is defined as:

$$\|\alpha_1 - \alpha_2\|_F^2 \equiv \mathrm{E}\left\{\left(\frac{d\ell(\boldsymbol{x_i}; \alpha_0)}{d\alpha}[\alpha_1 - \alpha_2]\right)^2\right\}. \tag{B.1}$$

We invoke the following assumptions to obtain a rate faster than $n^{-1/4}$.

Assumption B.6. $(k_n n^{-1/2} \ln n) \times \sup_{(\xi_1, \xi_2) \in \mathbb{R} \times \mathscr{X}^*} \|(h^{J_{k,n}}(\xi_1 - g^{J'_{k,n}}(\xi_2)))^2\|_E^2 = o(1)$, *where* $h^{J_{k,n}}(\varepsilon_k)$ *and* $g^{J'_{k,n}}(X^*)$ *are the sieve approximations of* $\sqrt{f^0_{\varepsilon_k}(\varepsilon_k)}$ *and* $g^0_k(X^*)$.

Assumption B.7. (i) *There exist a measurable function* $H(X)$ *with* $E\{H(X)^4\} < \infty$ *such that* $|\ell(\boldsymbol{x_i}; \alpha)| \le H(X)$ *for all* X *and* $\alpha \in \mathscr{A}_n$; (ii) $\ell(\boldsymbol{x_i}; \alpha) \in \Lambda_c^{\kappa,\omega}(\mathscr{X})$ *with* $\kappa > \dim X/2$, *for all* $\alpha \in \mathscr{A}_n$, *where* $\dim X$ *is the dimension of* X.

Assumption B.8. $\mathscr{A}$ *is convex in* α_0, *and* $f_k(X_k - g_k(x^*))$ *is pathwise differentiable at* g^0_k *for* $k = 2, \cdots, K$.

Assumption B.9. $\ln N(\delta, \mathscr{A}_n) = O(k_n \ln(k_n/\delta))$ *where* $N(\delta, \mathscr{A}_n)$ *is the minimum number of balls with radius* δ *under the* $\|\cdot\|_s$ *norm covering* $\mathscr{A}_n$.

Assumption B.10. *There exist* c_1, $c_2 > 0$,

$$c_1 E\left(\ln \frac{f_X(X_i;\alpha_0)}{f_X(X_i;\alpha)}\right) \leq \|\alpha - \alpha_0\|_F^2 \leq c_2 E\left(\ln \frac{f_X(X_i;\alpha_0)}{f_X(X_i;\alpha)}\right)$$

holds for all $\alpha \in \mathscr{A}_n$ *with* $\|\alpha - \alpha_0\|_s = o(1)$.

Assumption B.11. *For any* $\alpha \in \mathscr{A}$, *there exists* $\Pi_n \alpha \in \mathscr{A}_n$ *such that* $\|\Pi_n \alpha - \alpha\|_F = o(k_n^{-\mu_1})$ *and* $k_n^{-\mu_1} = o(n^{-1/4})$.

Assumption B.6 imposes conditions related to the sieve approximation of $\sqrt{f^0_{\varepsilon_k}(\varepsilon_k)}$ and $g^0_k(X^*)$. It is used to show that the residual part $\widehat{\mathscr{L}_n}(\alpha) - E[\widehat{\mathscr{L}_n}(\alpha)] = o(n^{-1/4})$ uniformly over $\alpha \in \mathscr{A}_n$. Assumption B.7(i) and (ii) impose a dominance condition and smoothness condition on $\ell(\boldsymbol{x_i};\alpha)$. Envelope conditions are imposed to limit changes in the objective function when the parameters change, ensuring stochastic equi-continuity. Assumption B.8 implies that the Fisher norm in Equation (B.1) is well defined. Assumption B.9 requires that the size of the sieve space $\mathscr{A}_n$ does not grow too fast in terms of the covering number. Assumption B.10 assumes that the population criterion function is locally equivalent to the Fisher norm. Assumption B.11 controls the approximation error of $\Pi_n \alpha$ to α and the selection of k_n such that the error goes to zero uniformly at the rate $o_p(n^{-1/4})$ over $\alpha \in \mathscr{A}$.

Theorem B.1. *If Assumptions B.1–B.11 hold,* $\|\widehat{\alpha}_n - \alpha_0\|_F = o_p(n^{-1/4})$.

Proof See Online Appendix C.4. *Q.E.D.*

B.4. Consistency of $\widehat{\boldsymbol{\beta}}_n$

To show that $\widehat{\boldsymbol{\beta}}_n$ is a consistent estimator, we first prove the consistency of the estimated conditional density $\widehat{f}_{X^*|\boldsymbol{X}}(X^*|\boldsymbol{X})$ under a sup norm. $\widehat{f}_{X^*|\boldsymbol{X}}(X^*|\boldsymbol{X})$ enters the sample analogue of the objective function for the optimal linear rank index. We use the consistency results in Lemma B.1 and Theorem B.1 to focus on a neighborhood of α_0. Define $\mathscr{N}_{\alpha_0 n} \equiv \{\alpha \in \mathscr{A}_n : \|\alpha - \alpha_0\|_s = o(1), \|\alpha - \alpha_0\|_F = o_p(n^{-1/4})\}$. We only use the sieve ML estimator $\widehat{\alpha}_n$ in this neighborhood to construct the conditional density estimator

$$\widehat{f}_{X^*|\boldsymbol{X}}(X^*|\boldsymbol{X}) = \frac{\widehat{f}_{\boldsymbol{X},X^*}(\boldsymbol{X}, X^*)}{\int_{\mathscr{X}^*} \widehat{f}_{\boldsymbol{X},X^*}(\boldsymbol{X}, x^*) dx^*}.$$

To achieve the consistency of $\widehat{f}_{X^*|\boldsymbol{X}}(X^*|\boldsymbol{X})$, we first show that

$$\left\|\widehat{f}_{\boldsymbol{X},X^*}(\boldsymbol{X},X^*)-f_{\boldsymbol{X},X^*}(\boldsymbol{X},X^*)\right\|_\infty=o_p(1) \text{ and } \left\|\widehat{f}_{\boldsymbol{X}}(\boldsymbol{X})-f_{\boldsymbol{X}}(\boldsymbol{X})\right\|_\infty=o_p(1).$$

We construct the upper bounds for $\left|\widehat{f}_{\boldsymbol{X},X^*}(\boldsymbol{X},X^*)-f_{\boldsymbol{X},X^*}(\boldsymbol{X},X^*)\right|$ and $\left|\widehat{f}_{\boldsymbol{X}}(\boldsymbol{X})-f_{\boldsymbol{X}}(\boldsymbol{X})\right|$ in Appendix C.1. Specifically,

$$\begin{aligned}
\left|\widehat{f}_{\boldsymbol{X},X^*}(\boldsymbol{X},X^*)-f_{\boldsymbol{X},X^*}(\boldsymbol{X},X^*)\right| &\leq h_2(\boldsymbol{X},X^*,\widehat{\alpha}_n,\bar{\omega})\|\widehat{\alpha}_n-\alpha_0\|_s,\\
\left|\widehat{f}_{\boldsymbol{X}}(\boldsymbol{X})-f_{\boldsymbol{X}}(\boldsymbol{X})\right| &\leq \int_{\mathscr{X}^*} h_2(\boldsymbol{X},x^*,\widehat{\alpha}_n,\bar{\omega})dx^*\|\widehat{\alpha}_n-\alpha_0\|_s,
\end{aligned}$$

where

$$\begin{aligned}
&h_2(\boldsymbol{X},X^*,\widehat{\alpha}_n,\bar{\omega})\\
&\equiv\Bigg|\sum_{l=0}^{K-1}\prod_{j=1}^{l}f^0_{\varepsilon_j}(\varepsilon^0_j)\times\left(\sqrt{\widehat{f}_{l+1}(\widehat{\varepsilon}_{l+1})}+\sqrt{f^0_{\varepsilon_{l+1}}(\varepsilon^0_{l+1})}\right)\omega^{-1}(\boldsymbol{X},X^*)\prod_{k=l+2}^{K}\widehat{f}_k(\widehat{\varepsilon}_k)\widehat{f}_{K+1}(X^*)\\
&+\prod_{k=1}^{K}f^0_{\varepsilon_k}(\varepsilon^0_k)\left(\sqrt{\widehat{f}_{K+1}(X^*)}+\sqrt{f^0_{X^*}(X^*)}\right)\omega^{-1}(X^*)\Bigg| \quad \text{with } \bar{\omega}(\boldsymbol{X},X^*)=(\omega^{-1}(\boldsymbol{X},X^*),\omega^{-1}(X^*)).
\end{aligned}$$

