Econometric Theory, **0**, 2024, 1–27.
doi:10.1017/S0266466624000318

NONPARAMETRIC IDENTIFICATION AND ESTIMATION OF A GENERALIZED ADDITIVE MODEL WITH A FLEXIBLE ADDITIVE STRUCTURE AND UNKNOWN LINK

Songnian Chen
Zhejiang University

Nianqing Liu
Xiamen University

Jian Zhang
Nankai University

Yahong Zhou
Shanghai University of Finance and Economics
and
Shanghai Institute for Mathematics and Interdisciplinary Sciences

This paper proposes a nonparametric approach to identify and estimate the generalized additive model with a flexible additive structure and with possibly discrete variables when the link function is unknown. Our approach allows for a flexible additive structure which provides applied researchers the flexibility to specify their model according to economic theory or practical experience. Motivated by the concerns from empirical research, our method also allows for multiple discrete variables in the covariates. By transforming our model into a generalized additive model with univariate component functions, our identification and estimation thereby follows a procedure adapted from the case with univariate components. The estimators converge to normal distributions in large sample with a one-dimensional convergence rate for the link function and a d_k-dimensional convergence rate for the component function $f_k(\cdot)$ defined on $\mathbb{R}^{d_k}$ for all k.

We thank the Co-Editor (Liangjun Su) and three anonymous referees for comments that have greatly improved the article. We also thank the Editor (Peter C.B. Phillips) for his help on improving the presentation of this article. We are very grateful to Chunrong Ai, Xiaohong Chen, Yanqin Fan, LungFei Lee, and Arthur Lewbel for their constructive comments and suggestions. We are also grateful to Liang Chen, Timothy Christensen, Qingfeng Liu, Ryo Okui, Quang Vuong, Hanghui Zhang, Xinyu Zhang, and seminar/conference participants at HKUST, SUFE, Xiamen University, 2019 Asian Meeting of Econometric Society, 2019 China Meeting of Econometric Society, and 2019 Young Econometrician Asia Pacific (YEAP) for helpful comments. The usual disclaimer applies. Liu acknowledges support from the National Natural Science Foundation of China (Grant No. 72394392). Zhou acknowledges support from the National Natural Science Foundation of China (Grant Nos. 72342034 and 72173083). Zequn Jin provided capable research assistance. Address correspondence to Jian Zhang, School of Economics, Nankai University, Tianjin, China; e-mail: jian.zhang@nankai.edu.cn

1. INTRODUCTION

Flexibility in model specification is one of the key features pursued by applied economists while using nonparametric method. This is because different applications need different model specifications that are often guided by economic theory or practical experience. In addition, there are potentially multiple discrete covariates in many economic datasets. It is hence an empirical concern to handle discrete covariates appropriately in the estimation procedures of economic applications.

In this paper, we address the above two concerns in estimating the generalized additive model with an unknown link function as

$$H(x) = G\Big(\sum_{k=1}^{K} f_k(x^k)\Big), \tag{1}$$

where $H(\cdot)$ is a function that can be consistently estimated (such as nonparametric regression), but the link function $G(\cdot)$ and the component functions $f_k(\cdot)$'s are unknown with $x \equiv (x^1, \ldots, x^K) \in \mathbb{R}^d$ and $x^k \in \mathbb{R}^{d_k}$. The first concern is addressed by allowing a flexible grouping of the covariates in the sub-vectors x^k for $k = 1, \ldots, K$.[1] In this way, a researcher can group the covariates according to economic theory or practical experience, instead of having to restrict one or all of sub-vectors to be univariate. The second concern is addressed by allowing discrete covariates in the estimation procedure. The parametric version of this functional restriction has been implemented in many economic applications, including the very popular specification of constant elasticity of substitution in the estimation of production function. See, e.g., Kmenta (1967), Hodges (1969), Paraskevopoulos (1979), Antras (2004), Klump, McAdam, and Willman (2007), and Berkowitz, Ma, and Nishioka (2017), among others.

To identify the model primitives of $G(\cdot)$ and $f_k(\cdot)$'s, we transform the model (1) by a known mapping into a new model with the link function $G(\cdot)$ and some univariate component functions $\tilde{f}_k(\cdot)$'s. We then identify the new model by applying some existing identification approach to the generalized additive model with univariate components. Closely following the identification strategy, we propose a three-step procedure to estimate the link $G(\cdot)$ and the original components $f_k(\cdot)$. The consistency and asymptotic normality is then established for the estimator of the link $G(\cdot)$ at a one-dimensional convergence rate and for the estimator of the component $f_k(\cdot)$ at a d_k-dimensional convergence rate.

Our paper contributes to the estimation of the generalized additive model. With a known link function and only univariate component functions, Chen et al. (1996), Linton and Härdle (1996), Horowitz and Mammen (2004), and Ma (2012), among others, estimated the univariate components at a one-dimensional convergence rate. Their estimators, hence, have no curse of dimensionality. With an unknown link and only univariate components, Horowitz (2001), Horowitz and Mammen

[1] The generalized additive model requires that the component functions are non-overlapping in their arguments, namely any two of $x^1, \ldots, x^K$ do not share common elements.

(2007, 2011), and Lin et al. (2018), among others, recovered the univariate components still at a one-dimensional convergence rate and hence avoided the curse of dimensionality. Jacho-Chávez, Lewbel, and Linton (2010; JLL hereafter) generalized the framework with only univariate components (and an unknown link) to allow multivariate components, as long as one component function is univariate. Our paper further generalizes the model to allow for a flexible specification of additivity, and the existence of a univariate component function is not needed. In a related area, Lewbel, Lu, and Su (2015) provided a nonparametric test of whether the monotonic transformation structure is correctly specified. With a weaker notion of separability, Pinkse (2001) developed the estimators of $\tilde{f}_1(\cdot),\ldots,\tilde{f}_K(\cdot)$ in a nonparametric regression with weak separability as $E(Y|X=x,Z=z)=\tilde{G}(x,\tilde{f}_1(z^1),\ldots,\tilde{f}_K(z^K))$ where $\tilde{G}$ is monotone in $\tilde{f}_1,\ldots,\tilde{f}_K$, and furthermore all of $\tilde{f}_1(\cdot),\ldots,\tilde{f}_K(\cdot)$ are monotone in their respective first arguments. He showed that the functions $\tilde{f}_1(\cdot),\ldots,\tilde{f}_K(\cdot)$ can be identified up to a monotonic transformation. The generalized additive model is in general identified up to location and sign-scale normalizations.[2] The papers most relevant to ours are Horowitz (2001) and JLL in this research line. To clarify our contributions relative to them, consider model (1) with $K=2$ and $d_1,d_2\geq 2$. Horowitz (2001) identified such a model by further imposing an additive structure on both $f_1(\cdot)$ and $f_2(\cdot)$ as $f_1(x^1)=\sum_{k=1}^{d_1}f_{1k}(x_k^1)$ and $f_2(x^2)=\sum_{k=1}^{d_2}f_{2k}(x_k^2)$. Although such an extra additive structure reduces the dimensionality of this problem to 1, it is vulnerable to misspecification error. Relevant economic theory might rule out any additional additive structure on the components of $f_1(\cdot)$ and $f_2(\cdot)$. JLL identified this model by imposing an additive structure on one of $f_1(\cdot)$ and $f_2(\cdot)$ as $f_1(x^1)=f_{11}(x_1^1)+f_{12}(x_2^1,\ldots,x_{d_1}^1)$ or $f_2(x^2)=f_{21}(x_1^2)+f_{22}(x_2^2,\ldots,x_{d_2}^2)$. The extra additive structure imposed by JLL is weaker than the one of Horowitz (2001), but their identification requires a large image/support condition (see Condition I2(iv) of their Assumption I) which substantially restricts its applicability in real empirical applications. Their identification strategy also rules out discrete elements in x^1 and x^2 (see condition I1 of their Assumption I), and hence further restricts their applicability in real applications.[3] In contrast, we identify such a model without imposing any extra additive structure or any large image/support condition. Our identification approach also allows for discrete elements in x^1 and x^2.

Our paper also contributes to the research on the identification of model primitives by exploiting the monotonicity restrictions on nonparametric functions. One of our key identification steps exploits the monotonicity of the unknown link $G(\cdot)$ to transform the original model into a new model with univariate components.

[2]Other related papers include Ma and Song (2015) who estimated the unknown link function of varying index coefficient models by the means of B-splines, as well as Kohler and Krzyżak (2017) and Schmidt-Hieber (2020). The latter two articles estimated nonparametric regression by deep neural network methods, and have natural links to the generalized additive model.

[3]Note that discrete regressors are still not allowed to enter any component functions in their extension to handle discrete regressors (see Section 6 of Jacho-Chávez, Lewbel, and Linton, 2010).

https://doi.org/10.1017/S0266466624000318 Published online by Cambridge University Press

Our identification is hence established by this connection between our model with a flexible grouping and the transformed model with univariate components. The identification of the latter has been well studied. The monotonicity of transformation function has been employed to identify the model primitives of different variants of transformation model by, e.g., Khan (2001), Chen (2002, 2010a, 2010b, 2012), and Chen and Zhang (2020). Moreover, the monotonicity of nonparametric function on latent random variables has been used to identify the non-separable models by, e.g., Chesher (2003) and Matzkin (2003). In the auction literature, the monotonicity of bidding strategy helps to identify the value distribution by, e.g., Guerre, Perrigne, and Vuong (2000, 2009), Athey and Haile (2002), Li and Zheng (2009), Marmer and Shneyerov (2012), Gentry and Li (2014), and Li and Liu (2018). The monotonicity of strategies is also used to identify discrete games by, e.g., Tang (2010), De Paula and Tang (2012), Grieco (2014), and Liu, Vuong, and Xu (2017), with a notable exception of Lewbel and Tang (2015). To test whether monotonicity restrictions hold, Hoderlein et al. (2016) provided a testing procedure in the structural model without strategic interaction; while Liu and Vuong (2020) proposed nonparametric tests for monotonicity of strategies in Bayesian games.[4]

The rest of this paper is organized in the following way. Section 2 presents our generalized additive model with two component functions. It also lays out our strategy to identify the link function $G(\cdot)$ and the component functions $f_k(\cdot)$ for $k = 1,2$. In Section 3, we propose a nonparametric estimation procedure closely following the identification strategy. Section 4 then establishes the large-sample properties of our estimators. In Section 5, a simulation is used to demonstrate the finite sample performance of our nonparametric estimators. Section 6 briefly discusses how to extend our framework to cases with discrete covariates and more than two components. Section 7 concludes. The appendix collects the proofs of our theorems. The Supplementary Material (SM) collects some notations and technical lemmas (as well as their proofs).

2. THE MODEL AND IDENTIFICATION

We consider the generalized additive model with an unknown link function as follows:

$$H(x) = G\Big(\sum_{k=1}^{K} f_k(x^k)\Big), \tag{2}$$

where $H(\cdot)$ is a function that can be identified directly by the joint distribution of observables and can therefore be consistently estimated, such as the mean regression function $E(Y|X = \cdot)$ or the quantile regression function $Q_{Y|X}(\tau_0|\cdot)$ for a given τ_0, and $x = (x^1, \cdots, x^K)$ such that $x^k \in \mathbb{R}^{d_k}$ for $d_k \geq 1$. The parameters of

[4] While maintaining the monotonicity restriction on bidding strategies, Liu and Luo (2017) proposed a nonparametric inference procedure to compare the valuation distributions in first price auctions.

interest include the unknown link function $G(\cdot)$ and the component functions $f_k(\cdot)$ for $k = 1,\ldots,K$.

For ease of exposition, we focus on the case of two component functions in the link, i.e., the model is simplified as

$$H(x) = G(f_1(x^1) + f_2(x^2)), \qquad \textbf{(M1)}$$

where the unknown link function $G(\cdot)$ is monotonic, $x \equiv (x^1, x^2) \in \mathbb{R}^d$ and $x^k \in \mathbb{R}^{d_k}$ for $k = 1,2$. We will return to the general case with more than two components in Section 6.2. Clearly, $d = d_1 + d_2$. Throughout the paper, we let $X \equiv (X^1, X^2)$ be a random vector in $\mathbb{R}^d$ with X^k denoting a random vector in $\mathbb{R}^{d_k}$. Let $x \equiv (x^1, x^2)$ be the realized value of X with $x^k \in \mathbb{R}^{d_k}$, for $k = 1,2$. In addition, let $p_V(\cdot)$ and $p_{V^s|V^t}(\cdot|v^t)$ denote the probability density function of any given random vector/variable V and the conditional density function of V^s given $V^t = v^t$, respectively.

In this paper, we provide the identification and estimation of $G(\cdot)$ and $f_k(\cdot)$'s under reasonably weak restrictions motivated by empirical concerns. Specifically, we allow for a flexible division of (x^1, x^2) guided by economic theory or practical experience,[5] and discrete variables in x^1 and/or x^2. The latter is motivated by the presence of discrete variables in many economic datasets. For presentation purposes, we first consider the case of (x^1, x^2) to only have continuous variables. We then return to the case with discrete variables in Section 6.1.

We obtain the nonparametric identification of (M1) in three steps. In the first step, we transform it into a new generalized additive model with univariate components. The new model has the same link function as (M1). In the second step, the transformed model is identified by a strategy adapted from Horowitz (2001). The original component functions are identified in the third step by applying the inverse of step-one transformation.

We first transform the original model (M1) into a new model with univariate components. Such a transformation is given by the following theorem.

