GLS UNDER MONOTONE HETEROSKEDASTICITY

YOICHI ARAI, TAISUKE OTSU, AND MENGSHAN XU

Abstract. The generalized least square (GLS) is one of the most basic tools in regression analyses. A major issue in implementing the GLS is estimation of the conditional variance function of the error term, which typically requires a restrictive functional form assumption for parametric estimation or tuning parameters for nonparametric estimation. In this paper, we propose an alternative approach to estimate the conditional variance function under nonparametric monotonicity constraints by utilizing the isotonic regression method. Our GLS estimator is shown to be asymptotically equivalent to the infeasible GLS estimator with knowledge of the conditional error variance, and is free from tuning parameters, not only for point estimation but also for interval estimation or hypothesis testing. Our analysis extends the scope of the isotonic regression method by showing that the isotonic estimates, possibly with generated variables, can be employed as first stage estimates to be plugged in for semiparametric objects. Simulation studies illustrate excellent finite sample performances of the proposed method. As an empirical example, we revisit Acemoglu and Restrepo's (2017) study on the relationship between an aging population and economic growth to illustrate how our GLS estimator effectively reduces estimation errors.

1. Introduction

The generalized least square (GLS) is one of the most basic tools in regression analyses. It yields the best linear unbiased estimator in the classical linear regression model, and has been studied extensively in econometrics and statistics literature; see e.g., Wooldridge (2010, Chapter 7) for a review. A major issue to implement the GLS is that the optimal weights given by the conditional error variance function (say, $\sigma^2(\cdot)$) are typically unknown to researchers and need to be estimated. One way to estimate $\sigma^2(\cdot)$ is to specify its parametric functional form and estimate it by a parametric regression for the squared OLS residuals of the original regression on the specified covariates. However, economic theory rarely provides exact functional forms of $\sigma^2(\cdot)$, and the feasible GLS using misspecified $\sigma^2(\cdot)$ is no longer asymptotically efficient (Cragg, 1983). To address this issue, Carroll (1982) and Robinson (1987) proposed to estimate $\sigma^2(\cdot)$ nonparametrically and established the asymptotic equivalence of the resulting feasible GLS estimator with the infeasible one under certain regularity conditions. This is a remarkable result, but it requires theoretically and practically judicious choices of tuning parameters, such as bandwidths, series lengths, or numbers of neighbors. It should be noted that such tuning parameters appear in not only the point estimator but also its standard error for inference, and their choices require separate investigations.

In this paper, we propose an alternative approach to estimate the conditional variance function to implement the GLS by exploring a shape constraint of $\sigma^2(\cdot)$ instead of its smoothness as in Robinson (1987). As argued by Matzkin (1994), economic theory often provides shape constraints for functions of economic variables, such as monotonicity, concavity, or symmetry. In particular, we focus on situations where $\sigma^2(\cdot)$ is known to be monotone in its argument even though its exact functional form is unspecified, and propose to estimate $\sigma^2(\cdot)$ by utilizing the method of isotonic regression (see, Groeneboom and Jongbloed, 2014, for a review). It is known that the conventional isotonic regression estimator typically yields piecewise constant function estimates and does not involve any tuning parameters. Although the limiting behavior of the isotonic regression estimator is less tractable (such as the $n^{1/3}$-consistency and complicated limiting distribution), we show that our feasible GLS estimator using the optimal weights by the isotonic estimator is asymptotically equivalent to the infeasible GLS estimator. Furthermore, we can plug-in this isotonic estimator to estimate the asymptotic variance of the GLS estimator so that statistical inference (confidence intervals or hypothesis testing) can be also free from tuning parameters.

For the linear model $Y = X'\beta + U$ in the presence of heteroskedasticity $\sigma^2(X) = E[U^2|X]$, using feasible GLS to improve the estimation efficiency has a long history. On the one hand, several parametric models have been proposed to estimate conditional error variance function $\sigma^2(\cdot)$. For example, for some constants $C > 0$ and λ, Box and Hill (1974) proposed $\sigma^2(X) = C(X'\beta)^{2-2\lambda}$, Bickel (1978) proposed $\sigma^2(X) = C\exp(\lambda(X'\beta))$, and Jobson and Fuller (1980) proposed $\sigma^2(X) = C\{1+\lambda(X'\beta)^2\}$; interestingly, all these parametric functions are monotone increasing (or decreasing) in the index of X. On the other hand, Carroll (1982) and Robinson (1987) estimated $\sigma^2(\cdot)$ with kernel and nearest neighbor estimator respectively, and they showed their semiparametric GLS estimators are asymptotically equivalent to the infeasible GLS estimator and thus efficient. Compared to existing parametric methods, our proposed method also imposes monotonicity, a feature implied by many above-mentioned parametric models, but it is nonparametric and does not rely on any specific parametric function form. Compared to existing nonparametric methods, our proposed method is tuning-parameter-free and its performance does not rely on a proper choice of tuning parameters. In the Monte Carlo simulations, we show that our proposed method outperforms the above-mentioned nonparametric methods at almost every choice of tuning parameters, while it performs as well as parametric feasible GLS estimators with correctly specified conditional error variance function.