Assumption B.12. *Suppose $\widehat{\alpha}_n\in\mathscr{N}_{\alpha_0 n}$. The following conditions hold: (i) $h_2(\boldsymbol{X},X^*,\widehat{\alpha}_n,\bar{\omega})$ is bounded; and (ii) $\int_{\mathscr{X}^*}h_2(\boldsymbol{X},x^*,\widehat{\alpha}_n,\bar{\omega})dx^*$ is bounded in probability.*

Assumption B.12 (i) and (ii) ensure that $\widehat{f}_{\boldsymbol{X},X^*}(\boldsymbol{X},X^*)$ and $\widehat{f}_{\boldsymbol{X}}(\boldsymbol{X})$ are Hölder continuous in a neighborhood of α_0, respectively. The following corollary provides a consistency result for the estimator $\widehat{f}_{X^*|\boldsymbol{X}}(X^*|\boldsymbol{X})$ under a sup norm.

Corollary B.1. *Suppose that all assumptions in Lemma B.1 and Theorem B.1 hold. Then, under Assumption B.12, $\widehat{f}_{X^*|\boldsymbol{X}}$ in Equation (3.3) satisfies $\left\|\widehat{f}_{X^*|\boldsymbol{X}}-f_{X^*|\boldsymbol{X}}\right\|_\infty=o_p(1)$.*

Proof See Appendix C.1. *Q.E.D.*

Note that $\boldsymbol{\beta}_0$ is assumed to be in the interior of a parameter space $\Theta=[0,1]^K$. Given the definition of the objective function, $Q_0(\boldsymbol{\beta})$ is continuous at $\boldsymbol{\beta}$, and therefore, the consistency of $\widehat{\boldsymbol{\beta}}_n$ follows.

Theorem B.2. *If Assumptions 2.9-2.10, and B.1, B.2–B.12 hold, then $\widehat{\boldsymbol{\beta}}_n\xrightarrow{p}\boldsymbol{\beta}_0$.*

Proof See Appendix C.2. *Q.E.D.*

B.5. Asymptotic Normality of $\widehat{\boldsymbol{\beta}}_n$

With the consistency of $\widehat{\boldsymbol{\beta}}_n$ in Theorem B.2, we consider estimators close to the population parameter $\boldsymbol{\beta}_0$. Let $\mathcal{N}$ be a neighborhood of $\boldsymbol{\beta}_0$. We follow the general method in Sherman (1993) to establish that the proposed optimal linear rank index estimator $\widehat{\boldsymbol{\beta}}_n$ is $\sqrt{n}$–consistent for $\boldsymbol{\beta}_0$ and asymptotically normally distributed. The key challenge we face is that the estimated conditional density $\widehat{f}_{X^*|\boldsymbol{X}}(X^*|\boldsymbol{X})$ enters the sample objective function.

For each $\boldsymbol{x}$ and for each $\boldsymbol{\beta} \in \Theta$, define

$$\Gamma(\boldsymbol{\beta}; f_{X^*|\boldsymbol{X}}) \equiv Q_0(\boldsymbol{\beta}) - Q_0(\boldsymbol{\beta}_0) = E\Big[\int [\mathbf{1}(\boldsymbol{X}_i\boldsymbol{\beta} > \boldsymbol{X}_j\boldsymbol{\beta}) - \mathbf{1}(\boldsymbol{X}_i\boldsymbol{\beta}_0 > \boldsymbol{X}_j\boldsymbol{\beta}_0)]\mathbf{1}(x_i^* > x_j^*) \times f_{X^*|\boldsymbol{X}}(x_i^*|\boldsymbol{X}_i) f_{X^*|\boldsymbol{X}}(x_j^*|\boldsymbol{X}_j) dx_i^* dx_j^*\Big]. \tag{B.2}$$

Assumptions 2.9 and 2.10 ensure that $Q_0(\boldsymbol{\beta})$ is uniquely maximized at $\boldsymbol{\beta}_0$, so that $\Gamma(\boldsymbol{\beta}; f_{X^*|\boldsymbol{X}})$ is also uniquely maximized at $\boldsymbol{\beta}_0$. We therefore rewrite the optimal linear rank index estimator as

$$\widehat{\boldsymbol{\beta}}_n = \arg\max_{\boldsymbol{\beta}} \Gamma_n(\boldsymbol{\beta}; \widehat{f}_{X^*|\boldsymbol{X}}), \tag{B.3}$$

where

$$\Gamma_n(\boldsymbol{\beta}; \widehat{f}_{X^*|\boldsymbol{X}}) = \frac{1}{n(n-1)} \sum_{i \neq j} \int [\mathbf{1}(\boldsymbol{x}_i\boldsymbol{\beta} > \boldsymbol{x}_j\boldsymbol{\beta}) - \mathbf{1}(\boldsymbol{x}_i\boldsymbol{\beta}_0 > \boldsymbol{x}_j\boldsymbol{\beta}_0)]\mathbf{1}(x_i^* > x_j^*) \times \widehat{f}_{X^*|\boldsymbol{X}}(x_i^*|\boldsymbol{x}_i) \widehat{f}_{X^*|\boldsymbol{X}}(x_j^*|\boldsymbol{x}_j) dx_i^* dx_j^*. \tag{B.4}$$

We decompose the sample objective function $\Gamma_n(\boldsymbol{\beta}; \widehat{f}_{X^*|\boldsymbol{X}})$ into two terms

$$\Gamma_n(\boldsymbol{\beta}; \widehat{f}_{X^*|\boldsymbol{X}}) = \underbrace{\Gamma_n(\boldsymbol{\beta}; f_{X^*|\boldsymbol{X}})}_{\text{Term A}} + \underbrace{\big[\Gamma_n(\boldsymbol{\beta}; \widehat{f}_{X^*|\boldsymbol{X}}) - \Gamma_n(\boldsymbol{\beta}; f_{X^*|\boldsymbol{X}})\big]}_{\text{Term B}}. \tag{B.5}$$

Term A in Equation (B.5) is the sample analogue of $\Gamma(\boldsymbol{\beta}; f_{X^*|\boldsymbol{X}})$. It does not depend on $\widehat{f}_{X^*|\boldsymbol{X}}$ and has the same structure as the sample objective in Sherman (1993). We apply the U-statistic decomposition to $\Gamma_n(\boldsymbol{\beta}; f_{X^*|\boldsymbol{X}})$ and analyze the properties of each term separately. To derive the asymptotic properties of Term B in Equation (B.5), we take the pathwise derivative of it along the vector $\widehat{f}_{X^*|\boldsymbol{X}} - f_{X^*|\boldsymbol{X}}$ and then apply a similar U-statistic decomposition. Combing the derivations for Term A and Term B together allows us to write the sample objective function $\Gamma_n(\boldsymbol{\beta};\widehat{f}_{X^*|\boldsymbol{X}})$ as a quadratic approximation, from which we derive the asymptotic distribution of $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0\right)$.