THEOREM 1. *Under a strictly monotonic link function $G(\cdot)$, the generalized additive model (M1) can be transformed equivalently to*

$$\mathcal{H}(z) = G(\tilde{f}_1(z^1) + \tilde{f}_2(z^2)), \qquad \textbf{(M2)}$$

where $\mathcal{H}(z) = E[H(X)|\zeta_1(X^1) = z^1, \zeta_2(X^2) = z^2]$, the inverse of $\tilde{f}_k(\cdot)$ is $\tilde{f}_k^{-1}(s) = \int G(s + f_{-k}(x^{-k})) \cdot w_{-k}(x^{-k})dx^{-k}$, and $\zeta_k(x^k) = \int H(x) \cdot w_{-k}(x^{-k})dx^{-k}$ with freely chosen nonnegative weight functions $w_k(\cdot)$ for $k = 1,2$ where $-k$ denotes the index other than k in $\{1,2\}$.

Theorem 1 transforms the original model (M1) into a new model (M2) which is easier to analyze for two reasons. First, the new function $\mathcal{H}(\cdot)$ can be identified,

[5] For example, let $(x^1, x^2) = (x_1, x_2, x_3, x_4)$. Our model allows all possible divisions, such as $x^1 = (x_1, x_2), x^2 = (x_3, x_4)$ or $x^1 = x_1, x^2 = (x_2, x_3, x_4)$.

https://doi.org/10.1017/S0266466624000318 Published online by Cambridge University Press

since the function $H(\cdot)$ and hence its weighted integrals $\zeta_k(\cdot)$'s are identified. Second, both of the new components $\tilde{f}_1(\cdot)$ and $\tilde{f}_2(\cdot)$ are univariate. Moreover, the functions $\tilde{f}_k(\cdot)$'s and their inverses are monotonic when the link $G(\cdot)$ is monotonic. To simplify the notation, hereafter let $Z=(Z^1,Z^2)$ with $Z^k=\zeta_k(X^k)$, and $z=(z^1,z^2)$ with $z^k\in\mathbb{R}$ for $k=1,2$.

Before proceeding with the identification of new model (M2), we give the identifying assumptions as follows.

Assumption I (Identification condition). (i) Location normalization: $\tilde{f}_1(z_0^1)=\tilde{f}_2(z_0^2)=0$ for some interior point (z_0^1,z_0^2) in the support of Z.

(ii) Scale normalization: $\int w_3(z^1)/\tilde{f}_1'(z^1)dz^1=1$ where $w_3(\cdot)$ is some nonnegative weight function by choice.

(iii) Monotonicity: the link function $G(\cdot)$ is strictly monotonic.

Parts (i) and (ii) of Assumption I specify the location and scale normalizations needed for the identification. Similar normalizations have been adopted by the literature (see, e.g., Horowitz, 2001) to identify the generalized additive model. Note that our identification strategy still works (with minor change) if the location normalization is relaxed to $\tilde{f}_k(z_0^k)=\tilde{f}_{k0}$ with some known constant $\tilde{f}_{k0}\in\mathbb{R}$ for $k=1,2$. We can also adopt other location and scale normalizations, such as the ones of JLL. Part (iii) imposes a monotonicity condition on the link function $G(\cdot)$. Such a monotonicity condition is used to guarantee the existence of $\tilde{f}_k(\cdot)$'s and their inverses.

In the second step, we turn to identify the new model (M2). Such an identification is achieved in two stages by applying a strategy adapted from Horowitz (2001). In the first stage, we identify the transformed components $\tilde{f}_1(\cdot)$ and $\tilde{f}_2(\cdot)$. In the second stage, the unknown link $G(\cdot)$ is identified.

We now turn to the identification of transformed components $\tilde{f}_k(\cdot)$'s. Let $\mathcal{H}(z)=E[H(X)|Z=z]$, and $\partial_k g(z)=\partial g(z)/\partial z_k$ for any multivariate function $g(z)$. The identification idea comes from the following two basic equations:

$$\partial_1\mathcal{H}(z)=G'(\tilde{f}_1(z^1)+\tilde{f}_2(z^2))\cdot\tilde{f}_1'(z^1), \tag{3}$$

$$\partial_2\mathcal{H}(z)=G'(\tilde{f}_1(z^1)+\tilde{f}_2(z^2))\cdot\tilde{f}_2'(z^2). \tag{4}$$

Letting (4) be divided by (3), we obtain

$$\frac{\tilde{f}_2'(z^2)}{\tilde{f}_1'(z^1)}=\frac{\partial_2\mathcal{H}(z)}{\partial_1\mathcal{H}(z)}. \tag{5}$$

We next multiply both sides by $w_3(z^1)$ and integrate (i) by z^1 on the whole support of $Z^1\equiv\zeta_1(X^1)$ and (ii) by z^2 from z_0^2 to z^2, and get

$$\tilde{f}_2(z^2)=\int_{z_0^2}^{z^2}\tilde{f}_2'(z^2)dz^2\cdot\int w_3(z^1)/\tilde{f}_1'(z^1)dz^1=\int_{z_0^2}^{z^2}\int\frac{\partial_2\mathcal{H}(z)}{\partial_1\mathcal{H}(z)}\cdot w_3(z^1)dz^1dz^2, \tag{C1}$$

https://doi.org/10.1017/S0266466624000318 Published online by Cambridge University Press

where the first equality comes from the location and scale normalizations imposed by Assumption I. *The second transformed component function* $\tilde{f}_2(\cdot)$ *is hence identified by (C1).*

The identification of $\tilde{f}_1(\cdot)$ follows a similar strategy. Specifically, applying previous strategy to $\tilde{f}_1'(z^1)/\tilde{f}_2'(z^2) = [\partial_1\mathcal{H}(z)]\big/[\partial_2\mathcal{H}(z)]$, we get

$$\tilde{f}_1(z^1) = \Big[\int_{z_0^1}^{z^1}\int \frac{\partial_1\mathcal{H}(z)}{\partial_2\mathcal{H}(z)}\cdot w_4(z^2)dz^2dz^1\Big]\Big/\Big[\int w_4(z^2)/\tilde{f}_2'(z^2)dz^2\Big]. \quad \textbf{(6)}$$

We can identify the first transformed component $\tilde{f}_1(\cdot)$ if the denominator $\int w_4(z^2)/\tilde{f}_2'(z^2)dz^2$ can be identified. This is achieved by the scale normalization and (5) as

$$\frac{1}{\int w_4(z^2)/\tilde{f}_2'(z^2)dz^2} = \frac{\int \frac{w_3(z^1)}{\tilde{f}_1'(z^1)}dz^1}{\int \frac{w_4(z^2)}{\tilde{f}_2'(z^2)}dz^2} = \int \frac{w_3(z^1)}{\int \frac{\tilde{f}_1'(z^1)}{\tilde{f}_2'(z^2)}\cdot w_4(z^2)dz^2}dz^1 = \int \frac{w_3(z^1)}{\int \frac{\partial_1\mathcal{H}(z)}{\partial_2\mathcal{H}(z)}\cdot w_4(z^2)dz^2}dz^1,$$

which introduces an expression to identify the first transformed component $\tilde{f}_1(\cdot)$ as follows:

$$\tilde{f}_1(z^1) = c\cdot\int_{z_0^1}^{z^1}\int \frac{\partial_1\mathcal{H}(z)}{\partial_2\mathcal{H}(z)}\cdot w_4(z^2)dz^2dz^1, \quad \textbf{(C2)}$$

where $c = \int \omega_3\big(z^1\big)\cdot\Big[\int\big[\partial_1\mathcal{H}(z)\big/\partial_2\mathcal{H}(z)\big]\cdot\omega_4\big(z^2\big)dz^2\Big]^{-1}dz^1$. Consequently, *the first transformed component function* $\tilde{f}_1(\cdot)$ *is identified by (C2).*

After identifying the transformed components $\tilde{f}_k(\cdot)$'s, we now investigate the identification of the unknown link $G(\cdot)$. The function $T(z) = \tilde{f}_1(z^1) + \tilde{f}_2(z^2)$ is identified once the transformed components $\tilde{f}_1(\cdot)$ and $\tilde{f}_2(\cdot)$ are identified. *The unknown link function* $G(\cdot)$ *is then identified* by the nonparametric regression of $H(X)$ on $T(Z)$, namely $E\big[H(X)\big|T(Z)\big]$, due to the following result:

$$E\big[H(X)\big|T(Z)=\tau\big] = E\big[\mathcal{H}(Z)\big|T(Z)=\tau\big] = G(\tau), \quad \textbf{(L1)}$$

where the first equality comes from the fact that, given $T(Z)$, the conditional expectation of $H(X)$ and $\mathcal{H}(Z) = E\big[H(X)\big|Z\big]$ are the same by the law of iterated expectation; and the second equality holds due to the restriction given by (M2). In particular, when $H(X)$ is a nonparametric regression $E(Y|X)$, by the law of iterated expectation, the identification equation (L1) for the link $G(\cdot)$ can be further simplified as

$$G(\tau) = E\big[Y\big|T(Z)=\tau\big]. \quad \textbf{(L2)}$$

In the final step, we use the inverse of step-one transformation to identify the original components $f_1(\cdot)$ and $f_2(\cdot)$. Notice that the original link $G(\cdot)$ has already been identified in step two. This is accomplished by the following mapping from

https://doi.org/10.1017/S0266466624000318 Published online by Cambridge University Press

the inverse of step-one transformation:

$$f_k(x^k) = \tilde{f}_k(\zeta_k(x^k)), \text{ for } k = 1,2, \tag{7}$$

which can be derived by replacing s with $f_k(x^k)$ in the expressions of $\tilde{f}_k^{-1}(\cdot)$ of Theorem 1 and exploring the equality of (M1). *Both of the original component functions $f_k(\cdot)$ for $k = 1,2$ are then identified*, since $\zeta_k(\cdot)$'s are identified functions by their definitions in Theorem 1, and $\tilde{f}_k(\cdot)$'s have been identified in step two.

We summarize the above discussion on the identification of the link function $G(\cdot)$ and the original component functions $f_k(\cdot)$'s in the following theorem whose proof is omitted.

THEOREM 2. *Suppose Assumption I holds. Given the expressions in (C1), (C2), and (L1) are well defined, the link function $G(\cdot)$ is identified by (L1), and the original component functions are identified by (7) where the transformed component functions $\tilde{f}_k(\cdot)$'s are given by (C2) and (C1), and the weighted integrals $\zeta_k(\cdot)$'s are defined by Theorem 1 for $k = 1,2$. In particular, when $H(x) = E(Y|X = x)$, the link function $G(\cdot)$ is identified by a simplified expression as (L2).*

Theorem 2 identifies the link $G(\cdot)$ and the original components $f_k(\cdot)$'s for $k = 1,2$ by applying Horowitz's (2001) strategy to the transformed model (M2) in Theorem 1. In addition, Theorem 2 establishes the identification of model primitives when there are only two components within the link. Such an identification strategy can be easily extended to the case of more than two components (i.e., $K > 2$). We will briefly discuss such an extension in Section 6.2.

Remark 1. Theorem 2 shows that the link $G(\cdot)$ and components $f_k(\cdot)$'s are identified under each chosen set of weights $w_k(\cdot)$, $k = 1,\ldots,4$. The choice of weights $w_k(\cdot)$ affects the efficiency of estimating $G(\cdot)$ and $f_k(\cdot)$'s.

3. ESTIMATION

This section only considers estimating the parameter of interest in the case of nonparametric (mean) regression for $H(\cdot)$, i.e., $H(x) = E(Y|X = x)$. We leave other cases of $H(\cdot)$ (such as the case of quantile regression) for future research. In the case of nonparametric regression, note that

$$E\big[Y\big|Z = z\big] = \mathcal{H}(z) = G\big(\tilde{f}_1(z^1) + \tilde{f}_2(z^2)\big) \tag{8}$$

by the law of iterated expectation. That is, our estimation problem is essentially the same as Horowitz (2001), in which all component functions are univariate, if the true $\zeta_1(\cdot)$ and $\zeta_2(\cdot)$ were used in our estimation. Consequently, we propose a three-step estimation procedure by the kernel method to recover the parameter of interest, namely the link function $G(\cdot)$ and the component functions $f_k(\cdot)$ for $k = 1,2$. We leave other nonparametric alternatives such as the sieve method proposed by, e.g., Ai and Chen (2003) and Chen (2007), for future research. In the first step, the nonparametric regression $H(\cdot)$ and its partial integrals $\zeta_k(\cdot)$'s are recovered by the local polynomial method, then the partial derivatives $\partial_k \mathcal{H}(\cdot)$ for $k=1,2$ are estimated by another local polynomial regression of Y on two generated regressors $\widehat{Z}^1 = \widehat{\zeta}_1(X^1)$ and $\widehat{Z}^2 = \widehat{\zeta}_2(X^2)$. In step two, the (transformed) components $\tilde{f}_k(\cdot)$'s are estimated through the expressions of (C2) and (C1) by replacing $\partial_k \mathcal{H}(\cdot)$ for $k=1,2$ with their step-one estimates, and the link $G(\cdot)$ is recovered through the local polynomial regression according to (L2). In the third step, the original components $f_k(\cdot)$ for $k=1,2$ are then recovered according to (7) by replacing $\tilde{f}_k(\cdot)$ and $\zeta_k(\cdot)$ with their nonparametric estimates.

Such a kernel estimation approach has several attractive features. First, the estimation strategy closely follows the identification idea laid out in Section 2. In particular, it transforms the estimation problem with multivariate components to the one with univariate components. The latter has been well studied in the literature. Second, it can group (X^1, X^2) in a flexible way. This flexibility can be important to adopt the generalized additive model in empirical applications, since many applications may specify some or even all component functions to be multivariate. Third, we only need one continuous variable in X^1 and X^2, namely, the other covariates in X^1 and X^2 can be all discrete. For presentation purposes, we consider the case of (X^1, X^2) to only have continuous variables here. We will return to the case with discrete variables in Section 6.1.