The isotonic estimator can date back to the middle of the last century. Earlier work includes Ayer *et al.* (1955), Grenander (1956), Rao (1969, 1970), and Barlow and Brunk (1972), among others. The isotonic estimator of a regression function can be formulated as a least square estimation with monotonicity constraint. Suppose that the conditional expectation $E[Y|X] = m(X)$ is monotone increasing, for an iid random sample $\{Y_i, X_i\}_{i=1}^n$, the isotonic estimator is the minimizer of the sum of squared errors, $\min_{m\in\mathcal{M}} \sum_{i=1}^n \{Y_i - m(X_i)\}^2$, where $\mathcal{M}$ is the class of monotone increasing functions. The minimizer can be calculated with the pool adjacent violators algorithm (Barlow and Brunk, 1972), or equivalently by solving the greatest convex minorant of the cumulative sum diagram $\{(0,0), (i, \sum_{j=1}^i Y_j), i = 1, \ldots, n\}$, where the corresponding $\{X_i\}_{i=1}^n$ are ordered sequence. See Groeneboom and Jongbloed (2014) for a comprehensive discussion of different aspects of isotonic regression. Moreover, recent developments in the monotone single index model provide convenient and flexible tools for combining monotonicity and multi-dimensional covariates, In a monotone single index model, the conditional mean of Y is modeled as $E[Y|X] = m(X'\alpha)$, and the monotone link function $m(\cdot)$ is solved with isotonic regression. Balabdaoui, Durot, and Jankowski (2019) studied the monotone single index model with the monotone least square method, and Groeneboom and Hendrickx (2018), Balabdaoui, Groeneboom, and Hendrickx (2019), and Balabdaoui and Groeneboom (2021) developed a score-type approach for the monotone single index model. Their approach can estimate the single index parameter α and the link function $m(\cdot)$ at $n^{-1/2}$-rate and $n^{-1/3}$-rate respectively. We contribute to this literature by showing that the isotonic estimates can be employed as first stage estimates to be plugged in for semiparametric objects. Furthermore, we note that our isotonic estimator involves generated variables (i.e. OLS residuals), which make theoretical developments substantially different from the existing ones.

This paper is organized as follows. In Section 2, we consider the case where $\sigma^2(\cdot)$ is monotone in one covariate, present our GLS estimator, and study its asymptotic properties. Section 3 extends our GLS approach to the case where $\sigma^2(\cdot)$ is specified by a monotone single-index function. Section 4 illustrates the proposed method by a simulation study and empirical example.

2. Heteroskedasticity by univariate covariate

We first consider the case where monotone heteroskedasticity is caused by a single covariate. In particular, consider the following multiple linear regression model

$$Y = \alpha + \beta X + Z'\gamma + U, \qquad E[U|X, Z] = 0, \tag{2.1}$$

where $X \in \mathscr{X} = [x_L, x_U]$ is a scalar covariate with compact support and Z is a vector of other covariates. In this section, we focus on the case where heteroskedasticity is caused by the covariate X, i.e.,

$$E[U^2|X, Z] = E[U^2|X] =: \sigma^2(X), \tag{2.2}$$

and $\sigma^2(\cdot)$ is a monotone increasing function. The monotone decreasing case is analogously (by setting U^2 as $-U^2$).

Let $\theta = (\alpha, \beta, \gamma')'$ be a vector of the slope parameters and $W = (1, X, Z')'$ so that the model in (2.1) can be written as $Y = W'\theta + U$. Based on an i.i.d. sample $\{Y_i, X_i, Z_i\}_{i=1}^n$, the infeasible GLS estimator for θ is written as

$$\hat{\theta}_{\text{IGLS}} = \left(\sum_{i=1}^n \sigma_i^{-2} W_i W_i'\right)^{-1} \left(\sum_{i=1}^n \sigma_i^{-2} W_i Y_i\right), \tag{2.3}$$

where $\sigma_i^2 = \sigma^2(X_i)$. In order to make this estimator feasible, various approaches have been proposed in the literature.