The following notations are introduced for convenience. Let ∇_k denote the k-th partial derivative operator with respect to $\boldsymbol{\beta}$, and $|\nabla_k|g(\boldsymbol{\beta})=\sum_{i_1,\dots,i_k}\left|\frac{\partial^k g(\boldsymbol{\beta})}{\partial\beta_{i_1}\cdots\partial\beta_{i_k}}\right|$. For the U-statistic decomposition for Term A, we define

$$\begin{aligned}
&f_{RC}(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{\beta}) \\
&\quad=\int\left[\mathbf{1}(\boldsymbol{x}_1\boldsymbol{\beta}>\boldsymbol{x}_2\boldsymbol{\beta})-\mathbf{1}(\boldsymbol{x}_1\boldsymbol{\beta}_0>\boldsymbol{x}_2\boldsymbol{\beta}_0)\right]\mathbf{1}(x_1^*>x_2^*)f_{X^*|\boldsymbol{X}}(x_1^*|\boldsymbol{x}_1)f_{X^*|\boldsymbol{X}}(x_2^*|\boldsymbol{x}_2)dx_1^*dx_2^*,
\end{aligned}\tag{B.6}$$

$$f_1(\boldsymbol{x},\boldsymbol{\beta})=E\left[f_{RC}(\boldsymbol{x},\cdot,\boldsymbol{\beta})\right]+E\left[f_{RC}(\cdot,\boldsymbol{x},\boldsymbol{\beta})\right]-2E\left[\Gamma_n(\boldsymbol{\beta};f_{X^*|\boldsymbol{X}})\right],\tag{B.7}$$

$$f_2(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{\beta})=f_{RC}(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{\beta})-E\left[f_{RC}(\boldsymbol{x}_1,\cdot,\boldsymbol{\beta})\right]-E\left[f_{RC}(\cdot,\boldsymbol{x}_2,\boldsymbol{\beta})\right]+E\left[\Gamma_n(\boldsymbol{\beta};f_{X^*|\boldsymbol{X}})\right],\tag{B.8}$$

$$\begin{aligned}
\tau(\boldsymbol{x},\boldsymbol{\beta})=&E_{\boldsymbol{X}_i}\left[\int\mathbf{1}(\boldsymbol{X}_i\boldsymbol{\beta}>\boldsymbol{x}\boldsymbol{\beta})\mathbf{1}(x_i^*>x_j^*)f_{X^*|\boldsymbol{X}}(x_i^*|\boldsymbol{X}_i)f_{X^*|\boldsymbol{X}}(x_j^*|\boldsymbol{x})dx_i^*dx_j^*\right] \\
&+E_{\boldsymbol{X}_j}\left[\int\mathbf{1}(\boldsymbol{x}\boldsymbol{\beta}>\boldsymbol{X}_j\boldsymbol{\beta})\mathbf{1}(x_i^*>x_j^*)f_{X^*|\boldsymbol{X}}(x_i^*|\boldsymbol{x})f_{X^*|\boldsymbol{X}}(x_j^*|\boldsymbol{X}_j)dx_i^*dx_j^*\right].
\end{aligned}\tag{B.9}$$

It follows that $E\left[\tau(\cdot,\boldsymbol{\beta})-\tau(\cdot,\boldsymbol{\beta}_0)\right]=2\Gamma(\boldsymbol{\beta};f_{X^*|\boldsymbol{X}})$. Define the pathwise derivative of $\Gamma_n(\boldsymbol{\beta};f_{X^*|\boldsymbol{X}})$ at the direction $[h-f_{X^*|\boldsymbol{X}}]$ evaluated at $f_{X^*|\boldsymbol{X}}$:

$$\begin{aligned}
\kappa_n(\boldsymbol{\beta},h,f_{X^*|\boldsymbol{X}})=&\frac{1}{n(n-1)}\sum_{i\neq j}\int[\mathbf{1}(\boldsymbol{x}_i\boldsymbol{\beta}>\boldsymbol{x}_j\boldsymbol{\beta})-\mathbf{1}(\boldsymbol{x}_i\boldsymbol{\beta}_0>\boldsymbol{x}_j\boldsymbol{\beta}_0)]\mathbf{1}(x_i^*>x_j^*) \\
&\times\Big(\bar{h}(x_i^*|\boldsymbol{x}_i)(h(x_j^*|\boldsymbol{x}_j)-f_{X^*|\boldsymbol{X}}(x_j^*|\boldsymbol{x}_j)) \\
&+\bar{h}(x_j^*|\boldsymbol{x}_j)(h(x_i^*|\boldsymbol{x}_i)-f_{X^*|\boldsymbol{X}}(x_i^*|\boldsymbol{x}_i))\Big)dx_i^*dx_j^*,
\end{aligned}\tag{B.10}$$

where $\bar{h}$ is a mean value between h and $f_{X^*|\boldsymbol{X}}$. For the U-statistic decomposition for Term B, we define

$$\begin{aligned}
\tau_2(\boldsymbol{x},\boldsymbol{\beta},h,f_{X^*|\boldsymbol{X}})=&E_{\boldsymbol{X}_i}\Big[\int\mathbf{1}(\boldsymbol{X}_i\boldsymbol{\beta}>\boldsymbol{x}\boldsymbol{\beta})\mathbf{1}(x_i^*>x_j^*)\Big(\bar{h}(x_i^*|\boldsymbol{X}_i)(h(x_j^*|\boldsymbol{x})-f_{X^*|\boldsymbol{X}}(x_j^*|\boldsymbol{x})) \\
&+\bar{h}(x_j^*|\boldsymbol{x})(h(x_i^*|\boldsymbol{X}_i)-f_{X^*|\boldsymbol{X}}(x_i^*|\boldsymbol{X}_i))\Big)dx_i^*dx_j^*\Big], \\
&+E_{\boldsymbol{X}_j}\Big[\int\mathbf{1}(\boldsymbol{x}\boldsymbol{\beta}>\boldsymbol{X}_j\boldsymbol{\beta})\mathbf{1}(x_i^*>x_j^*)\Big(\bar{h}(x_i^*|\boldsymbol{x})(h(x_j^*|\boldsymbol{X}_j)-f_{X^*|\boldsymbol{X}}(x_j^*|\boldsymbol{X}_j)) \\
&+\bar{h}(x_j^*|\boldsymbol{X}_j)(h(x_i^*|\boldsymbol{x})-f_{X^*|\boldsymbol{X}}(x_i^*|\boldsymbol{x}))\Big)dx_i^*dx_j^*\Big].
\end{aligned}\tag{B.11}$$

We invoke the following assumptions to show the asymptotic normality of $\widehat{\boldsymbol{\beta}}_n$.

Assumption B.13. *The following conditions hold:*

(i) For each $\boldsymbol{x}$*, all mixed second partial derivatives of* $\tau(\boldsymbol{x},\boldsymbol{\beta})$ *exist on* $\mathcal{N}$*.*

(ii) There is an integrable function $M(\boldsymbol{x})$ *such that for all* $\boldsymbol{x}$*, and* $\boldsymbol{\beta} \in \mathcal{N}$*,*

$$\|\nabla_2\tau(\boldsymbol{x},\boldsymbol{\beta}) - \nabla_2\tau(\boldsymbol{x},\boldsymbol{\beta}_0)\| \leq M(\boldsymbol{x})|\boldsymbol{\beta}-\boldsymbol{\beta}_0|, \tag{B.12}$$

where $\|\cdot\|$ *denotes the matrix norm* $\|(a_{ij})\| = (\sum_{i,j} a_{ij}^2)^{1/2}$.