Specifically, our estimation approach proceeds in three steps as follows.

Step 1. Estimation of $\partial_k \mathcal{H}(\cdot)$. We first use a local rth-order polynomial method to estimate $H(x) = E[Y|X^1 = x^1, X^2 = x^2]$[6]. We use a leave-one-out estimator $\widehat{H}_{-j}(x)$, namely, the intercept of

$$\widehat{\alpha} = \arg\min_{\alpha} \sum_{i \neq j} \Big(Y_i - \sum_{0 \le |\mathbf{k}| \le r} \alpha_{\mathbf{k}} (X_i - x)^{\mathbf{k}}\Big)^2 K\Big(\frac{X_i - x}{h_H}\Big),$$

where $\mathbf{k} = (k_1, k_2, \ldots, k_d)$ is a d-tuple of integers, $|\mathbf{k}| = k_1 + k_2 + \cdots + k_d$, $(X_i - x)^{\mathbf{k}} = (X_i^1 - x^1)^{k_1} \times (X_i^2 - x^2)^{k_2} \times \cdots \times (X_i^d - x^d)^{k_d}$, and $K(x_1, \ldots, x_d) = \Pi_{\ell=1}^d k(x_\ell)$ with $k(\cdot)$ being a univariate kernel function (i.e., a multiplicative kernel is used in the multivariate case). More details of local polynomial regression can be found in Appendix S.1 of SM. The generated regressors are estimated by $\widehat{\zeta}_1\big(X_i^1\big) = \big(1/n\big) \cdot \sum_{j=1}^n \widehat{H}_{-j}\big(X_i^1, X_j^2\big)$ and $\widehat{\zeta}_2\big(X_i^2\big) = \big(1/n\big) \cdot \sum_{j=1}^n \widehat{H}_{-j}\big(X_j^1, X_i^2\big)$ with the weights $w_k(\cdot)$ to be the marginal densities of X^k on $\mathcal{S}_{X^k}$, namely $w_k(\cdot) = p_{X^k}(\cdot)$, for $k = 1,2$.

Finally, the partial derivatives $\partial_k \mathcal{H}(\cdot)$ can be recovered by another local rth-order polynomial estimation, i.e., the slope coefficients of

$$\widehat{\beta} = \arg\min_{\beta} \sum_{i=1}^n \Big(Y_i - \sum_{0 \le k_1 + k_2 \le r} \beta_{k_1,k_2} (\widehat{\zeta}_1\big(X_i^1\big) - z^1)^{k_1} (\widehat{\zeta}_2\big(X_i^2\big) - z^2)^{k_2}\Big)^2 \cdot k\Big(\frac{\widehat{\zeta}_1\big(X_i^1\big) - z^1}{h_{\mathcal{H}}}\Big) k\Big(\frac{\widehat{\zeta}_2\big(X_i^2\big) - z^2}{h_{\mathcal{H}}}\Big).$$

Denote the derivative estimators by $\partial_k \widehat{\mathcal{H}}(z)$ for $k = 1,2$.

[6]Here, r is also the smoothness of unknown functions and densities. See Assumption 3.

https://doi.org/10.1017/S0266466624000318 Published online by Cambridge University Press

Step 2. Estimation of the transformed model. The transformed component functions $\tilde{f}_k(\cdot)$'s are estimated by the sample analogue of (C2) and (C1) as follows:

$$\widehat{\tilde{f}}_1(z^1) = \widehat{c}\int_{z_0^1}^{z^1}\int \frac{\partial_1\widehat{\mathcal{H}}(z)}{\partial_2\widehat{\mathcal{H}}(z)}\omega_4(z^2)dz^2dz^1, \quad \widehat{\tilde{f}}_2(z^2) = \int_{z_0^2}^{z^2}\int \frac{\partial_2\widehat{\mathcal{H}}(z)}{\partial_1\widehat{\mathcal{H}}(z)}\omega_3(z^1)dz^1dz^2,$$

where $\widehat{c} = \int \omega_3(z^1)\left[\int \left[\partial_1\widehat{\mathcal{H}}(z)/\partial_2\widehat{\mathcal{H}}(z)\right]\cdot\omega_4(z^2)dz^2\right]^{-1}dz^1$.

The link function $G(\cdot)$ is then estimated by the intercept of

$$\widehat{\gamma} = \arg\min_{\gamma}\sum_{i=1}^{n}\Big(Y_i - \sum_{0\le k\le r}\gamma_k(\widehat{\tilde{f}}_1(\widehat{Z}_i^1) + \widehat{\tilde{f}}_2(\widehat{Z}_i^2) - \tau)^k\Big)^2 k\Big(\frac{\widehat{\tilde{f}}_1(\widehat{Z}_i^1) + \widehat{\tilde{f}}_2(\widehat{Z}_i^2) - \tau}{h_G}\Big),$$

where $\widehat{Z}_i^k = \widehat{\zeta}_k(X_i^k)$ for $k = 1,2$.

Step 3. Estimation of the original component functions $f_k(\cdot)$'s. Lastly, the original component functions $f_1(\cdot)$ and $f_2(\cdot)$ are estimated by

$$\widehat{f}_k(x^k) = \widehat{\tilde{f}}_k\Big(\widehat{\zeta}_k(x^k)\Big), \quad \text{for } k = 1,2.$$

Three remarks are in order. First, our estimators essentially have similar asymptotic properties to Horowitz's (2001) estimators if the true partial integrals $\zeta_k(\cdot)$'s were used so that the first step is not needed. Second, we use local polynomial regressions instead of local constant ones to address the boundary bias issue (see also Fan and Gijbels, 1992). Third, the step-two estimation of the link $G(\cdot)$ can be viewed as a result of estimating it by a sample analogue of (L2). It can also be viewed as a result of recovering $G(\cdot)$ by a sample analogue of a moment condition of $G(\tau) = E[Y|f_1(X^1) + f_2(X^2) = \tau]$ which comes from (M1) and the law of iterated expectation.

4. LARGE SAMPLE PROPERTIES

In this section, we study the large sample properties of the estimators proposed in Section 3. Let $d_1 \geq d_2$ only for presentation purposes.[7] We first state the assumptions under which the large sample properties of our estimators are established. Let $int(\Theta)$ denote the interior of any given set Θ. Let $\mathcal{S}_W$ be the support of a random vector/variable W, and let $\mathcal{S}_G$ be defined as $\{\tau : \tau = f_1(x^1) + f_2(x^2)$ for some $(x^1,x^2) \in \mathcal{S}_{(X^1,X^2)}\}$.

Assumption 1 (DGP). $\{(Y_i,X_i)\}_{i=1}^n$ is an i.i.d. sample from the distribution of (Y,X) which satisfies (M1) and (i) $E(|Y|^{4+s}|X = x) \leq C$ for some finite C, positive s, and all $x \in \mathcal{S}_X$; (ii) $Var(Y|X = x)$ is continuous in x.

[7] If $d_1 < d_2$, we can define $\bar{x}^1 = x^2$ and $\bar{x}^2 = x^1$. It then follows that $\bar{d}_1 > \bar{d}_2$ where $\bar{d}_k$ denotes the dimension of $\bar{x}^k$ for $k = 1,2$. We then study the new model of $\overline{H}(\bar{x}^1,\bar{x}^2) = G\big(\bar{f}_1(\bar{x}^1) + \bar{f}_2(\bar{x}^2)\big)$ where $\overline{H}(\bar{x}^1,\bar{x}^2) = H(x^1,x^2), \bar{f}_1(\bar{x}^1) = f_2(x^2)$, and $\bar{f}_2(\bar{x}^2) = f_1(x^1)$.

https://doi.org/10.1017/S0266466624000318 Published online by Cambridge University Press

Assumption 2 (Distribution of X). The random vector X satisfies (i) $\mathcal{S}_X$ is compact; (ii) the distribution of X is absolutely continuous with respect to Lebesgue measure and has density of $p_X(\cdot) > 0$ in the interior of $\mathcal{S}_X$; (iii) there exist some compact intervals $\mathcal{I}_1 \subset \text{int}(\mathcal{S}_{Z^1})$, $\mathcal{I}_2 \subset \text{int}(\mathcal{S}_{Z^2})$ and some $\underline{c} > 0$ such that (a) $\tilde{f}_k'(z^k) \geq \underline{c}$ for all $z^k \in \mathcal{I}_k$ and $k = 1,2$, (b) $P(X : Z^k \in \mathcal{I}_k, k = 1,2) > 0$, (c) $z_0^k \in \mathcal{I}_k$ where z_0^k is defined in Assumption I for $k = 1,2$, (d) $|G'(\cdot)| \geq \underline{c}$ on $\mathcal{S}_G$.

Assumption 3 (Smoothness of G, f_k, and p_X). (i) The link function $G(\cdot)$ is $(r+1)$ times continuously differentiable. (ii) The component functions $f_k(\cdot)$ for $k = 1,2$ and density $p_X(\cdot)$ are $(r+1)$ times differentiable with respect to any mixture of its arguments with uniformly continuous derivatives on their supports $\mathcal{S}_{X^k}$ and $\mathcal{S}_X$.

Assumption 4 (Weights). (i) For $k = 1,2$, the weight function $w_k(\cdot) = p_{X^k}(\cdot)$.
(ii) For $k = 3,4$, the weight function $w_k(\cdot)$ is nonnegative and bounded with support $\mathcal{S}_{w_k} \subset \mathcal{I}_{k-2}$ such that $w_k(\cdot)$ has $(r+1)$th continuous derivatives on $\mathcal{S}_{w_k}$ with $\int w_k(z^{k-2})dz^{k-2} = 1$.

Assumption 5 (Kernel). The univariate kernel function $k(\cdot)$ is symmetric, bounded, and continuously differentiable on its support $[-1,1]$ For any $d' \geq 1$ and a kernel function $K(\cdot)$ on $[-1,1]^{d'}$, there is $K(s_1,\ldots,s_{d'}) = \Pi_{j=1}^{d'} k(s_j)$. Let $H_{\mathbf{j}}(u) = u^{\mathbf{j}} K(u)$ for all integers $\mathbf{j} = (j_1, j_2, \ldots, j_d)$ and $u \in \mathbb{R}^d$. Then $H_{\mathbf{j}}(u)$ is Lipschitz continuous on $[-1,1]^d$ for all $\mathbf{j}$ with $0 \leq |\mathbf{j}| \leq 2r+1$.

Assumption 6 (Bandwidth). As $n \to \infty$, the bandwidth sequences h_H, $h_{\mathcal{H}}$, and h_G go to zero and satisfy:

(i) $nh_H^{d+r+1}/log(n) \to \infty$, $nh_{\mathcal{H}}^6/log(n) \to \infty$, $nh_G^3/log(n) \to \infty$,

(ii) $h_H^{d_2}/h_{\mathcal{H}} \to 0$, $log(n)^2/[nh_H^{d_1/2+r+1} h_{\mathcal{H}}^3] \to \gamma_1$, $n \cdot h_H^{d_2} \cdot h_{\mathcal{H}}^{2r} \to \tilde{\gamma}_1$,

(iii) $h_H^{r+1}/h_G \to 0$, $n \cdot h_H^{d_1} \cdot h_G^2 \to \infty$, $nh_G^{2r+3} \to \gamma_2$, $n \cdot h_H^{2r+2} \cdot h_G \to \gamma_3$, $nh_H^{d_2+2r+2} \to \tilde{\gamma}_2$,

(iv) $h_{\mathcal{H}}^r/h_G \to 0$, $h_G/h_{\mathcal{H}} \to \delta_G$, $n \cdot h_{\mathcal{H}}^{2r} \cdot h_G \to \gamma_4$,

where $\gamma_1, \ldots, \gamma_4$, $\tilde{\gamma}_1$, $\tilde{\gamma}_2$, and δ_G are some nonnegative constants.

Assumption 1 describes the model and data generating process (DGP). Assumption 2(i) and (ii) gives some regularity conditions on the support and density function of the random vector X. With the normalization conditions in Assumption I, Assumption 2(iii) provides sufficient conditions to identify the component functions $f_k(\cdot)$'s and the link function $G(\cdot)$.

Assumption 3 contains some smoothness conditions on the link function $G(\cdot)$, the component functions $f_k(\cdot)$'s, and the density function $p_X(\cdot)$. They require those functions having a smoothness of $(r+1)$. This ensures that our Taylor-series expansions have proper orders. In addition, they imply that the transformed component functions $\tilde{f}_k(\cdot)$'s also have $(r+1)$ derivatives which are uniformly continuous on their supports.

Assumption 4 describes the condition on the weight functions $w_k(\cdot)$ for $k = 1,\ldots,4$. For $k = 1,2$, it uses the marginal density of X^k on $\mathcal{S}_{X^k}$ as the weight $w_k(\cdot)$

https://doi.org/10.1017/S0266466624000318 Published online by Cambridge University Press

to estimate the partial integrations $\zeta_k(\cdot)$ in step one of our estimation approach laid out in Section 3. Other weights for $w_1(\cdot)$ and $w_2(\cdot)$ can also be used. For $k=3,4$, it requires the weight function $w_k(\cdot)$ to be $(r+1)$ times continuously differentiable on its support.