In this paper, we are concerned with the situation where the researcher knows $\sigma^2(\cdot)$ is monotone in a particular regressor X but its exact functional form is unspecified. In particular, by utilizing knowledge of monotonicity of $\sigma^2(\cdot)$, we propose to estimate $\sigma^2(\cdot)$ by the isotonic regression from the squared OLS residual on the regressor X. More precisely, let $\hat{\theta}_{\text{OLS}} =$

$\left(\sum_{i=1}^n W_i W_i'\right)^{-1}\left(\sum_{i=1}^n W_i Y_i\right)$ be the OLS estimator for (2.1), and $\hat{U}_j = Y_j - W_j'\hat{\theta}_{\mathrm{OLS}}$ be its residual. Then we estimate $\sigma^2(\cdot)$ by

$$\hat{\sigma}^2(\cdot) = \text{isotonic regression function from } \{\hat{U}_j^2\}_{j=1}^n \text{ on } \{X_j\}_{j=1}^n. \tag{2.4}$$

Although this estimator is shown to be consistent for $\sigma^2(\cdot)$ at the interior of support $[x_L, x_U]$ of X, it is generally biased at the lower boundary x_L, which may cause inconsistency of the resulting GLS estimator. Therefore, we propose to trim observations whose X_i's are too close to x_L, and develop the following feasible GLS estimator

$$\hat{\theta} = \left(\sum_{i=1}^n \mathbb{I}\{X_i \geq q_n\}\hat{\sigma}_i^{-2} W_i W_i'\right)^{-1}\left(\sum_{i=1}^n \mathbb{I}\{X_i \geq q_n\}\hat{\sigma}_i^{-2} W_i Y_i\right), \tag{2.5}$$

where $\mathbb{I}\{\cdot\}$ is the indicator function, and the trimming term q_n is set as the $(n^{-1/3})$-th sample quantile of $\{X_i\}_{i=1}^n$.

To study asymptotic properties of the proposed estimator $\hat{\theta}$, we impose the following assumptions. Let $\mathcal{B}(a, R)$ be a ball around a with radius R.

Assumption.

- **A1:** $\{Y_i, X_i, Z_i\}_{i=1}^n$ *is an iid sample of* (Y, X, Z). *The support of* (X, Z) *is convex with non-empty interiors and is a subset of* $\mathcal{B}(0, R)$ *for some* $R > 0$. *The support of* X *is a compact interval* $\mathscr{X} = [x_L, x_U]$.
- **A2:** $\sigma^2 : \mathscr{X} \to \mathbb{R}$ *is a monotone increasing function defined on* $\mathscr{X}$, *and* $0 < \sigma^2(x_L) < \sigma^2(x_U) < \infty$. *There exist positive constants* a_0 *and* M *such that* $E[|U|^{2s}|X = x] \leq a_0 s! M^{s-2}$ *for all integers* $s \geq 2$ *and* $x \in \mathscr{X}$. $\sigma^2(\cdot)$ *is continuously differentiable in* $(x_L, x_L + \delta)$ *for some* $\delta > 0$.
- **A3:** X *has a continuous density function* $f(\cdot)$ *on* $\mathscr{X}$, *and there exists a positive constant* b *such that* $b < f(x) < \infty$ *for all* $x \in \mathscr{X}$.

Assumption A1 is standard. Compact support of X is required to apply the isotonic regression on X. Assumption A2 is on the error term. Monotonicity of $\sigma^2(\cdot)$ is the main assumption. Assumption A3 contains additional mild conditions on the density of X.

We first present asymptotic properties of the conditional error variance estimator $\hat{\sigma}^2(\cdot)$ in (2.4). Let q_n^* be the $(n^{-1/3})$-th population quantile of X, $D_A^L[f](a)$ be the left derivative of the greatest convex minorant of a function f evaluated at $a \in A$, $\{\mathcal{W}_t\}$ be the standard Brownian motion, $\varepsilon = U^2 - \sigma^2(X)$, and $\sigma_\varepsilon^2(x) = E[\varepsilon^2|X = x]$. Also define $c^* = \lim_{n\to\infty} n^{1/3}(q_n^* - x_L)$. Assumption A3 guarantees $0 < c^* < \infty$. Then we obtain the following lemma for the behavior of $\hat{\sigma}^2(\cdot)$ around the boundary x_L, which extends the result by Babbi and Kumar (2021, Theorem 2.1(ii)) by allowing the generated variable $\hat{U}_i$ as a regressand for $\hat{\sigma}^2(\cdot)$.

Lemma 1. *Under Assumptions A1-A3, it holds*

$$n^{1/3}\{\hat{\sigma}^2(q_n) - \sigma^2(q_n)\} \xrightarrow{d} D_{[0,\infty)}^L\left[\sqrt{\frac{\sigma_\varepsilon^2(x_L)}{c^* f(x_L)}}\mathcal{W}_t + \left(\lim_{x\downarrow x_L}\frac{d\sigma^2(x)}{dx}\right)c^*\left(\frac{1}{2}t^2 - t\right)\right](1). \tag{2.6}$$

Based on this lemma, the asymptotic distribution of our feasible GLS estimator $\hat{\theta}$ is obtained as follows.