(iii) $E\left[|\nabla_1\tau(\cdot,\boldsymbol{\beta}_0)|^2\right] < \infty$.

(iv) $E\left[|\nabla_2|\tau(\cdot,\boldsymbol{\beta}_0)\right] < \infty$.

(v) The matrix $E\left[\nabla_2\tau(\cdot,\boldsymbol{\beta}_0)\right] = E\left[\frac{\partial^2\tau(\cdot,\boldsymbol{\beta}_0)}{\partial\beta_i\partial\beta_j}\right]$ *is negative definite and uniformly bounded away from zero.*

Assumption B.14. $\{f_2(\cdot,\cdot,\boldsymbol{\beta}) : \boldsymbol{\beta} \in \Theta\}$ *is Euclidean with a finite envelop.* [21]

Assumption B.15. *For* $\boldsymbol{\beta} \in \mathcal{N}$ *and* $h \in \mathcal{N}_{f_{X^*|\boldsymbol{X}},n}$*,*

$$\kappa_n(\boldsymbol{\beta},h,f_{X^*|\boldsymbol{X}}) = \frac{1}{\sqrt{n}}\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right)' W_{2n} + o_p(|\boldsymbol{\beta}-\boldsymbol{\beta}_0|^2) + o_p(\frac{1}{n}),$$

where $W_{2n} = \sqrt{n}P_n\nabla_1\tau_2(\boldsymbol{x},\boldsymbol{\beta}_0,h;f_{X^*|\boldsymbol{X}})$ *denotes an average of the first derivative terms with* $E\left[|\nabla_1\tau(\cdot,\boldsymbol{\beta}_0)+\nabla_1\tau_2(\boldsymbol{x},\boldsymbol{\beta}_0,f_{X^*|\boldsymbol{X}};f_{X^*|\boldsymbol{X}})|^2\right] < \infty$.

Assumption B.13 contains regularity conditions used for a Taylor expansion of $\tau(\boldsymbol{x},\cdot)$ around $\boldsymbol{\beta}_0$. Assumption B.14 is used to ensure that the remainder term of a Taylor expansion of $\tau(\boldsymbol{x},\cdot)$ around $\boldsymbol{\beta}_0$ has an order $o_p(\frac{1}{n})$. Assumption B.15 underscores the quadratic and linear relationships between the deviation of $\boldsymbol{\beta}$ from $\boldsymbol{\beta}_0$ and the pathwise derivative κ_n, with adjustments for the sample size n. The terms $o_p(|\boldsymbol{\beta}-\boldsymbol{\beta}_0|^2)$ and $o_p(\frac{1}{n})$ denote the rates at which the remaining components become negligible as n grows. This captures the effect of the first-step estimation on the linear rank index estimator $\widehat{\boldsymbol{\beta}}_n$ in a neighborhood of $\boldsymbol{\beta}_0$.

Theorem B.3. *Under Assumptions 2.9–2.10, and B.1–B.15, we obtain*

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0\right) \xrightarrow{d} N\left(0,V^{-1}\Delta V^{-1}\right),$$

where $V = \frac{1}{2}E\left[\nabla_2\tau(\cdot,\boldsymbol{\beta}_0)\right]$ *and* $W_n \xrightarrow{d} N(0,\Delta)$.

[21] The definition of a Euclidean class of functions can be found in Section 2 of Pakes and Pollard (1989).

Proof See Appendix C.3. *Q.E.D.*

Theorem B.3 states our main asymptotic results that the proposed optimal linear rank index estimator $\widehat{\boldsymbol{\beta}}_n$ is $\sqrt{n}$–consistent for $\boldsymbol{\beta}_0$ and asymptotically normally distributed. To make inference, since we have shown the asymptotic normality of $\widehat{\boldsymbol{\beta}}_n$, we suggest a bootstrap procedure for $\widehat{\boldsymbol{\beta}}_n$. More detailed discussions on the consistency of the bootstrap procedure can be found in Chen, Linton, and Van Keilegom (2003).

C. Proofs

C.1. Proof of Corollary B.1

Proof Define $\widehat{f}_{\boldsymbol{X},X^*}(\boldsymbol{X},X^*) = \prod_{k=1}^K \widehat{f}_k(X_k - \widehat{g}_k(X^*))\widehat{f}_{K+1}(X^*)$ and $\widehat{f}_{\boldsymbol{X}}(\boldsymbol{X}) = \int_{\mathcal{X}^*} \widehat{f}_{\boldsymbol{X},X^*}(\boldsymbol{X},x^*)dx^*$. We first show that $\left\|\widehat{f}_{\boldsymbol{X},X^*}(\boldsymbol{X},X^*) - f_{\boldsymbol{X},X^*}(\boldsymbol{X},X^*)\right\|_\infty = o_p(1)$ and $\left\|\widehat{f}_{\boldsymbol{X}}(\boldsymbol{X}) - f_{\boldsymbol{X}}(\boldsymbol{X})\right\|_\infty = o_p(1)$ and then use them to show that

$$\left\|\widehat{f}_{X^*|\boldsymbol{X}}(X^*|\boldsymbol{X}) - f_{X^*|\boldsymbol{X}}(X^*|\boldsymbol{X})\right\|_\infty = o_p(1).$$

Denote $\widehat{\varepsilon}_k = X_k - \widehat{g}_k(X^*)$, $\varepsilon_k^0 = X_k - g_k^0(X^*)$, and $\prod_{j=1}^0 f_{\varepsilon_j}^0(\varepsilon_j^0)\prod_{k=1}^K \widehat{f}_k(\widehat{\varepsilon}_k)\widehat{f}_{K+1}(X^*) = \prod_{k=1}^K \widehat{f}_k(\widehat{\varepsilon}_k)\widehat{f}_{K+1}(X^*)$.

Consider

$$\begin{aligned}
&\left|\widehat{f}_{\boldsymbol{X},X^*}(\boldsymbol{X},X^*) - f_{\boldsymbol{X},X^*}(\boldsymbol{X},X^*)\right| \\
&= \Big|\prod_{k=1}^K \widehat{f}_k(\widehat{\varepsilon}_k)\widehat{f}_{K+1}(X^*) - \prod_{k=1}^K f_{\varepsilon_k}^0(\varepsilon_k^0) f_{X^*}^0(X^*)\Big| \\
&= \Big|\sum_{l=0}^{K-1}\left(\prod_{j=1}^l f_{\varepsilon_j}^0(\varepsilon_j^0)\prod_{k=l+1}^K \widehat{f}_k(\widehat{\varepsilon}_k)\widehat{f}_{K+1}(X^*) - \prod_{j=1}^{l+1} f_{\varepsilon_j}^0(\varepsilon_j^0)\prod_{k=l+2}^K \widehat{f}_k(\widehat{\varepsilon}_k)\widehat{f}_{K+1}(X^*)\right) \\
&\qquad + \left(\prod_{k=1}^K f_{\varepsilon_k}^0(\varepsilon_k^0)\widehat{f}_{K+1}(X^*) - \prod_{k=1}^K f_{\varepsilon_k}^0(\varepsilon_k^0) f_{X^*}^0(X^*)\right)\Big| \\
&= \Big|\sum_{l=0}^{K-1}\prod_{j=1}^l f_{\varepsilon_j}^0(\varepsilon_j^0)\left(\widehat{f}_{l+1}(\widehat{\varepsilon}_{l+1}) - f_{\varepsilon_{l+1}}^0(\varepsilon_{l+1}^0)\right)\prod_{k=l+2}^K \widehat{f}_k(\widehat{\varepsilon}_k)\widehat{f}_{K+1}(X^*) \\
&\qquad + \prod_{k=1}^K f_{\varepsilon_k}^0(\varepsilon_k^0)\left(\widehat{f}_{K+1}(X^*) - f_{X^*}^0(X^*)\right)\Big| \\
&\leq \Big|\sum_{l=0}^{K-1}\prod_{j=1}^l f_{\varepsilon_j}^0(\varepsilon_j^0)\left(\sqrt{\widehat{f}_{l+1}(\widehat{\varepsilon}_{l+1})} + \sqrt{f_{\varepsilon_{l+1}}^0(\varepsilon_{l+1}^0)}\right)\omega^{-1}(\boldsymbol{X},X^*)\prod_{k=l+2}^K \widehat{f}_k(\widehat{\varepsilon}_k)\widehat{f}_{K+1}(X^*)
\end{aligned}$$