Assumption 5 gives the restrictions on the univariate kernel function $k(\cdot)$ which builds all multivariate kernel functions throughout this paper in a multiplicative way. This assumption is also used in other local polynomial literature. See, e.g., Kong, Linton, and Xia (2010) and JLL. This assumption is utilized to derive the uniform asymptotic representation of local polynomial estimators.

Assumption 6 specifies the conditions on the choices of bandwidths used in our kernel estimation. These conditions permit various combinations of bandwidths h_H, $h_{\mathcal{H}}$, and h_G. For example, they are satisfied when $h_H \in \left(n^{-1/(r+1+d)}, n^{-(r+1)/[r\cdot(2r+3)]}\right)$, and $h_{\mathcal{H}} = h_G = n^{-(r+1)/[r\cdot(2r+3)]}$ for large enough r. They ensure that the remainder terms are negligible in each stage of our estimation. In particular, conditions (ii)–(iv) control the contributions from the previous estimation steps to the asymptotic variances of $\widehat{f}_k(\cdot)$ and $\widehat{G}(\cdot)$ for $k=1,2$.

We now present the asymptotic results of our estimators of the component functions $f_k(\cdot)$ and the link function $G(\cdot)$ for $k=1,2$. We first consider the estimation of original component functions $f_k(\cdot)$ for $k=1,2$. Our third theorem gives the asymptotic properties of the estimators $\widehat{f}_k(\cdot)$ for $k=1,2$.

THEOREM 3. *Suppose that Assumptions I and 1–6 hold. Then, for every $k=1,2$, as $n\to\infty$: (i) $\sup_{x^k\in\mathcal{S}_{X^k}}\left|\widehat{f}_k(x^k)-f_k(x^k)\right|\to 0$ in probability, and (ii) for any $x^k\in\mathcal{S}_{X^k}$, $\sqrt{nh_H^{d_k}}\left(\widehat{f}_k\left(x^k\right)-f_k\left(x^k\right)-B_{nf_k}(x^k)\right)\xrightarrow{d} N\left(0,\sigma_k^2(x^k)\right)$ with*

$$B_{nf_k}(x^k) = h_{\mathcal{H}}^r\mathfrak{B}_k(\zeta_k(x^k)) + h_H^{r+1}\left[\tilde{f}_k'(\zeta_k(x^k))D_k(x^k)+\tilde{\mathcal{B}}_k(\zeta_k(x^k))\right], \tag{9}$$

$$\sigma_k^2(x^k) = \tilde{f}_k'(\zeta_k(x^k))^2\cdot\left[\int\left(e_1'S_r^{-1}V_k^{\mu}(t)\right)^2K_k(t)^2dt\right]\int\frac{E\left[\left(Y-H(x)\right)^2|X=x\right]}{p_{X^k|X^{-k}}(x^k|x^{-k})^2}\cdot p_X(x)dx^{-k}, \tag{10}$$

where $\mathfrak{B}_k(\cdot)$, $\tilde{\mathcal{B}}_k(\cdot)$, and $D_k(x^k)$ are given by Appendix A, and e_1, S_r, and $V_k^{\mu}(t)$ are defined in Appendix S.1 of SM.

Theorem 3 establishes the uniform consistency and asymptotic normality of our estimators of original component functions $f_k(\cdot)$ for $k=1,2$. It shows that the only contributions from previous estimation steps are in the resulting biases of $\widehat{f}_k(\cdot)$ for $k=1,2$ in the final step. The variances of previous steps do not contribute into the variances of $\widehat{f}_k(\cdot)$, namely the asymptotic variances of $\widehat{\tilde{f}}_k(\cdot)$ do not enter the ones of $\widehat{f}_k(\cdot)$. In particular, since the estimator can be represented as $\widehat{f}_k(\cdot)=\widehat{\tilde{f}}_k(\widehat{\zeta}_k(\cdot))$, the asymptotic bias term $B_{nf_k}(x^k)$ consists of two parts. The first part $h_{\mathcal{H}}^r\mathfrak{B}_k(\zeta_k(x^k))$ is the bias of the infeasible estimator $\check{\tilde{f}}_k(\zeta_k(\cdot))$ of $f_k(\cdot)$ if the (unobserved) true $\zeta_k(\cdot)$'s were used in all three steps. Specifically, the infeasible estimator $\check{\tilde{f}}_k(\cdot)$ of the transformed component function $\tilde{f}_k(\cdot)$ is obtained by using the true $\zeta_k(\cdot)$'s, instead of their estimators $\widehat{\zeta}_k(\cdot)$'s, to recover $\partial_k\mathcal{H}(\cdot)$ in the first step. The second part $h_H^r \cdot [\tilde{f}_k'(\zeta_k(x^k))D_k(x^k) + \tilde{\mathcal{B}}_k(\zeta_k(x^k))]$ is the additional bias brought by using the estimators $\widehat{\zeta}_k(\cdot)$'s, instead of the true functions $\zeta_k(\cdot)$'s, in all three steps.

Two additional remarks are in order. First, the asymptotic bias terms $B_{nf_k}(x^k)$ for $k = 1,2$ is controllable in general when we use bandwidths satisfying Assumption 6, i.e., $\limsup_{n\to\infty}\sqrt{nh_H^{d_k}}B_{nf_k}(x^k) < \infty$ holds. Second, there are two ways to consistently estimate the asymptotic variances $\sigma_k^2(x^k)$ for $k = 1,2$. The first way exploits the expression of $\sigma_k^2(x^k)$ and replaces its population terms with their nonparametric consistent estimators. The other way is to estimate $\sigma_k^2(x^k)$ by adapting the bootstrap method for nonparametric regression. See, e.g., Härdle and Bowman (1988), Hall (1992), and Hall and Horowitz (2013), among others.

We next consider the estimation of link function $G(\cdot)$. Our next theorem summarizes the large sample properties of our link estimator $\widehat{G}(\cdot)$. Let $\mathcal{S}_G$ be the compact set $\{\tau : \tau = f_1(x^1) + f_2(x^2)$ for some $(x^1,x^2) \in \mathcal{S}_X\}$ where $\mathcal{S}_X$ is the support of X.

THEOREM 4. *Let Assumptions I and 1–6 hold. Then, as* $n \to \infty$*: (i)* $\sup_{\tau\in\mathcal{S}_G}\left|\widehat{G}(\tau) - G(\tau)\right| \to 0$ *in probability, and (ii) for any* $\tau \in \mathcal{S}_G$, $\sqrt{nh_G}\cdot\left(\widehat{G}(\tau) - G(\tau) - B_{nG}(\tau)\right) \xrightarrow{d} N\left(0,\sigma_G^2(\tau)\right)$ *with*

$$B_{nG}(\tau) = h_G^{r+1}e_{1G}'\{\mathcal{S}_r^G\}^{-1}S_r^{G,r+1}G_{r+1}(\tau) - h_H^{r+1}G'(\tau)\sum_{k=1}^{2}E\left[\tilde{f}_k'(\zeta_k(X^k))D_k(X^k) + \tilde{\mathcal{B}}_k(\zeta_k(X^k))\middle|T=\tau\right] - h_{\mathcal{H}}^rG'(\tau)\sum_{k=1}^{2}E\left[\mathfrak{B}_k(\zeta_k(X^k))\middle|T=\tau\right], \quad \textbf{(11)}$$

$$\sigma_G^2(\tau) = \frac{Var(Y|T=\tau)}{p_T(\tau)}\int\left(e_{1G}'\{S_r^G\}^{-1}\mu_G(u)\right)^2k(t)^2dt + \delta_G\cdot\sigma_{G2}^2(\tau), \quad \textbf{(12)}$$

where δ_G *is given by Assumption 6,* $\mathfrak{B}_k(\cdot)$, $\tilde{\mathcal{B}}_k(\cdot)$, $D_k(x^k)$, *and* $\sigma_{G2}^2(\tau)$ *are defined in Appendix A, and* e_{1G}, $\{S_r^G\}$, $S_r^{G,r+1}$, $G_{r+1}(\tau)$, *and* $\mu_G(u)$ *are provided by Appendix S.1 of SM.*

Theorem 4 shows the uniform convergence and asymptotic normality of our kernel estimator of link $G(\cdot)$. Several remarks are in order. First, the asymptotic bias $B_{nG}(\tau)$ consists of three terms. The first term $h_G^{r+1}e_{1G}'\{\mathcal{S}_r^G\}^{-1}S_r^{G,r+1}G_{r+1}(\tau)$ comes from the infeasible estimation of link $G(\cdot)$ when the (unobserved) true $\zeta_k(\cdot)$ and $\tilde{f}_k(\cdot)$ for $k = 1,2$, instead of their estimators, were used in the second step to recover $G(\cdot)$. It is a bias term of a standard nonparametric regression. The other two terms are the additional biases caused by using the feasible estimators $\widehat{\zeta}_k(\cdot)$

and $\widehat{\tilde{f}}_k(\cdot)$ for $k = 1,2$, instead of their true functions, in the second step to estimate $G(\cdot)$. Second, similar to the case of $\widehat{f}_k(\cdot)$'s, the asymptotic bias is controllable under Assumption 6. Third, our asymptotic variance $\sigma_G^2(\tau)$ can be estimated through replacing its population quantities with their consistent estimators.

5. A SIMULATION STUDY

This section demonstrates the finite sample performance of our estimator by some Monte Carlo experiments. We adopt the following DGP with the sample sizes of 400 and 800, each replicated 200 times:

$$Y = 1\{f_1(X^1) + f_2(X_1^2, X_2^2) - U > 0\},$$

where the regressors X^1, X_1^2, and X_2^2 are independent truncated normal on $[-3,3]$ with mean 0 and standard deviation of 2, and the error term U is independent of all regressors and distributed according to standard normal $N(0,1)$. The true link and component functions are specified as

$$G(\tau) = \Phi(\tau), \quad f_1(x^1) = x^1, \quad f_2(x_1^2, x_2^2) = x_1^2 \cdot x_2^2,$$

where $\Phi(\cdot)$ is the distribution function of standard normal.

Two remarks are in order. First, under this specification, the partial integrals are $\zeta_1(x^1) = E[\Phi(x^1 + X_1^2 \cdot X_2^2)]$ and $\zeta_2(x_1^2, x_2^2) = E[\Phi(X^1 + x_1^2 \cdot x_2^2)]$, and the transformed components are the correspondent inverse functions with $\tilde{f}_1(\zeta_1(x^1)) = x^1$ and $\tilde{f}_2(\zeta_2(x_1^2, x_2^2)) = x_1^2 \cdot x_2^2$. Second, the location normalization then requires $z_0^1 = \zeta_1(0)$ and $z_0^2 = \zeta_2(0,0)$ since $\tilde{f}_1(\zeta_1(0)) = 0$ and $\tilde{f}_2(\zeta_2(0,0)) = 0$. The symmetry of distributions of X^1 and X^2 implies that $\zeta_1(0) = \zeta_2(0,0) = \Phi(0) = 0.5$, which is used in the simulation. The scale normalization holds in the model with a constant weight function $w_3(z_1) = \left(\int_{0.3}^{0.7} \left[\tilde{f}_1'(z_1)\right]^{-1} dz_1\right)^{-1} \cdot 1\{0.3 \le z_1 \le 0.7\}$.

We next provide the implementation details of our estimation method. Let $\widehat{\sigma}(W)$ denote the standard error of a given random variable W. To estimate $f_1(\cdot)$, $f_2(\cdot,\cdot)$, and $G(\cdot)$, we use local linear regressions with a second-order Gaussian kernel and the bandwidths of $h_H = \min\left\{\widehat{\sigma}(X^1),\ \widehat{\sigma}(X_1^2),\ \widehat{\sigma}(X_2^2)\right\} \cdot n^{-1/7}$, $h_{\mathcal{H}} = \min\left\{\widehat{\sigma}(\widehat{Z}^1),\ \widehat{\sigma}(\widehat{Z}^2)\right\} \cdot n^{-1/8}$, and $h_G = \widehat{\sigma}\left(\widehat{f}_1(X^1) + \widehat{f}_2(X_1^2, X_2^2)\right) \cdot n^{-1/5}$ following the simplified Silverman's rule of thumb (Silverman, 1986; Hansen, 2009). The weight function $w_4(\cdot)$ is chosen according to $w_4(z^2) = \frac{5}{3} \cdot 1\{0.2 \le z^2 \le 0.8\}$. Meanwhile, we replicate the estimators of JLL (a.k.a. "JLL estimators" in our paper) to do a side-by-side comparison. The details are given as follows. In the estimation of JLL, we also choose the second-order Gaussian kernel to do local linear regressions in all stages, use linear extrapolation to extend the integrand function when we do numerical integration and apply Silverman's rule of thumb to pick the bandwidths. To compute the integrals in our and JLL's estimators, we adopt the midpoint rule to calculate them numerically.