$$+\prod_{k=1}^{K} f^0_{\varepsilon_k}(\varepsilon^0_k)\left(\sqrt{\widehat{f}_{K+1}(X^*)}+\sqrt{f^0_{X^*}(X^*)}\right)\omega^{-1}(X^*)\Big|\|\widehat{\alpha}_n-\alpha_0\|_s$$

$$\equiv h_2(\boldsymbol{X},X^*,\widehat{\alpha}_n,\bar{\omega})\|\widehat{\alpha}_n-\alpha_0\|_s, \tag{C.1}$$

where $\bar{\omega}(\boldsymbol{X},X^*)=(\omega^{-1}(\boldsymbol{X},X^*),\omega^{-1}(X^*))$. Since Assumption B.12(i) ensures the boundedness of the function $h_2(\boldsymbol{X},X^*,\widehat{\alpha}_n,\bar{\omega})$, we have

$$\sup_{\boldsymbol{X},X^*}\left|\widehat{f}_{\boldsymbol{X},X^*}(\boldsymbol{X},X^*)-f_{\boldsymbol{X},X^*}(\boldsymbol{X},X^*)\right|=\left\|\widehat{f}_{\boldsymbol{X},X^*}-f_{\boldsymbol{X},X^*}\right\|_\infty=o_p(1) \tag{C.2}$$

by the consistency result in Lemma B.1. Integrating out X^* in Equation (C.1) yields the following inequality

$$\begin{aligned}\left|\widehat{f}_{\boldsymbol{X}}(\boldsymbol{X})-f_{\boldsymbol{X}}(\boldsymbol{X})\right|&=\left|\int_{\mathscr{X}^*}\widehat{f}_{\boldsymbol{X},X^*}(\boldsymbol{X},x^*)-f_{\boldsymbol{X},X^*}(\boldsymbol{X},x^*)dx^*\right|\\&\le\int_{\mathscr{X}^*}\left|\widehat{f}_{\boldsymbol{X},X^*}(\boldsymbol{X},x^*)-f_{\boldsymbol{X},X^*}(\boldsymbol{X},x^*)\right|dx^*\\&\le\int_{\mathscr{X}^*}h_2(\boldsymbol{X},x^*,\widehat{\alpha}_n,\bar{\omega})dx^*\cdot\|\widehat{\alpha}_n-\alpha_0\|_s.\end{aligned} \tag{C.3}$$

Applying Assumption B.12(ii) and Lemma B.1 to the inequality in Equation (C.3), we obtain $\|\widehat{f}_{\boldsymbol{X}}-f_{\boldsymbol{X}}\|_\infty=o_p(1)$. Consider

$$\begin{aligned}&\left|\widehat{f}_{X^*|\boldsymbol{X}}(X^*|\boldsymbol{X})-f_{X^*|\boldsymbol{X}}(X^*|\boldsymbol{X})\right|\\&=\left|\frac{\widehat{f}_{\boldsymbol{X},X^*}(\boldsymbol{X},X^*)}{\widehat{f}_{\boldsymbol{X}}(\boldsymbol{X})}-\frac{f_{\boldsymbol{X},X^*}(\boldsymbol{X},X^*)}{f_{\boldsymbol{X}}(\boldsymbol{X})}\right|\\&=\left|\frac{f_{\boldsymbol{X}}(\boldsymbol{X})\big(\widehat{f}_{\boldsymbol{X},X^*}(\boldsymbol{X},X^*)-f_{\boldsymbol{X},X^*}(\boldsymbol{X},X^*)\big)}{\widehat{f}_{\boldsymbol{X}}(\boldsymbol{X})\cdot f_{\boldsymbol{X}}(\boldsymbol{X})}+\frac{f_{\boldsymbol{X},X^*}(\boldsymbol{X},X^*)\big(f_{\boldsymbol{X}}(\boldsymbol{X})-\widehat{f}_{\boldsymbol{X}}(\boldsymbol{X})\big)}{\widehat{f}_{\boldsymbol{X}}(\boldsymbol{X})\cdot f_{\boldsymbol{X}}(\boldsymbol{X})}\right|\\&=\left|\frac{f_{\boldsymbol{X}}(\boldsymbol{X})\big(\widehat{f}_{\boldsymbol{X},X^*}(\boldsymbol{X},X^*)-f_{\boldsymbol{X},X^*}(\boldsymbol{X},X^*)\big)}{f_{\boldsymbol{X}}(\boldsymbol{X})\cdot\big(\widehat{f}_{\boldsymbol{X}}(\boldsymbol{X})-f_{\boldsymbol{X}}(\boldsymbol{X})\big)+f_{\boldsymbol{X}}(\boldsymbol{X})^2}\right.\\&\qquad\left.+\frac{f_{\boldsymbol{X},X^*}(\boldsymbol{X},X^*)\big(f_{\boldsymbol{X}}(\boldsymbol{X})-\widehat{f}_{\boldsymbol{X}}(\boldsymbol{X})\big)}{f_{\boldsymbol{X}}(\boldsymbol{X})\cdot\big(\widehat{f}_{\boldsymbol{X}}(\boldsymbol{X})-f_{\boldsymbol{X}}(\boldsymbol{X})\big)+f_{\boldsymbol{X}}(\boldsymbol{X})^2}\right|\\&\le\left|\frac{\widehat{f}_{\boldsymbol{X},X^*}(\boldsymbol{X},X^*)-f_{\boldsymbol{X},X^*}(\boldsymbol{X},X^*)}{\big(\widehat{f}_{\boldsymbol{X}}(\boldsymbol{X})-f_{\boldsymbol{X}}(\boldsymbol{X})\big)+f_{\boldsymbol{X}}(\boldsymbol{X})}\right|+\left|\frac{f_{\boldsymbol{X},X^*}(\boldsymbol{X},X^*)\big(f_{\boldsymbol{X}}(\boldsymbol{X})-\widehat{f}_{\boldsymbol{X}}(\boldsymbol{X})\big)}{f_{\boldsymbol{X}}(\boldsymbol{X})\cdot\big(\widehat{f}_{\boldsymbol{X}}(\boldsymbol{X})-f_{\boldsymbol{X}}(\boldsymbol{X})\big)+f_{\boldsymbol{X}}(\boldsymbol{X})^2}\right|\\&\le\frac{\left|\widehat{f}_{\boldsymbol{X},X^*}(\boldsymbol{X},X^*)-f_{\boldsymbol{X},X^*}(\boldsymbol{X},X^*)\right|}{f_{\boldsymbol{X}}(\boldsymbol{X})-\|\widehat{f}_{\boldsymbol{X}}-f_{\boldsymbol{X}}\|_\infty}+\frac{f_{\boldsymbol{X},X^*}(\boldsymbol{X},X^*)\left|f_{\boldsymbol{X}}(\boldsymbol{X})-\widehat{f}_{\boldsymbol{X}}(\boldsymbol{X})\right|}{f_{\boldsymbol{X}}(\boldsymbol{X})^2-f_{\boldsymbol{X}}(\boldsymbol{X})\cdot\|\widehat{f}_{\boldsymbol{X}}(\boldsymbol{X})-f_{\boldsymbol{X}}(\boldsymbol{X})\|_\infty}.\end{aligned}$$

For each $\boldsymbol{X}$ in the domain of $f_{X^*|\boldsymbol{X}}$, we have $f_{\boldsymbol{X}}(\boldsymbol{X}) > 0$. Because $\|\widehat{f}_{\boldsymbol{X}} - f_{\boldsymbol{X}}\|_\infty = o_p(1)$, for sufficient large n, we have $\|\widehat{f}_{\boldsymbol{X}} - f_{\boldsymbol{X}}\|_\infty < \frac{1}{2} f_{\boldsymbol{X}}(\boldsymbol{X})$. This implies that for sufficient large n,

$$\begin{aligned}
&\left|\widehat{f}_{X^*|\boldsymbol{X}}(X^*|\boldsymbol{X}) - f_{X^*|\boldsymbol{X}}(X^*|\boldsymbol{X})\right| \\
&\le \frac{2\left|\widehat{f}_{\boldsymbol{X},X^*}(\boldsymbol{X},X^*) - f_{\boldsymbol{X},X^*}(\boldsymbol{X},X^*)\right|}{f_{\boldsymbol{X}}(\boldsymbol{X})} + \frac{2f_{\boldsymbol{X},X^*}(\boldsymbol{X},X^*)\left|f_{\boldsymbol{X}}(\boldsymbol{X}) - \widehat{f}_{\boldsymbol{X}}(\boldsymbol{X})\right|}{f_{\boldsymbol{X}}(\boldsymbol{X})^2}.
\end{aligned}$$

Applying the results $\|\widehat{f}_{\boldsymbol{X},X^*}(\boldsymbol{X},X^*) - f_{\boldsymbol{X},X^*}(\boldsymbol{X},X^*)\|_\infty = o_p(1)$ and $\|\widehat{f}_{\boldsymbol{X}}(\boldsymbol{X}) - f_{\boldsymbol{X}}(\boldsymbol{X})\|_\infty = o_p(1)$ with $f_{\boldsymbol{X}}(\boldsymbol{X}) > 0$, we obtain the desired result $\|\widehat{f}_{X^*|\boldsymbol{X}}(X^*|\boldsymbol{X}) - f_{X^*|\boldsymbol{X}}(X^*|\boldsymbol{X})\|_\infty = o_p(1)$. *Q.E.D.*

C.2. Proof of Theorem B.2

Proof With the consistency result of $\hat{\alpha}_n$ in Lemma B.1, we prove the consistency of the linear rank index estimator $\widehat{\boldsymbol{\beta}}_n$ using Theorem 2.1 in Newey and McFadden (1994).

We show in Section B.5 that it is equivalent to write the optimal linear rank index estimator as

$$\widehat{\boldsymbol{\beta}}_n = \arg\max_{\boldsymbol{\beta}} \ \Gamma_n(\boldsymbol{\beta}; \widehat{f}_{X^*|\boldsymbol{X}}),$$

where the sample objective function $\Gamma_n(\boldsymbol{\beta}; \widehat{f}_{X^*|\boldsymbol{X}})$ is defined in Equation (B.4). After checking the conditions in Theorem 2.1 in Newey and McFadden (1994), one thing left to verify is the uniform convergence of the sample objective function $\Gamma_n(\boldsymbol{\beta}; \widehat{f}_{X^*|\boldsymbol{X}})$ to $\Gamma(\boldsymbol{\beta}; f_{X^*|\boldsymbol{X}})$ (the population objective function is defined in Equation (B.2)). Using the triangular inequality, we have

$$\begin{aligned}
&\left|\Gamma_n(\boldsymbol{\beta}; \widehat{f}_{X^*|\boldsymbol{X}}) - \Gamma(\boldsymbol{\beta}; f_{X^*|\boldsymbol{X}})\right| \\
&\le \left|\Gamma_n(\boldsymbol{\beta}; \widehat{f}_{X^*|\boldsymbol{X}}) - \Gamma_n(\boldsymbol{\beta}; f_{X^*|\boldsymbol{X}})\right| + \left|\Gamma_n(\boldsymbol{\beta}; f_{X^*|\boldsymbol{X}}) - \Gamma(\boldsymbol{\beta}; f_{X^*|\boldsymbol{X}})\right| \\
&\le \left|\Gamma_n(\boldsymbol{\beta}; \widehat{f}_{X^*|\boldsymbol{X}}) - \Gamma_n(\boldsymbol{\beta}; f_{X^*|\boldsymbol{X}})\right| + \left|\Gamma_n(\boldsymbol{\beta}; f_{X^*|\boldsymbol{X}}) - E\left[\Gamma_n(\boldsymbol{\beta}; f_{X^*|\boldsymbol{X}})\right]\right|.
\end{aligned}$$

The second term on the right-hand is the second-order U−process with zero mean, and Euclidean with a constant envelope. Applying Corollary 7 in Sherman (1994) to the term yields

$$\text{(C.4)} \qquad \sup_{\boldsymbol{\beta}} \left| \Gamma_n(\boldsymbol{\beta}; f_{X^*|\boldsymbol{X}}) - E\left[\Gamma_n(\boldsymbol{\beta}; f_{X^*|\boldsymbol{X}})\right] \right| = O_p(\frac{1}{\sqrt{n}}) = o_p(1).$$

As for the first term, we rewrite $\Gamma_n(\boldsymbol{\beta}; \widehat{f}_{X^*|\boldsymbol{X}}) - \Gamma_n(\boldsymbol{\beta}; f_{X^*|\boldsymbol{X}})$ as follows

$$\begin{aligned}
&\left|\Gamma_n(\boldsymbol{\beta}; \widehat{f}_{X^*|\boldsymbol{X}}) - \Gamma_n(\boldsymbol{\beta}; f_{X^*|\boldsymbol{X}})\right| \\
&= \Big| \frac{1}{n(n-1)} \sum_{i \neq j} \int [\mathbf{1}(\boldsymbol{x}_i \boldsymbol{\beta} > \boldsymbol{x}_j \boldsymbol{\beta}) - \mathbf{1}(\boldsymbol{x}_i \boldsymbol{\beta}_0 > \boldsymbol{x}_j \boldsymbol{\beta}_0)] \mathbf{1}(x_i^* > x_j^*) \\
&\qquad \times \left(\widehat{f}_{X^*|\boldsymbol{x}}(x_i^*|\boldsymbol{x}_i) \widehat{f}_{X^*|\boldsymbol{X}}(x_j^*|\boldsymbol{x}_j) - f_{X^*|\boldsymbol{X}}(x_i^*|\boldsymbol{x}_i) f_{X^*|\boldsymbol{X}}(x_j^*|\boldsymbol{x}_j) \right) dx_i^* dx_j^* \Big| \\
&< \sup_{\boldsymbol{X}_i, \boldsymbol{X}_j} \int \left| \widehat{f}_{X^*|\boldsymbol{X}}(x_i^*|\boldsymbol{X}_i) \widehat{f}_{X^*|\boldsymbol{X}}(x_j^*|\boldsymbol{X}_j) - f_{X^*|\boldsymbol{X}}(x_i^*|\boldsymbol{X}_i) f_{X^*|\boldsymbol{X}}(x_j^*|\boldsymbol{X}_j) \right| dx_i^* dx_j^* \\
&\le \sup_{\boldsymbol{X}_i, \boldsymbol{X}_j} \int \left| \left(\widehat{f}_{X^*|\boldsymbol{X}}(x_i^*|\boldsymbol{X}_i) - f_{X^*|\boldsymbol{X}}(x_i^*|\boldsymbol{X}_i) \right) \widehat{f}_{X^*|\boldsymbol{X}}(x_j^*|\boldsymbol{X}_j) \right| \\
&\qquad + \left| f_{X^*|\boldsymbol{X}}(x_i^*|\boldsymbol{X}_i) \left(\widehat{f}_{X^*|\boldsymbol{X}}(x_j^*|\boldsymbol{X}_j) - f_{X^*|\boldsymbol{X}}(x_j^*|\boldsymbol{X}_j) \right) \right| dx_i^* dx_j^* \\
\text{(C.5)} \quad &\le c^* \left(\sup_{x_i^*, \boldsymbol{X}_i} \left| \widehat{f}_{X^*|\boldsymbol{X}}(x_i^*|\boldsymbol{X}_i) - f_{X^*|\boldsymbol{X}}(x_i^*|\boldsymbol{X}_i) \right| + \sup_{x_j^*, \boldsymbol{X}_j} \left| \widehat{f}_{X^*|\boldsymbol{X}}(x_j^*|\boldsymbol{X}_j) - f_{X^*|\boldsymbol{X}}(x_j^*|\boldsymbol{X}_j) \right| \right),
\end{aligned}$$