We now show the performance of our estimators and JLL estimators of $f_1(\cdot)$, $f_2(\cdot,\cdot)$, and $G(\cdot)$ to demonstrate how well our estimation procedure can recover the

TABLE 1. Simulation results for the estimation of component function $f_1(x^1)$.

n	x^1	Ours			JLL		
		Bias	SD	RMSE	Bias	SD	RMSE
	−1	0.109	0.162	0.194	0.141	0.249	0.286
400	0	−0.005	0.121	0.121	0.018	0.246	0.246
	1	−0.104	0.161	0.191	−0.148	0.272	0.309
	−1	0.095	0.115	0.149	0.129	0.220	0.254
800	0	0.002	0.101	0.100	−0.007	0.165	0.164
	1	−0.089	0.126	0.154	−0.122	0.185	0.221

TABLE 2. Simulation results for the estimation of component function $f_2(x_1^2, x_2^2)$.

n	x_1^2	x_2^2	Ours			JLL		
			Bias	SD	RMSE	Bias	SD	RMSE
	−1	−1	−0.115	0.237	0.263	−0.107	0.346	0.361
	0	−1	−0.007	0.199	0.199	0.017	0.253	0.253
400	1	−1	0.083	0.239	0.252	0.084	0.315	0.326
	−1	1	0.065	0.251	0.258	0.095	0.288	0.303
	0	1	−0.008	0.201	0.201	−0.006	0.248	0.248
	1	1	−0.078	0.242	0.254	−0.097	0.301	0.315
	−1	−1	−0.103	0.168	0.197	−0.061	0.257	0.263
	0	−1	−0.005	0.151	0.151	−0.028	0.191	0.193
800	1	−1	0.086	0.182	0.201	0.030	0.245	0.247
	−1	1	0.069	0.182	0.194	0.037	0.266	0.268
	0	1	−0.022	0.152	0.153	0.023	0.210	0.210
	1	1	−0.095	0.176	0.200	−0.056	0.267	0.273

component and link functions at different locations. In particular, we report the bias (Bias), the standard deviation (SD), and the root mean square error (RMSE) for all estimators. Table 1 summarizes the simulation results for the estimation of components $f_1(\cdot)$, Table 2 is for $f_2(\cdot,\cdot)$, and Table 3 for the estimation of link $G(\cdot)$. We report in Tables 1–3 the simulation results for ours and JLL estimators at different points in the interior of the support of each function, where the left sections display the results for our estimators and right sections for JLL.

Tables 1 and 2 show the estimation of components $f_1(\cdot)$ and $f_2(\cdot,\cdot)$, respectively. Table 1 shows the performance of our component estimator $\widehat{f_1}(x^1)$ for $x^1 = -1, 0, 1$. They show that our estimator $\widehat{f_1}(\cdot)$ performs reasonably well even under the moderate sample size of 400. When the sample size increases from 400 to 800, the RMSEs of $\widehat{f_1}(\cdot)$ decline significantly. Moreover, the estimation biases are relatively small under both sample sizes of 400 and 800. Table 2 reports the estimation

https://doi.org/10.1017/S0266466624000318 Published online by Cambridge University Press

TABLE 3. Simulation results for the estimation of link function $G(\tau)$.

		Ours			JLL		
n	τ	Bias	SD	RMSE	Bias	SD	RMSE
	−3	0.000	0.006	0.006	0.013	0.039	0.041
400	0	0.000	0.062	0.061	0.002	0.049	0.049
	3	0.001	0.010	0.010	−0.016	0.045	0.047
	−3	0.001	0.005	0.005	0.010	0.027	0.028
800	0	0.002	0.043	0.042	0.001	0.041	0.041
	3	0.000	0.004	0.004	−0.007	0.026	0.026

results for $f_2(x_1^2, x_2^2)$ for all $x_1^2 = -1, 0, 1$ and $x_2^2 = -1, 1$. We first look at the case of $x_2^2 = -1$, which is shown in the upper sections of Table 2. The biases are relatively small under both $n = 400$ and $n = 800$. In addition, the decrease of RMSEs is significant when the sample size increases from 400 to 800. Our estimation of the two-dimensional function $f_2(\cdot,\cdot)$ hence performs reasonably well. We then look at the case of $x_2^2 = 1$ shown in the lower sections of Table 2. Similar to the case of $x_2^2 = -1$, it confirms that (i) the biases are relatively satisfactory under both sample sizes of 400 and 800 and (ii) our estimator becomes closer to its true value as the sample size increases.

Table 3 gives the performance of our link estimator $\widehat{G}(\tau)$ for $\tau = -3, 0, 3$. In general, our link estimator performs relatively well, although it is given by a nonparametric regression with a regressor generated by a two-step nonparametric estimation. The biases are reasonably small under $n = 400$ and $n = 800$. In addition, the RMSEs decrease when the sample size increases from 400 to 800.

Tables 1–3 also compare our results with JLL estimators. We can see that (i) our estimators have smaller variances and RMSEs than JLL estimators with two exceptions in the estimation of $f_1(\cdot), f_2(\cdot)$, and $G(\cdot)$ and (ii) our estimators have biases in a magnitude similar to JLL. Thus, our estimators perform well in finite sample even though we do not require the existence of an univariate component like JLL. We also compare our method with Pinkse (2001) in finite sample before concluding our simulation section. In the context of Pinkse (2001), the above model can be represented as

$$H(x^1, x^2) = \tilde{G}\big(x^1, \tilde{f}_2(x^2)\big),$$

where $H(x^1, x_1^2, x_2^2) = \Phi(x^1 + x_1^2 x_2^2)$, $\tilde{f}_2(x^2) = x_1^2 x_2^2$, and $\tilde{G}(x^1, t) = \Phi(x^1 + t)$. The estimation of $H(\cdot)$ is the object of comparison here. To implement his approach, we apply local linear method for the first-step estimation and a weighted local constant regression for the second-step estimation. The second-step estimation closely follows the definition of his estimator. We also use the second-order Gaussian kernel and bandwidths following the rule of thumb. The weight is chosen according to the simulation study of Pinkse (2001). We report the simulation results of his third estimator, namely S_π, here.

Table 4 shows the simulation results for ours, JLL, and Pinkse's estimators of the overall function $H(x^1, x_1^2, x_2^2)$ under sample sizes 400 and 800. The comparison shows that (i) the RMSEs of our estimator decline significantly when the sample size increases from 400 to 800; (ii) our estimator has smaller variances and RMSEs than both of JLL and Pinkse's (2001) estimators in most cases; and (iii) our biases are comparable to the best ones between JLL and Pinkse (2001).

6. EXTENSIONS

6.1. Discrete Covariates

We now turn to the case with discrete covariates in (x^1, x^2). Let $X^k = (X_d^k, X_c^k)$ with discrete regressors $X_d^k \in \mathbb{R}^{a_k}$ ($a_k \geq 1$) and continuous regressors $X_c^k \in \mathbb{R}^{b_k}$ ($b_k \geq 1$) for $k = 1, 2$.

With mixed data of discrete and continuous regressors, our transformation and identification results, namely Theorems 1 and 2, still hold under proper choices of weight functions $w_k(\cdot)$ for $k = 1, \ldots, 4$ and proper definition of integration with respect to discrete variables. We follow Li and Racine (2007) to accommodate both discrete and continuous regressors in our estimation. We mainly need to modify the kernel regression estimators of $\hat{\zeta}_k(x^k)$ for $k = 1, 2$ and $\widehat{H}_{-j}(x)$ in Step 1 (outlined in Section 3) as follows:

$$\widehat{\zeta}_1(x^1) = \frac{1}{n}\sum_{j=1}^{n}\widehat{H}_{-j}\big(x^1, X_j^2\big), \quad \widehat{\zeta}_2(x^2) = \frac{1}{n}\sum_{j=1}^{n}\widehat{H}_{-j}\big(X_j^1, x^2\big),$$

where $\widehat{H}_{-j}\big(x\big)$ comes from the intercept of a leave-one-out local polynomial estimation

$$\min_{\alpha}\sum_{i\neq j}\Big(Y_i - \sum_{0\leq|\mathbf{k}|\leq r}\alpha_{\mathbf{k}}(X_{ci} - x_c)^{\mathbf{k}}\Big)^2 K_{h_H,\lambda}(x, X_i),$$

where $\mathbf{k} = (k_1, k_2, \ldots, k_{b_1+b_2})$, $K_{h_H,\lambda}(x, X) = \Pi_{\ell=1}^{b_1} k\big(\frac{x_{c\ell}^1 - X_{c\ell}^1}{h_H}\big) \cdot \Pi_{\ell=1}^{b_2} k\big(\frac{x_{c\ell}^2 - X_{c\ell}^2}{h_H}\big) \cdot \Pi_{\ell=1}^{a_1} \lambda_{1\ell}^{N_\ell^1(x,X)} \cdot \Pi_{\ell=1}^{a_2} \lambda_{2\ell}^{N_\ell^2(x,X)}$, and $N_\ell^k(x, X) = 1\{X_{d\ell}^k \neq x_{d\ell}^k\}$. Here, we use a multiplicative kernel function for the multivariate regressors. A univariate kernel of $k(\cdot)$ is adopted for the continuous regressors, and another univariate kernel of $l(X_{d\ell}^k, x_{d\ell}^k, \lambda_{k\ell}) = \lambda_{k\ell}^{1\{X_{d\ell}^k \neq x_{d\ell}^k\}}$ is employed for the discrete (and unordered) regressors with a bandwidth $\lambda_{k\ell} \in [0, 1]$.[8]

With the above adaption in our estimation (to accommodate the mixed data of discrete and continuous regressors), we can obtain the large sample properties of

[8] If the discrete regressors are ordered, then a univariate kernel of $l(X_{d\ell}^k, x_{d\ell}^k, \lambda_{k\ell}) = \lambda_{k\ell}^{|X_{d\ell}^k - x_{d\ell}^k|}$ can be applied in this case. See Li and Racine (2007) for more details.

https://doi.org/10.1017/S0266466624000318 Published online by Cambridge University Press

TABLE 4. Simulation results for estimating original regression function $H(x^1, x_1^2, x_2^2)$.

x^1	x_1^2	x_2^2	Ours Bias	Ours SD	Ours RMSE	JLL Bias	JLL SD	JLL RMSE	Pinkse (2001) Bias	Pinkse (2001) SD	Pinkse (2001) RMSE
						$n = 400$					
−1	−1	−1	−0.004	0.095	0.095	0.012	0.109	0.109	−0.012	0.105	0.106
0	−1	−1	−0.046	0.076	0.089	−0.081	0.102	0.130	−0.075	0.071	0.103
1	−1	−1	−0.024	0.036	0.043	−0.073	0.082	0.110	−0.049	0.034	0.060
−1	0	−1	0.039	0.073	0.083	0.101	0.095	0.138	0.052	0.062	0.081
0	0	−1	−0.004	0.084	0.084	0.014	0.106	0.106	0.003	0.079	0.079
1	0	−1	−0.045	0.070	0.083	−0.092	0.100	0.135	−0.057	0.061	0.083
−1	1	−1	0.020	0.039	0.044	0.068	0.065	0.094	0.050	0.035	0.061
0	1	−1	0.035	0.075	0.082	0.093	0.097	0.134	0.078	0.073	0.107
1	1	−1	−0.010	0.093	0.094	−0.010	0.108	0.108	0.008	0.107	0.107
−1	−1	1	0.019	0.038	0.042	0.068	0.063	0.093	0.050	0.033	0.060
0	−1	1	0.031	0.072	0.078	0.095	0.088	0.129	0.078	0.075	0.109
1	−1	1	−0.016	0.098	0.099	−0.007	0.092	0.092	0.007	0.100	0.100
−1	0	1	0.040	0.072	0.082	0.096	0.089	0.131	0.054	0.060	0.081
0	0	1	−0.004	0.087	0.087	0.008	0.096	0.096	0.004	0.078	0.078
1	0	1	−0.046	0.074	0.087	−0.095	0.095	0.134	−0.055	0.058	0.080
−1	1	1	0.011	0.092	0.093	0.011	0.101	0.101	−0.016	0.115	0.116
0	1	1	−0.036	0.079	0.086	−0.082	0.092	0.122	−0.078	0.075	0.108
1	1	1	−0.021	0.038	0.043	−0.075	0.079	0.108	0.051	0.036	0.062
						$n = 800$					
−1	−1	−1	0.001	0.070	0.070	0.020	0.092	0.094	0.000	0.088	0.088
0	−1	−1	−0.034	0.059	0.068	−0.066	0.081	0.105	−0.065	0.058	0.088
1	−1	−1	−0.018	0.032	0.037	−0.048	0.048	0.048	0.042	0.026	0.049
−1	0	−1	0.036	0.060	0.069	0.072	0.085	0.111	0.056	0.049	0.074
0	0	−1	−0.001	0.065	0.064	−0.008	0.090	0.090	0.003	0.065	0.065
1	0	−1	−0.033	0.058	0.066	−0.083	0.075	0.112	−0.048	0.048	0.068
−1	1	−1	0.018	0.032	0.037	0.050	0.066	0.083	0.044	0.026	0.052
0	1	−1	0.034	0.059	0.068	0.057	0.078	0.096	0.068	0.060	0.091
1	1	−1	0.001	0.075	0.074	−0.026	0.082	0.085	0.005	0.081	0.081
−1	−1	1	0.016	0.028	0.032	0.051	0.064	0.082	0.046	0.028	0.054
0	−1	1	0.031	0.061	0.069	0.061	0.078	0.099	0.071	0.061	0.093
1	−1	1	−0.005	0.077	0.077	−0.021	0.086	0.078	0.008	0.085	0.085
−1	0	1	0.031	0.055	0.063	0.085	0.090	0.123	0.057	0.045	0.072
0	0	1	−0.007	0.068	0.068	0.006	0.089	0.089	0.005	0.066	0.066
1	0	1	−0.038	0.059	0.070	−0.072	0.073	0.103	−0.047	0.049	0.067
−1	1	1	0.003	0.073	0.072	0.025	0.098	0.101	−0.004	0.088	0.088
0	1	1	−0.033	0.061	0.069	−0.064	0.084	0.106	−0.068	0.059	0.090
1	1	1	−0.018	0.032	0.032	−0.049	0.052	0.072	−0.043	0.025	0.050

https://doi.org/10.1017/S0266466624000318 Published online by Cambridge University Press

$\widehat{f}_k(\cdot)$ for $k = 1, 2$ and $\widehat{G}(\cdot)$ similar to those summarized by Theorems 3 and 4. In particular, the asymptotic variance of $\widehat{f}_k(\cdot)$ has an order of $O\big(1/(nh_H^{b_k})\big)$ instead of $O\big(1/(nh_H^{d_k})\big)$ where $b_k < d_k$.

6.2. Multiple Component Functions

We next briefly discuss the extension of our method from the baseline model with two components to the case with more than two components.

Let $K > 2$. For any $k = 2, \ldots, K$, let $H_k(x^1, x^k) = \int H(x) \cdot p_{\tilde{X}^{-k}}(\tilde{x}^{-k})\, d\tilde{x}^{-k}$ where $\tilde{X}^{-k}$ is obtained by excluding X^1 and X^k from X, and $\tilde{x}^{-k}$ is obtained by excluding x^1 and x^k from x. This constructed $H_k(x^1, x^k)$ is identified if the original $H(x)$ is identified. We can transform the original model (2) with K components into the following new model with two components as

$$H_k(x^1, x^k) = G_k(f_1(x^1) + f_k(x^k)), \tag{13}$$

where $G_k(\tau) = \int G\big(\tau - f_k(x^k) + \sum_{\ell=2}^{K} f_\ell(x^\ell)\big) \cdot p_{\tilde{X}^{-k}}(\tilde{x}^{-k}) d\tilde{x}^{-k}$ is monotonic if the original link function $G(\cdot)$ is monotonic.

Our previous idea can be applied directly to the new model (13) to identify $f_1(\cdot)$ and $f_k(\cdot)$. Specifically, for any $k = 2, \ldots, K$, we use an idea similar to Theorem 1 to transform the new model (13) into the following model with two univariate components:

$$\mathcal{H}_k(z^1, z^k) = G_k(\tilde{f}_1(z^1) + \tilde{f}_k(z^k)), \tag{14}$$

where $\mathcal{H}_k(z^1, z^k) = E[H_k(X^1, X^k) | \zeta_1(X^1) = z^1, \zeta_k(X^k) = z^k]$, the inverse of $\tilde{f}_\ell(\cdot)$ is $\tilde{f}_\ell^{-1}(s) = \int G_k(s + f_{-\ell}(x^{-\ell})) \cdot w_{-\ell}(x^{-\ell}) dx^{-\ell}$, and $\zeta_\ell(x^\ell) = \int H_k(x^1, x^k) \cdot w_{-\ell}(x^{-\ell}) dx^{-\ell}$ with freely chosen weight functions $w_{-\ell}(\cdot)$ for $\ell = 1, k$ where $x^{-\ell}$ is x^k if $\ell = 1$ and is x^1 if $\ell = k$. The transformed components $\tilde{f}_1(\cdot)$ and $\tilde{f}_k(\cdot)$ can then be identified by (C2) and (C1), respectively, where $\mathcal{H}(\cdot)$ is replaced by $\mathcal{H}_k(\cdot)$. The original components are identified as $f_k(x^k) = \tilde{f}_k(\zeta_k(x^k))$ for all $k = 1, \ldots, K$. Once all of $f_k(\cdot)$, $k = 1, \ldots, K$, are identified, the original link $G(\cdot)$ is identified by $G(\tau) = E\big[H(X) \big| \sum_{k=1}^{K} f_k(X^k) = \tau\big]$. Similar to the case with two components (i.e., $K = 2$), we can closely follow the above identification strategy to estimate the link $G(\cdot)$ and the components $f_\ell(\cdot)$ in three steps for $\ell = 1, \ldots, K$. Let $k = 2, \ldots, K$. In the first step, we estimate the transformed function $\mathcal{H}_k(z^1, z^k)$ by the nonparametric sample analogue of its definition as $\widehat{E}[H_k(X^1, X^k) | \widehat{\zeta}_1(X^1) = z^1, \widehat{\zeta}_k(X^k) = z^k]$, where $H_k(x^1, x^k) = \int H(x) \cdot p_{\tilde{X}^{-k}}(\tilde{x}^{-k})\, d\tilde{x}^{-k}$ and $\widehat{\zeta}_\ell(X^\ell)$'s are also given by the sample analogues of $\zeta_\ell(X^\ell) = \int H_k(X^1, X^k) \cdot w_{-\ell}(X^{-\ell}) dX^{-\ell}$ for $\ell = 1, k$. Given the first-step estimator $\widehat{\mathcal{H}}_k(z^1, z^k)$, the second step follows Horowitz's (2001) estimation procedure to estimate the transformed components $\tilde{f}_1(\cdot)$ and $\tilde{f}_k(\cdot)$ according to (C2) and (C1), respectively, with $\mathcal{H}(\cdot)$ replaced by $\mathcal{H}_k(\cdot)$ in the transformed model (14). Moreover, the link $G(\cdot)$ is recovered by $\widehat{G}(\tau) = \widehat{E}\big[Y \big| \sum_{\ell=1}^{K} \widehat{\tilde{f}}_\ell(\widehat{\zeta}_\ell(X^\ell)) = \tau\big]$. In step three, the original components are then recovered by $\widehat{f}_\ell(\cdot) = \widehat{\tilde{f}}_\ell(\widehat{\zeta}_\ell(\cdot))$ for

$\ell = 1,\dots,K$. Note that we will obtain $(K-1)$ estimates for the first component $f_1(\cdot)$. We hence aggregate them by their average to estimate $f_1(\cdot)$.

7. CONCLUSION

In this paper, we consider estimating the generalized additive model with a flexible grouping and an unknown link function. To identify the model primitives, we transform the model into a new model with univariate components. We then identify the new model by applying the existing strategy for the generalized additive model with univariate components. Closely following the identification strategy, we propose a three-step procedure to estimate the link and original components. The consistency and asymptotic normality are then established for the link estimator at a one-dimensional convergence rate and for the component estimators at the convergence rates corresponding to their own dimensions.

This paper adopts a multi-step kernel method to estimate the component and link functions in the generalized additive model with a flexible additive structure and unknown link. Hahn, Liao, and Ridder (2018) studied nonparametric two-step sieve M estimation in a general class of semi/nonparametric models. As the sieve method is convenient to implement in practice, it is interesting to use a multi-step sieve method to estimate the component and link functions in our framework. This is an interesting topic for future research.

Appendix

Appendix A defines some key notation on convergence rates, bias, and variance terms. Appendix B proves the theorems given in the text. Appendix S.1 of SM introduces some additional notation on local polynomial regression for the convenience of discussion in the text and proofs. All technical lemmas are stated and shown in Appendix S.2 of SM.

A. Key Notation on Convergence Rates, Bias, and Variance Terms

For $k = 1,2,$

$$\sigma^2_{G2}(\tau) = G'(\tau)^2 \sum_{k=1}^{2} \int c^{2(2-k)} \omega_{5-k}(Z_i^{-k})^2 \cdot \left\{ \int \left(q_k(\mathcal{Z}^0_{ki})' e'_d \tilde{S}_r^{-1} V_k^{\tilde{\mu}}(t)\right)^2 \mathcal{K}_k(t)^2 dt \right\} \cdot \frac{\mathrm{Var}(Y|Z = \mathcal{Z}^0_{ki})}{p_Z(\mathcal{Z}^0_{ki})}\, dZ_i^{-k}, \tag{A.1}$$

$$J_{nk}(x^k) = \frac{1}{nh_H^{d_k}} \sum_{i=1}^{n} K_k\left(\frac{X_i^k - x^k}{h_H}\right) \cdot \frac{Y_i - H(x_i)}{p_{X^k|X^{-k}}(x^k|X_i^{-k})} e'_1 S_r^{-1} V_k^{\mu}\left(\frac{X_i^k - x^k}{h_H}\right), \tag{A.2}$$

$$D_k(x^k) = e'_1 S_r^{-1} S_r^{r+1} \int H_{r+1}(x^k, x^{-k}) p_{X^{-k}}(x^{-k}) dx^{-k}, \tag{A.3}$$

https://doi.org/10.1017/S0266466624000318 Published online by Cambridge University Press

$$\tilde{\mathfrak{I}}_{nk}(z^k) = \frac{c^{2-k}}{nh_{\mathcal{H}}} \sum_{i=1}^{n} \omega_{5-k}(Z_i^{-k}) \Big[q_k(\mathcal{Z}_{ki})' e_d' \tilde{S}_r^{-1} \frac{Y_i - H(X_i)}{p_Z(\mathcal{Z}_{ki})} V_k^{\tilde{\mu}} \Big(\frac{\zeta_k(X_i^k) - z^k}{h_{\mathcal{H}}} \Big) \mathcal{K}_k \big(\frac{z^k - \zeta_k(X_i^k)}{h_{\mathcal{H}}} \big)$$

$$- q_k(\mathcal{Z}_{ki}^0)' e_d' \tilde{S}_r^{-1} \frac{Y_i - H(X_i)}{p_Z(\mathcal{Z}_{ki}^0)} V_k^{\tilde{\mu}} \Big(\frac{\zeta_k(X_i^k) - z_0^k}{h_{\mathcal{H}}} \Big) \mathcal{K}_k \big(\frac{z_0^k - \zeta_k(X_i^k)}{h_{\mathcal{H}}} \big) \Big]$$

$$\mathfrak{B}_k(z^k) = c^{2-k} \int_{z_0^k}^{z^k} \int q_k(\nu)' \mathfrak{D}(\nu) \omega_{5-k}(\nu^{-k}) d\nu^{-k} d\nu^k,$$

$$\tilde{\mathcal{B}}_k(z^k) = c^{2-k} \int_{z_0^k}^{z^k} \int q_k(\nu)' \mathcal{D}(\nu) \omega_{5-k}(\nu^{-k}) d\nu^{-k} d\nu^k,$$

$$\mathfrak{D}(z) = e_d' \tilde{S}_r^{-1} \tilde{S}_r^{r+1} \mathcal{H}_{r+1}(z),$$

$$\mathcal{D}(z) = \tilde{e}_1 \mathcal{D}_1(z) + \tilde{e}_2 \mathcal{D}_2(z),$$

where $c = \int \omega_3(z^1) \cdot \Big[\int [\partial_1 \mathcal{H}(z) / \partial_2 \mathcal{H}(z)] \cdot \omega_4(z^2) dz^2 \Big]^{-1} dz^1$, $V_k^{\mu}(u^k) = \int \mu(u^k, t^{-k}) K_{-k} (t^{-k}) dt^{-k}$, $V_k^{\tilde{\mu}}(\tilde{u}^k) = \int \tilde{\mu}(\tilde{u}^k, \tilde{t}^{-k}) k_{-k}(\tilde{t}^{-k}) d\tilde{t}^{-k}$, $T = f_1(X^1) + f_2(X^2)$, $\mathcal{X}_{1i} = (x^1, X_i^2)$, $\mathcal{X}_{2i} = (X_i^1, x^2)$, $\mathcal{Z}_{1i} = (z^1, Z_i^2)$, $\mathcal{Z}_{2i} = (Z_i^1, z^2)$, $\mathcal{Z}_{1i}^0 = (z_0^1, Z_i^2)$, $\mathcal{Z}_{2i}^0 = (Z_i^1, z_0^2)$, $\mathcal{K}_K(u^k) = \int_{-\infty}^{u^k} k_k(t^k) dt^k$, x_s^k denotes the sth element of x^k, $e_1 = (1, 0, 0, \dots, 0)'$ is an $N_r \times 1$ vector, $e_{1G} = (1, 0, 0, \dots, 0)'$ is an $(r+1) \times 1$ vector, $\tilde{e}_1 = (1, 0)'$, $\tilde{e}_2 = (0, 1)'$, $q_2(\nu) = \Big[-\frac{\partial_2 \mathcal{H}(\nu)}{[\partial_1 \mathcal{H}(\nu)]^2}, \frac{1}{\partial_1 \mathcal{H}(\nu)} \Big]'$, $q_1(\nu) = \Big[\frac{1}{\partial_2 \mathcal{H}(\nu)}, -\frac{\partial_1 \mathcal{H}(\nu)}{[\partial_2 \mathcal{H}(\nu)]^2} \Big]'$, and

$$\mathcal{D}_k(z) = -p_Z(z)^{-1} \frac{\partial}{\partial z^k} \Big\{ \sum_{\ell=1}^{2} \frac{\partial}{\partial z^\ell} \mathcal{H}(z) \int D_\ell(x^\ell) p_{X^\ell|Z}(x^\ell | z) dx^\ell \cdot p_Z(z) \Big\},$$

where $D_k(x^k)$'s are given by (A.3), respectively.

Furthermore, let $\xi_H = h_H^{r+1} + \sqrt{log(n)/(nh_H^d)}$, $\xi_{\mathcal{H}} = h_{\mathcal{H}}^{r+1} + \sqrt{log(n)/(nh_{\mathcal{H}}^2)}$, $\xi_{\mathcal{H}}' = h_{\mathcal{H}}^r + \sqrt{log(n)/(nh_{\mathcal{H}}^4)}$, and $\xi_{Hk} = h_H^{r+1} + \sqrt{log(n)/(nh_H^{d_k})}$ for $k = 1, 2$. Let $\mathcal{S}_Z$ be a compact set range of $\{(z^1, z^2) : z^1 = \zeta_1(x^1)$ and $z^2 = \zeta_2(x^2)$ for some $(x^1, x^2) \in \mathcal{S}_X\}$, and let $\mathcal{S}_{Z^k}$ be a compact set range of $\{z^k : z^k = \zeta_k(x^k)$ for some $x^k \in \mathcal{S}_{X^k}\}$ for $k = 1, 2$.