for some constant c^*, where we have used Assumptions B.2(ii) and (iii). We have $\left|\Gamma_n(\boldsymbol{\beta}; \widehat{f}_{X^*|\boldsymbol{X}}) - \Gamma_n(\boldsymbol{\beta}; f_{X^*|\boldsymbol{X}})\right| = o_p(1)$ by the consistency result of Corollary B.1. Therefore, we obtain the uniform convergence result,

$$\text{(C.6)} \qquad \sup_{\boldsymbol{\beta}} \left| \Gamma_n(\boldsymbol{\beta}; \widehat{f}_{X^*|\boldsymbol{X}}) - \Gamma(\boldsymbol{\beta}; f_{X^*|\boldsymbol{X}}) \right| = o_p(1).$$

Q.E.D.

C.3. Proof of Theorem B.3

Proof We first apply the U-statistic decomposition to $\Gamma_n(\boldsymbol{\beta}; f_{X^*|\boldsymbol{X}})$, that is Term A in Equation (B.5).

$$\text{(C.7)} \qquad \Gamma_n(\boldsymbol{\beta}; f_{X^*|\boldsymbol{X}}) = E\left[\Gamma_n(\boldsymbol{\beta}; f_{X^*|\boldsymbol{X}})\right] + P_n f_1(\cdot, \boldsymbol{\beta}) + U_n f_2(\cdot, \cdot, \boldsymbol{\beta}),$$

where $P_n f_1(\cdot,\boldsymbol{\beta}) = \frac{1}{n}\sum_i f_1(\boldsymbol{x}_i,\boldsymbol{\beta})$ and $U_n f_2(\cdot,\cdot,\boldsymbol{\beta}) = \frac{1}{n(n-1)}\sum_{i\neq j} f_2(\boldsymbol{x}_i,\boldsymbol{x}_j,\boldsymbol{\beta})$ are degenerate $U-$ processes of the first and the second order, respectively, with f_1 and f_2 defined in Equations (B.7) and (B.8) respectively. We investigate each term on the right-hand side of Equation (C.7) following Sherman (1993).

The **first term** $E\left[\Gamma_n(\boldsymbol{\beta}; f_{X^*|\boldsymbol{X}})\right]$ is connected to the function $\tau(\boldsymbol{x},\boldsymbol{\beta})$ defined in Equation (B.9). Under Assumption B.13(i), for each $\boldsymbol{x}$, a second order Taylor expansion of $\tau(\boldsymbol{x},\boldsymbol{\beta})$ at $\boldsymbol{\beta}_0$ yields

$$\tau(\boldsymbol{x},\boldsymbol{\beta}) - \tau(\boldsymbol{x},\boldsymbol{\beta}_0) = \left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right)' \nabla_1\tau(\boldsymbol{x},\boldsymbol{\beta}_0) + \frac{1}{2}\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right)' \nabla_2\tau(\boldsymbol{x},\beta^*)\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right), \tag{C.8}$$

where β^* lies between $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_0$. Taking the expectation of Equation (C.8) under Assumption B.13(ii) yields

$$2E\left[\Gamma_n(\boldsymbol{\beta}; f_{X^*|\boldsymbol{X}})\right] = 2\Gamma(\boldsymbol{\beta}; f_{X^*|\boldsymbol{X}}) = \left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right)' V\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right) + o_p(|\boldsymbol{\beta}-\boldsymbol{\beta}_0|^2) \tag{C.9}$$

for $\boldsymbol{\beta}\in\mathcal{N}$ and $\boldsymbol{\beta}$ close to $\boldsymbol{\beta}_0$. Note that $E\left[\nabla_1\tau(\cdot,\boldsymbol{\beta}_0)\right] = 0$ because $\Gamma(\boldsymbol{\beta}; f_{X^*|\boldsymbol{X}})$ is maximized at $\boldsymbol{\beta}_0$ and the exchange of the partial derivatives and expectation.

The **second term** $P_n f_1(\cdot,\boldsymbol{\beta})$ can be represented as

$$P_n f_1(\cdot,\boldsymbol{\beta}) = P_n\left[\tau(\cdot,\boldsymbol{\beta}) - \tau(\cdot,\boldsymbol{\beta}_0)\right] - 2\Gamma(\boldsymbol{\beta}; f_{X^*|\boldsymbol{X}}). \tag{C.10}$$

Applying Equations (C.8) and (C.9) to (C.10) and with Assumption B.13(ii), we obtain

(C.11)
$$P_n f_1(\cdot,\boldsymbol{\beta}) = \left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right)' P_n\nabla_1\tau(\boldsymbol{x},\boldsymbol{\beta}_0) + \frac{1}{2}\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right)'\left[P_n\nabla_2\tau(\boldsymbol{x},\boldsymbol{\beta}_0) - 2V\right]\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right) + o_p(|\boldsymbol{\beta}-\boldsymbol{\beta}_0|^2),$$

uniformly over a neighborhood of $\boldsymbol{\beta}_0$. By Assumption B.13(iv) and a weak law of large numbers, $P_n\nabla_2\tau(\boldsymbol{x},\boldsymbol{\beta}_0) - 2V = o_p(1)$ as $n\to\infty$. This implies that

$$P_n f_1(\cdot,\boldsymbol{\beta}) = \frac{1}{\sqrt{n}}\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right)' W_{1n} + o_p(|\boldsymbol{\beta}-\boldsymbol{\beta}_0|^2). \tag{C.12}$$

As for the **third term** $U_n f_2(\cdot,\cdot,\boldsymbol{\beta})$, by Assumption B.14, the collection of functions $\{f_2(\cdot,\cdot,\boldsymbol{\beta}) : \boldsymbol{\beta}\in\Theta\}$ is zero mean, and is Euclidean with a finite envelop. In addition, we have $E\left[f_2(\cdot,\cdot,\boldsymbol{\beta})^2\right] \to$

0 as $\boldsymbol{\beta} \to \boldsymbol{\beta}_0$. Theorem 3 in Sherman (1993) therefore implies that

$$U_n f_2(\cdot,\cdot,\boldsymbol{\beta}) = o_p(\frac{1}{n}) \tag{C.13}$$

uniformly over $o_p(1)$ neighborhoods of $\boldsymbol{\beta}_0$.

Combining the results of the three terms in Equation (C.7), we obtain

$$\Gamma_n(\boldsymbol{\beta}; f_{X^*|\boldsymbol{X}}) = \frac{1}{2}\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right)' V\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right) + \frac{1}{\sqrt{n}}\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right)' W_{1n} + o_p(|\boldsymbol{\beta}-\boldsymbol{\beta}_0|^2) + o_p(\frac{1}{n}), \tag{C.14}$$

where $\boldsymbol{\beta}$ in $o_p(1)$ neighborhoods of $\boldsymbol{\beta}_0$.