B. Proofs of Theorems

B.1. Proof of Theorem 1

Proof. By definition, for $k = 1, 2$, we have

$$\zeta_k(x^k) = \int H(x) w_{-k}(x^{-k}) dx^{-k} = \int G\big(f_1(x^1) + f_2(x^2)\big) w_{-k}(x^{-k}) dx^{-k} = \delta_k\big(f_k(x^k)\big), \quad \textbf{(B.1)}$$

where the second equality comes from the model restriction (M1). Here, the dependence of $\delta_k(\cdot)$ on the function $f_{-k}(\cdot)$ is abbreviated for simplicity of notation. It is easy to verify that $\delta_k(\cdot)$ is strictly monotonic and hence has an inverse function $\delta_k^{-1}(\cdot)$ if $G(\cdot)$ is strictly monotonic. Thus, $f_k(x^k) = \delta_k^{-1}(\zeta_k(x^k))$. Because $\mathcal{H}(z) = E[H(X) | \zeta_1(X^1) = z^1, \zeta_2(X^2) = z^2]$

by definition, it follows that

$$\begin{aligned}\mathcal{H}(z) &= E\left[G\big(f_1(X^1)+f_2(X^2)\big)\Big|\zeta_1(X^1)=z^1,\zeta_2(X^2)=z^2\right]\\ &= E\left[G\big(\delta_1^{-1}(\zeta_1(X^1))+\delta_2^{-1}(\zeta_2(X^2))\big)\Big|\zeta_1(X^1)=z^1,\zeta_2(X^2)=z^2\right]\\ &= G\big(\delta_1^{-1}(z^1)+\delta_2^{-1}(z^2)\big).\end{aligned}$$

The desired conclusion is therefore established by letting $\tilde{f}_k(z^k)=\delta_k^{-1}(z^k)$ for $k=1,2$. □

B.2. Proof of Theorem 3

Proof. Only the case for $k=2$ is proved. The proof for $k=1$ is similar. The definition of $\widehat{f}_2(x^2)$ gives the following decomposition:

$$\widehat{f}_2(x^2)-f_2(x^2)=\big[\widehat{\tilde{f}}_2\big(\widehat{\zeta}_2(x^2)\big)-\tilde{f}_2\big(\widehat{\zeta}_2(x^2)\big)\big]+\big[\tilde{f}_2\big(\widehat{\zeta}_2(x^2)\big)-\tilde{f}_2\big(\zeta_2(x^2)\big)\big], \tag{B.2}$$

where both terms on the right-hand side of equality converge to 0 uniformly over $x^2\in\mathcal{S}_{X^2}$ in probability by SM Lemmas S.3 and S.6. Part (i) is hence established. The rest of the proof is to show part (ii). The first term on the right-hand side of (B.2) can be simplified as

$$\widehat{\tilde{f}}_2\big(\widehat{\zeta}_2(x^2)\big)-\tilde{f}_2\big(\widehat{\zeta}_2(x^2)\big)=\widehat{\tilde{f}}_2\big(\zeta_2(x^2)\big)-\tilde{f}_2\big(\zeta_2(x^2)\big)+O_p\big(\xi_{H2}(\xi'_{\mathcal{H}}+\xi_{H1})\big), \tag{B.3}$$

uniformly over x^2 as $n\to\infty$, where the third (remaining) term on the right-hand side is due to $\int_{\zeta_2(x^2)}^{\widehat{\zeta}_2(x^2)}\int\big[\frac{\partial_2\widehat{\mathcal{H}}(z)}{\partial_1\widehat{\mathcal{H}}(z)}-\frac{\partial_2\mathcal{H}(z)}{\partial_1\mathcal{H}(z)}\big]w_3(z^1)dz^1dz^2=O_p\big(\xi_{H2}(\xi'_{\mathcal{H}}+\xi_{H1})\big)$, which is derived by applying a Taylor expansion similar to (17) of SM Lemma S.6 on the (unweighted) integrand and SM Lemmas S.3 and S.5. Take a Taylor expansion to the second term on the right-hand side of (B.2) to obtain

$$\tilde{f}_2\big(\widehat{\zeta}_2(x^2)\big)-\tilde{f}_2\big(\zeta_2(x^2)\big)=\tilde{f}'_2\big(\zeta_2(x^2)\big)\big(\widehat{\zeta}_2(x^2)-\zeta_2(x^2)\big)+O_p\Big(\xi_{H2}^2\Big),$$

uniformly over x^2 as $n\to\infty$. Consequently, with bandwidths satisfying Assumption 6, the asymptotic representations of $\widehat{\tilde{f}}_2(\cdot)$ given by SM Lemma S.6, and $\widehat{\zeta}_2(\cdot)$ given by SM Lemma S.3 imply that

$$\begin{aligned}\widehat{f}_2(x^2)-f_2(x^2) = {} & \tilde{f}'_2(\zeta_2(x^2))\cdot J_{n2}(x^2)+\tilde{\mathfrak{J}}_{n2}(\zeta_2(x^2))-E[\tilde{\mathfrak{J}}_{n2}(\zeta_2(x^2))]\\ &+h_H^{r+1}\big[\tilde{f}'_2(\zeta_2(x^2))D_2(x^2)+\tilde{\mathcal{B}}_2(\zeta_2(x^2))\big]+h_{\mathcal{H}}^r\mathfrak{B}_2(\zeta_2(x^2))\\ &+o_p(h_{\mathcal{H}}^r+h_H^{r+1}),\end{aligned} \tag{B.4}$$

uniformly over x^2 as $n\to\infty$. The asymptotic normality of part (ii) then follows by applying the Lindeberg–Feller central limit theorem (see Theorem 7.2.1 of Chung, 2001) to (B.4). The asymptotic bias is an immediate consequence of (B.4), and the asymptotic variance $\mathrm{Var}\Big(\sqrt{nh_H^{d_2}}\cdot\Big[\tilde{f}'_2(\zeta_2(x^2))J_{n2}(x^2)+\tilde{\mathfrak{J}}_{n2}(\zeta_2(x^2))\Big]\Big)=\sigma_2^2(x^2)+o(1)$ is obtained by a calculation similar to the one of asymptotic variance of a kernel density estimator. This completes the whole proof. □

https://doi.org/10.1017/S0266466624000318 Published online by Cambridge University Press

B.3. Proof of Theorem 4

Proof. For any $i = 1,\dots,n$, let $T = T(x) = f_1(X^1) + f_2(X^2)$, $T_i = T(X_i) = f_1(X_i^1) + f_2(X_i^2)$, $\widehat{T}_i = \widehat{T}(x_i) = \widehat{f}_{1i}(X_i^1) + \widehat{f}_{2i}(X_i^2)$, and $p_T(\cdot)$ be the probability density function of T, where $\widehat{f}_{ki}(\cdot)$ is the estimator of $f_k(\cdot)$ leaving observation i out for $k = 1,2$. Since we have $\sup_{x\in\mathcal{S}_X}\left|\widehat{T}(x) - T(x)\right|$ similar to SM Lemma S.4, we can derive the asymptotic representation for any $\tau \in \mathcal{S}_G$,

$$\begin{aligned}
&\widehat{G}(\tau) - G(\tau)\\
&= \frac{1}{nh_G}\sum_{i=1}^{n} e_{1G}'\mathcal{S}_{n,r}^{G}(\tau)^{-1} k\Big(\frac{T_i-\tau}{h_G}\Big)\Big\{Y_i - \mu_G(T_i-\tau)'\beta_G(\tau)\Big\}\mu_G\Big(\frac{T_i-\tau}{h_G}\Big) + \frac{1}{nh_G^2}\sum_{i=1}^{n} e_{1G}'\mathcal{S}_{n,r}^{G}(\tau)^{-1}\cdot\\
&\qquad \Big[\Big(\frac{\partial}{\partial u}t_G(u,Y_i;\tau)k(u) + t_G(u,Y_i;\tau)k'(u)\Big)\Big|_{u=\frac{T_i-\tau}{h_G}}\Big]\cdot\big(\widehat{T}(X_i) - T_i\big) + o_p\Big(h_G^{r+1} + \sqrt{log(n)/(nh_G)}\Big)\\
&=: \Gamma_{1n}(\tau) + \Gamma_{2n}(\tau) + o_p\Big(h_G^{r+1} + \sqrt{log(n)/(nh_G)}\Big)
\end{aligned} \tag{B.5}$$

uniformly over $\tau \in \mathcal{S}_G$ as $n \to \infty$, where $t_G(u,Y_i;\tau) = \mu_G(u)\big(Y_i - \mu(u)'B_{h_G}\beta_G(\tau)\big)$. The first term $\Gamma_{1n}(\tau)$ is the uniform Bahadur representation for local polynomial regression in Kong, Linton, and Xia (2010). The second term $\Gamma_{2n}(\tau)$ represents the error caused by using generated regressor $\widehat{T}_i$. Thus, we have the uniform convergence of $\sup_{\tau\in\mathcal{S}_G}\left|\widehat{G}(\tau) - G(\tau)\right|$ and thus part (i) is proved. Similar to SM Lemma S.5, $\Gamma_{1n}(\tau)$ can be decomposed into a bias leading term and a stochastic leading term, i.e.,[9]

$$\Gamma_{1n}(\tau) = \frac{1}{nh_G}\sum_{i=1}^{n} e_{1G}'\{\mathcal{S}_r^G\}^{-1}\frac{Y_i - G(T_i)}{p_T(\tau)}\mu_G\Big(\frac{T_i-\tau}{h_G}\Big)k\Big(\frac{T_i-\tau}{h_G}\Big) + B_0(\tau) + R_{Gn}, \tag{B.6}$$

where $R_{Gn} = o_p\big(h_G^{r+1} + \sqrt{1/(nh_G)}\big)$, and $B_0(\tau) = e_{1G}'\{\mathcal{S}_r^G\}^{-1}S_r^{G,r+1}G_{r+1}(\tau)\cdot h_G^{r+1}$. As for $\Gamma_{2n}(\tau)$, we can further decompose as under Assumption 6,

$$\begin{aligned}
\Gamma_{2n}(\tau) &= \frac{1}{nh_G^2}\sum_{i=1}^{n} e_{1G}'\{\mathcal{S}_r^G\}^{-1}p_T(\tau)^{-1}B_1(T_i,Y_i,\tau)\cdot\Big(E[\widehat{T}(X_i)|X_i] - T_i\Big) + \frac{1}{nh_G^2}\sum_{i=1}^{n} e_{1G}'\{\mathcal{S}_r^G\}^{-1}\cdot\\
&\qquad p_T(\tau)^{-1}B_1(T_i,Y_i,\tau)\cdot\Big(\widehat{T}(X_i) - E[\widehat{T}(X_i)|X_i]\Big) + o_p\Big(h_G^{r+1} + h_H^{r+1} + h_{\mathcal{H}}^r + \sqrt{1/(nh_G)}\Big)\\
&=: \Gamma_{21n}(\tau) + \Gamma_{22n}(\tau) + o_p\Big(h_G^{r+1} + h_H^{r+1} + h_{\mathcal{H}}^r + \sqrt{1/(nh_G)}\Big),
\end{aligned} \tag{B.7}$$

where $B_1(T_i,Y_i,\tau) = \Big(\frac{\partial}{\partial u}t_G(u,Y_i;\tau)k(u) + t_G(u,Y_i;\tau)k'(u)\Big)\Big|_{u=\frac{T_i-\tau}{h_G}}$, $E[\widehat{T}(X_i)|X_i] - T_i = \sum_{k=1}^{2}B_{nfk}(X_i^k) = \sum_{k=1}^{2}\big\{h_{\mathcal{H}}^r\mathfrak{B}_k(\zeta_k(X_i^k)) + h_H^{r+1}\big[\tilde{f}_k'(\zeta_k(X_i^k))D_k(X_i^k) + \tilde{\mathcal{B}}_k(\zeta_k(X_i^k))\big]\big\}$, and $\widehat{T}(X_i) - E[\widehat{T}(X_i)|X_i] = \sum_{k=1}^{2}\big(\tilde{f}_k'(\zeta_k(X_i^k))\cdot J_{nk}(X_i^k) + \tilde{\mathfrak{J}}_{nk}(\zeta_k(X_i^k)) - E[\tilde{\mathfrak{J}}_{nk}(\zeta_k(X_i^k))|X_i]\big)$. $\Gamma_{21n}(\tau)$ is the additional bias due to the generated regressor $\widehat{T}(X_i)$. Similar to the arguments in (13) of SM Lemma S.5, we get

$$\begin{aligned}
\Gamma_{21n}(\tau) &= -e_{1G}'\{\mathcal{S}_r^G\}^{-1}p_T(\tau)^{-1}\frac{1}{h_G}\int\mu_G\Big(\frac{T-\tau}{h_G}\Big)k\Big(\frac{T-\tau}{h_G}\Big)G'(T)\cdot E\Big[\sum_{k=1}^{2}B_{nfk}(X_i^k)\Big|T_i = T\Big]p_T(T)dT + R_{0n}\\
&= -G'(\tau)\cdot E\Big[\sum_{k=1}^{2}\big\{h_{\mathcal{H}}^r\mathfrak{B}_k(\zeta_k(x^k)) + h_H^{r+1}\big[\tilde{f}_k'(\zeta_k(x^k))D_k(x^k) + \tilde{\mathcal{B}}_k(\zeta_k(x^k))\big]\big\}\Big|T=\tau\Big] + R_{0n},
\end{aligned} \tag{B.8}$$

[9]Here, We derive a weaker, pointwise representation rather than the uniform representation in SM Lemma S.5.

https://doi.org/10.1017/S0266466624000318 Published online by Cambridge University Press

where $R_{0n} = o_p(h_H^{r+1} + h_{\mathcal{H}}^r)$, and the last equality is due to (i) change of variables, (ii) Taylor expansion, and (iii) the fact that $e_{1G}'\{\mathcal{S}_r^G\}^{-1}\int \mu_G(u)k(u)du = e_{1G}'e_{1G} = 1$.