Next, we analyze Term B in Equation (B.5).

$$\begin{aligned}
\text{Term B} &= \left[\Gamma_n(\boldsymbol{\beta}; \widehat{f}_{X^*|\boldsymbol{X}}) - \Gamma_n(\boldsymbol{\beta}; f_{X^*|\boldsymbol{X}})\right] \\
&= \frac{d\Gamma_n(\boldsymbol{\beta}; \bar{f}_{X^*|\boldsymbol{X}} + t(\widehat{f}_{X^*|\boldsymbol{X}} - f_{X^*|\boldsymbol{X}}))}{dt}\Big|_{t=0} \\
&= \frac{1}{n(n-1)}\sum_{i\neq j}\int [\mathbf{1}(\boldsymbol{x}_i\boldsymbol{\beta} > \boldsymbol{x}_j\boldsymbol{\beta}) - \mathbf{1}(\boldsymbol{x}_i\boldsymbol{\beta}_0 > \boldsymbol{x}_j\boldsymbol{\beta}_0)]\mathbf{1}(x_i^* > x_j^*) \\
&\qquad \times \Big(\bar{f}_{X^*|\boldsymbol{X}}(x_i^*|\boldsymbol{x}_i)(\widehat{f}_{X^*|\boldsymbol{X}}(x_j^*|\boldsymbol{x}_j) - f_{X^*|\boldsymbol{X}}(x_j^*|\boldsymbol{x}_j)) \\
&\qquad + \bar{f}_{X^*|\boldsymbol{X}}(x_j^*|\boldsymbol{x}_j)(\widehat{f}_{X^*|\boldsymbol{X}}(x_i^*|\boldsymbol{x}_i) - f_{X^*|\boldsymbol{X}}(x_i^*|\boldsymbol{x}_i))\Big) dx_i^* dx_j^* \\
&\equiv \kappa_n(\boldsymbol{\beta}, \widehat{f}_{X^*|\boldsymbol{X}}, f_{X^*|\boldsymbol{X}}),
\end{aligned}$$

where $\bar{f}_{X^*|\boldsymbol{X}}$is a mean value between $\widehat{f}_{X^*|\boldsymbol{X}}$ and $f_{X^*|\boldsymbol{X}}$. We can see that $\kappa_n(\boldsymbol{\beta}, \widehat{f}_{X^*|\boldsymbol{X}}, f_{X^*|\boldsymbol{X}})$ is also a U-statistics of order two and discontinuous at $\boldsymbol{\beta}$. We can adopt the similar U-statistic decomposition to the Term A to deal with $\kappa_n(\boldsymbol{\beta}, \widehat{f}_{X^*|\boldsymbol{X}}, f_{X^*|\boldsymbol{X}})$. Under Assumption B.15, an expansion of $\kappa_n(\boldsymbol{\beta}, \widehat{f}_{X^*|\boldsymbol{X}}, f_{X^*|\boldsymbol{X}})$ in $o_p(1)$ neighborhoods of $\boldsymbol{\beta}_0$ is

$$\kappa_n(\boldsymbol{\beta}, \widehat{f}_{X^*|\boldsymbol{X}}, f_{X^*|\boldsymbol{X}}) = \frac{1}{\sqrt{n}}\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right)' W_{2n} + o_p(|\boldsymbol{\beta}-\boldsymbol{\beta}_0|^2) + o_p(\frac{1}{n}).$$

Combining the derivation for Term A and Term B allows us to write the sample objective function $\Gamma_n(\boldsymbol{\beta}; \widehat{f}_{X^*|\boldsymbol{X}})$ by a quadratic approximation as follows

$$\begin{aligned}
&\Gamma_n(\boldsymbol{\beta}; \widehat{f}_{X^*|\boldsymbol{X}}) \\
&= \frac{1}{2}\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right)' V\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right) + \frac{1}{\sqrt{n}}\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right)' (W_{1n}+W_{2n}) + o_p(|\boldsymbol{\beta}-\boldsymbol{\beta}_0|^2) + o_p(\frac{1}{n})
\end{aligned} \tag{C.15}$$

$$= \frac{1}{2}\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right)' V \left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right) + \frac{1}{\sqrt{n}}\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right)' W_n + o_p(|\boldsymbol{\beta}-\boldsymbol{\beta}_0|^2) + o_p(\frac{1}{n}),$$

where $\boldsymbol{\beta}$ in $o_p(1)$ neighborhoods of $\boldsymbol{\beta}_0$, $\widehat{f}_{X^*|\boldsymbol{X}} \in \mathcal{N}_{f_{X^*|\boldsymbol{X}},n}$ and $W_n \equiv W_{1n} + W_{2n}$. Since the expected value of the sum of $\nabla_1\tau(\cdot,\boldsymbol{\beta}_0)$ and $\nabla_1\tau_2(\boldsymbol{x},\boldsymbol{\beta}_0,f_{X^*|\boldsymbol{X}};f_{X^*|\boldsymbol{X}})$ is zero, and the expected value of the square of this sum is finite (as stated in Assumption B.15), it follows that W_n converges in distribution to a normal distribution with mean zero and variance Δ.

We then derive the expression for $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0\right)$ and obtain the influence function from the expression to show the asymptotic properties of our estimator. Because $\widehat{\boldsymbol{\beta}}_n$ maximizes the sample objective function $\Gamma_n(\boldsymbol{\beta};\widehat{f}_{X^*|\boldsymbol{X}})$ and the parameters $\boldsymbol{\beta}_0 + \frac{V^{-1}W_n}{\sqrt{n}}$ is in $o_p(1)$ neighborhoods of $\boldsymbol{\beta}_0$ for sufficient large n, we have $\Gamma_n(\widehat{\boldsymbol{\beta}}_n;\widehat{f}_{X^*|\boldsymbol{X}}) \geq \Gamma_n(\boldsymbol{\beta}_0 - \frac{V^{-1}W_n}{\sqrt{n}};\widehat{f}_{X^*|\boldsymbol{X}})$. Plugging Equation (C.15) into this inequality yields

$$\begin{aligned}
&\frac{1}{2}\left(\widehat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0\right)' V \left(\widehat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0\right) + \frac{1}{\sqrt{n}}\left(\widehat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0\right)' W_n + o_p(\frac{1}{n}) \\
&= \frac{1}{2n} t_n' V t_n + \frac{1}{n} t_n' W_n + o_p(\frac{1}{n}) \\
&\geq \frac{1}{2n}\left(V^{-1}W_n\right)' V \left(V^{-1}W_n\right) + \frac{1}{n}\left(V^{-1}W_n\right)' W_n + o_p(\frac{1}{n}) \\
&= \frac{1}{2n} t_n^{*\prime} V t_n^* + \frac{1}{n} t_n^{*\prime} W_n + o_p(\frac{1}{n}),
\end{aligned}$$

where $t_n = \sqrt{n}\left(\widehat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0\right)$ and $t_n^* = -V^{-1}W_n$. We multiply this expression by $2n$, rearrange terms, and use the facts that $\frac{-V^{-1}W_n}{\sqrt{n}}$ maximizes $\frac{1}{2n}\theta' V\theta + \frac{1}{n}\theta' W_n$ and V is negative definite by Assumption B.13(v) to get

$$0 \leq -\left(t_n - t_n^*\right)' V \left(t_n - t_n^*\right) \leq o_p(1).$$

This implies that $t_n = t_n^* + o_p(1)$ or $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0\right) = -V^{-1}W_n + o_p(1)$, which leads to the desired asymptotic result, $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0\right) \xrightarrow{d} N\left(0, V^{-1}\Delta V^{-1}\right)$. *Q.E.D.*