Next consider $\Gamma_{22n}(\tau)$. It represents the additional stochastic term induced by $\widehat{T}(X_i)$. Similar to SM Lemma S.5, under Assumption 6, $\Gamma_{22n}(\tau)$ can be written as

$$\Gamma_{22n}(\tau) = -e_{1G}'\{\mathcal{S}_r^G\}^{-1} p_T(\tau)^{-1}\frac{1}{nh_G}\sum_{i=1}^{n}\mu_G\Big(\frac{T_i-\tau}{h_G}\Big)k\Big(\frac{T_i-\tau}{h_G}\Big)G'(T_i)$$
$$\cdot\sum_{k=1}^{2}\big(\tilde{f}_k'(\zeta_k(X_i^k))\cdot J_{nk}(X_i^k) + \tilde{\mathfrak{J}}_{nk}(\zeta_k(X_i^k)) - E[\tilde{\mathfrak{J}}_{nk}(\zeta_k(X_i^k))|X_i]\big) + R_{1n}, \tag{B.9}$$

where $R_{1n} = o_p\big(h_G^{r+1} + h_H^{r+1} + h_{\mathcal{H}}^r + \sqrt{1/(nh_G)}\big)$. Follow the U-Statistics arguments similar to Lemma 8 of Horowitz (1998),[10] (B.9) can be represented as

$$\Gamma_{22n}(\tau) = e_{1G}'\{\mathcal{S}_r^G\}^{-1}\Big(\int\mu_G(u)k(u)du\Big)G'(\tau)\sum_{k=1}^{2}\{\Gamma_{22n,k}(\tau) - E[\Gamma_{22n,k}(\tau)]$$
$$+\widetilde{\Gamma}_{22n,k}(\tau) - E[\widetilde{\Gamma}_{22n,k}(\tau)]\} + R_{1n}$$
$$= G'(\tau)\sum_{k=1}^{2}\Gamma_{22n,k}(\tau) + G'(\tau)\sum_{k=1}^{2}\widetilde{\Gamma}_{22n,k}(\tau) + R_{1n},$$

for all $\tau \in \mathcal{S}_G$, where $\Gamma_{22n,k}(\tau) = \frac{1}{nh_{\mathcal{H}}}\sum_{i=1}^n c^{2-k}\omega_{5-k}(Z_i^{-k})q_k(\mathcal{Z}_{ki}^0)'e_d'\tilde{S}_r^{-1}V_k^{\tilde{\mu}}\Big(\frac{Z_i^k - z_0^k}{h_{\mathcal{H}}}\Big)\frac{Y_i - \mathcal{H}(Z_i)}{p_Z(\mathcal{Z}_{ki}^0)}\mathcal{K}_k\Big(\frac{z_0^k - Z_i^k}{h_{\mathcal{H}}}\Big)$, $\widetilde{\Gamma}_{22n,k}(\tau) = -\frac{1}{n}\sum_{i=1}^n \tilde{f}_k'\big(\zeta_k(X_i^k)\big)\frac{p_{X^1|T}(X_i^k|\tau)}{p_{X^k|X^{-k}}(X_i^k|X_i^{-k})}\big(Y_i - H(X_i)\big)$, and $E[\Gamma_{22n,k}(\tau)] = E[\widetilde{\Gamma}_{22n,k}(\tau)] = 0$. $\Gamma_{22n,k}(\tau)$ is the stochastic term due to the estimation of $\tilde{f}_k(\cdot)$, i.e., $\tilde{\mathfrak{J}}_{nk}(\zeta_k(X_i^k))$, and has an order of $O_p(1/\sqrt{nh_{\mathcal{H}}})$. $\widetilde{\Gamma}_{22n,k}(\tau)$ is induced by the estimation of $\zeta_k(X_i^k)$, i.e., $\tilde{f}_k'(\zeta_k(X_i^k))\cdot J_{nk}(X_i^k)$ with a variance of order $O(1/\sqrt{n})$. Therefore, $\widetilde{\Gamma}_{22n,k}(\tau)$ is of smaller order than $\Gamma_{22n,k}(\tau)$ and we conclude that

$$\Gamma_{22n}(\tau) = G'(\tau)\sum_{k=1}^{2}\Gamma_{22n,k}(\tau) + R_{1n}. \tag{B.10}$$

By combining the bias leading terms of (B.6) and (B.8), the asymptotic bias of $\widehat{G}(\tau)$ can be established. By the stochastic parts of (B.6) and (B.10), the asymptotic normality and correspondent variance follow from Lindeberg–Feller central limit theorem. This complete the whole proof. □

SUPPLEMENTARY MATERIAL

The supplementary material for this article can be found at https://doi.org/10.1017/S0266466624000318.

[10]Note that we use our SM Lemma S.1 instead of Horowitz's (1998) Lemma 5 to characterize the projection error of U-statistics.

REFERENCES

Ai, C., & Chen, X. (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica*, 71(6), 1795–1843.

Antras, P. (2004). Is the US aggregate production function Cobb–Douglas? New estimates of the elasticity of substitution. *Contributions in Macroeconomics*, 4(1), Article 20121005.

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

Berkowitz, D., Ma, H., & Nishioka, S. (2017). Recasting the iron rice bowl: The reform of China's state-owned enterprises. *The Review of Economics and Statistics*, 99(4), 735–747.

Chen, R., Härdle, W., Linton, O. B., & Severance-Lossin, E. (1996). Nonparametric estimation of additive separable regression models. In *Statistical theory and computational aspects of smoothing* (pp. 247–265). Springer.

Chen, S. (2002). Rank estimation of transformation models. *Econometrica*, 70(4), 1683–1697.

Chen, S. (2010a). An integrated maximum score estimator for a generalized censored quantile regression model. *Journal of Econometrics*, 155(1), 90–98.

Chen, S. (2010b). Root-*N*-consistent estimation of fixed-effect panel data transformation models with censoring. *Journal of Econometrics*, 159(1), 222–234.

Chen, S. (2012). Distribution-free estimation of the Box–Cox regression model with censoring. *Econometric Theory*, 28(3), 680–695.

Chen, S., & Zhang, H. (2020). Root-*N*-prediction of generalized heteroscedastic transformation regression models. *Journal of Econometrics*, 215(2), 305–340.

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

Chesher, A. (2003). Identification in nonseparable models. *Econometrica*, 71(5), 1405–1441.

Chung, K. L. (2001). *A course in probability theory*. Academic Press.

De Paula, A., & Tang, X. (2012). Inference of signs of interaction effects in simultaneous games with incomplete information. *Econometrica*, 80(1), 143–172.

Fan, J., & Gijbels, I. (1992). Variable bandwidth and local linear regression smoothers. *Annals of Statistics*, 20(4), 2008–2036.

Gentry, M., & Li, T. (2014). Identification in auctions with selective entry. *Econometrica*, 82(1), 315–344.

Grieco, P. L. (2014). Discrete games with flexible information structures: An application to local grocery markets. *The RAND Journal of Economics*, 45(2), 303–340.

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

Guerre, E., Perrigne, I., & Vuong, Q. (2009). Nonparametric identification of risk aversion in first-price auctions under exclusion restrictions. *Econometrica*, 77(4), 1193–1227.

Hahn, J., Liao, Z., & Ridder, G. (2018). Nonparametric two-step sieve *M* estimation and inference. *Econometric Theory*, 34(6), 1281–1324.

Hall, P. (1992). On bootstrap confidence intervals in nonparametric regression. *Annals of Statistics*, 20(2), 695–711.

Hall, P., & Horowitz, J. (2013). A simple bootstrap method for constructing nonparametric confidence bands for functions. *Annals of Statistics*, 41(4), 1892–1921.

Hansen, B. E. (2009). *Lecture notes on nonparametrics*. University of Wisconsin–Madison.

Härdle, W., & Bowman, A. W. (1988). Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands. *Journal of the American Statistical Association*, 83(401), 102–110.

Hoderlein, S., Su, L., White, H., & Yang, T. T. (2016). Testing for monotonicity in unobservables under unconfoundedness. *Journal of Econometrics*, 193(1), 183–202.

Hodges, D. J. (1969). A note on estimation of Cobb–Douglas and CES production function models. *Econometrica*, 37(4), 721–725.

https://doi.org/10.1017/S0266466624000318 Published online by Cambridge University Press

Horowitz, J. (1998). Nonparametric estimation of a generalized additive model with an unknown link function [Working paper]. University of Iowa.

Horowitz, J. L. (2001). Nonparametric estimation of a generalized additive model with an unknown link function. *Econometrica*, 69(2), 499–513.

Horowitz, J. L., & Mammen, E. (2004). Nonparametric estimation of an additive model with a link function. *Annals of Statistics*, 32(6), 2412–2443.

Horowitz, J. L., & Mammen, E. (2007). Rate-optimal estimation for a general class of nonparametric regression models with unknown link functions. *Annals of Statistics*, 35(6), 2589–2619.

Horowitz, J. L., & Mammen, E. (2011). Oracle-efficient nonparametric estimation of an additive model with an unknown link function. *Econometric Theory*, 27(3), 582–608.

Jacho-Chávez, D., Lewbel, A., & Linton, O. (2010). Identification and nonparametric estimation of a transformed additively separable model. *Journal of Econometrics*, 156(2), 392–407.

Khan, S. (2001). Two-stage rank estimation of quantile index models. *Journal of Econometrics*, 100(2), 319–355.

Klump, R., McAdam, P., & Willman, A. (2007). Factor substitution and factor-augmenting technical progress in the United States: A normalized supply-side system approach. *The Review of Economics and Statistics*, 89(1), 183–192.

Kmenta, J. (1967). On estimation of the CES production function. *International Economic Review*, 8(2), 180–189.

Kohler, M., & Krzyżak, A. (2017). Nonparametric regression based on hierarchical interaction models. *IEEE Transactions on Information Theory*, 63(3), 1620–1630.

Kong, E., Linton, O., & Xia, Y. (2010). Uniform Bahadur representation for local polynomial estimates of *M*-regression and its application to the additive model. *Econometric Theory*, 26(5), 1529–1564.

Lewbel, A., Lu, X., & Su, L. (2015). Specification testing for transformation models with an application to generalized accelerated failure-time models. *Journal of Econometrics*, 184(1), 81–96.

Lewbel, A., & Tang, X. (2015). Identification and estimation of games with incomplete information using excluded regressors. *Journal of Econometrics*, 189(1), 229–244.

Li, H., & Liu, N. (2018). Nonparamametric identification and estimation of double auctions with bargaining [Working paper]. Shanghai University of Finance and Economics.

Li, Q., & Racine, J. S. (2007). *Nonparametric econometrics: Theory and practice*. Princeton University Press.

Li, T., & Zheng, X. (2009). Entry and competition effects in first-price auctions: Theory and evidence from procurement auctions. *The Review of Economic Studies*, 76(4), 1397–1429.

Lin, H., Pan, L., Lv, S., & Zhang, W. (2018). Efficient estimation and computation for the generalised additive models with unknown link function. *Journal of Econometrics*, 202(2), 230–244.

Linton, O., & Härdle, W. (1996). Estimation of additive regression models with known links. *Biometrika*, 83(3), 529–540.

Liu, N., & Luo, Y. (2017). A nonparametric test for comparing valuation distributions in first-price auctions. *International Economic Review*, 58(3), 857–888.

Liu, N., & Vuong, Q. (2020). Nonparametric tests for monotonicity of strategies in games of incomplete information [Working paper]. New York University.

Liu, N., Vuong, Q., & Xu, H. (2017). Rationalization and identification of binary games with correlated types. *Journal of Econometrics*, 201(2), 249–268.

Ma, S. (2012). Two-step spline estimating equations for generalized additive partially linear models with large cluster sizes. *Annals of Statistics*, 40(6), 2943–2972.

Ma, S., & Song, P. X.-K. (2015). Varying index coefficient models. *Journal of the American Statistical Association*, 110(509), 341–356.

Marmer, V., & Shneyerov, A. (2012). Quantile-based nonparametric inference for first-price auctions. *Journal of Econometrics*, 167(2), 345–357.

Masry, E. (1996). Multivariate local polynomial regression for time series: uniform strong consistency and rates. *Journal of Time Series Analysis*, 17(6), 571–599.

Matzkin, R. L. (2003). Nonparametric estimation of nonadditive random functions. *Econometrica*, 71(5), 1339–1375.

Paraskevopoulos, C. C. (1979). Alternative estimates of the elasticity of substitution: An intermetropolitan CES production function analysis of US manufacturing industries, 1958–1972. *The Review of Economics and Statistics*, 61(3), 439–442.

Pinkse, J. (2001). Nonparametric regression estimation using weak separability [Working paper]. University of British Columbia.

Pollard, D. (1984). *Convergence of stochastic processes*. Springer-Verlag.

Powell, J. L., Stock, J. H., & Stoker, T. M. (1989). Semiparametric estimation of index coefficients. *Econometrica*, 57(6), 1403–1430.

Schmidt-Hieber, J. (2020). Nonparametric regression using deep neural networks with ReLU activation function. *Annals of Statistics*, 48(4), 1875–1897.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Chapman and Hall/CRC.

Tang, X. (2010). Estimating simultaneous games with incomplete information under median restrictions. *Economics Letters*, 108(3), 273–276.