Theorem 1. *Under Assumptions A1-A3, it holds*

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, E[\sigma_i^{-2} W_i W_i']^{-1}),$$

and the asymptotic variance matrix is consistently estimated by $\left(\frac{1}{n}\sum_{i=1}^n \hat{\sigma}_i^{-2} W_i W_i'\right)^{-1}$.

This theorem implies that our estimator $\hat{\theta}$ has the same limiting distribution as the infeasible GLS estimator $\hat{\theta}_{\text{IGLS}}$ and thus achieves the semiparametric efficiency bound. This result extends the scope of the isotonic regression method by showing that the isotonic estimates, possibly with generated variables, can be employed as first stage estimates to be plugged in for semiparametric objects. We re-emphasize that $\hat{\theta}$ does not involve any tuning parameter (since q_n is completely determined by the sample).

Remark 1. Monotonicity is an assumption that can be tested. For observable random variables (Y, X), several methods have been developed to test whether $E[Y|X]$ is monotone increasing in X. See, e.g., Ghosal, Sen and van der Vaart (2000), Hall and Heckman (2000), Dumbgen and Spokoiny (2001), Chetverikov (2019), and Hsu, Liu and Shi (2019), among others. All these tests can be adapted for our case, testing the monotonicity of $\sigma^2(\cdot)$ with generated $\{\hat{U}_j^2\}_{j=1}^n$ and observed $\{X_j\}_{j=1}^n$. Since Assumptions A1-A2 and $\hat{\theta}_{\text{OLS}} - \theta = O_p(n^{-1/2})$ imply $\hat{U}_j^2 - U_j^2 = O_p(n^{-1/2}\log n)$ uniformly over $j = 1, \ldots, n$, the critical values of these tests can be adjusted accordingly to maintain a proper asymptotic size. We emphasize that in contrast to such nonparametric testing problems, it is highly nontrivial to guarantee that the estimation errors by generated variables are asymptotically negligible for the $\sqrt{n}$-consistent object like our GLS estimator $\hat{\theta}$.

Remark 2. We want to note that even if the true $\sigma^2(\cdot)$ is not monotone increasing in X, our feasible GLS estimator $\hat{\theta}$ in (2.5) is still consistent for θ due to $E[U|X, Z] = 0$. In this case, the asymptotic distribution of $\hat{\theta}$ is

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, E[p(X_i)^{-1} W_i W_i']^{-1} E[\sigma_i^2 p(X_i)^{-2} W_i W_i'] E[p(X_i)^{-1} W_i W_i']^{-1}),$$

where $p(\cdot)$ denotes the projection of $E[U^2|X, Z]$ onto the space of the monotone increasing functions supported on $\mathscr{X}$, and it can be estimated by the same isotonic regression as (2.4). Let $\hat{p}(\cdot)$ denote the isotonic estimator of $p(\cdot)$, then the asymptotic variance matrix can be consistently estimated by

$$\left(\frac{1}{n}\sum_{i=1}^n \hat{p}(X_i)^{-1} W_i W_i'\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^n \hat{p}(X_i)^{-2} \hat{U}_i^2 W_i W_i'\right) \left(\frac{1}{n}\sum_{i=1}^n \hat{p}(X_i)^{-1} W_i W_i'\right)^{-1}.$$

3. Heteroskedasticity by multivariate covariates

We now consider the model

$$Y = \alpha + X'\beta + Z'\gamma + U, \qquad E[U|X, Z] = 0,$$

where X is a vector of covariates. This section focuses on the case where heteroskedasticity takes the form of a monotone single-index function of X with unknown parameters η_0, i.e., $E[U^2|X,Z] = E[U^2|X] = \sigma^2(X'\eta_0)$ for a monotone increasing function $\sigma^2(\cdot)$. As we mentioned, the monotone index model $\sigma^2(X'\eta_0)$ covers existing parametric models (e.g., Box and Hill, 1974, Bickel, 1978, and Jobson and Fuller, 1980) as special cases. For identification, η_0 is normalized as $||\eta_0|| = 1$. Define

$$\sigma^2_\eta(a) := E[\sigma^2(X'\eta_0)|X'\eta = a]. \tag{3.1}$$

We show in Lemma 4 that $\sigma^2(\cdot)$ and η_0 can be consistently estimated by extending the method proposed in Balabdaoui, Groeneboom, and Hendrickx (2019) (BGH hereafter) and Balabdaoui and Groeneboom (2021) to allow generated variables. In particular, for a fixed η, define the isotonic regression of $\{\hat{U}_i^2\}_{i=1}^n$ on $\{X_i'\eta\}_{i=1}^n$ as

$$\hat{\sigma}^2_\eta = \arg\min_{m\in\mathcal{M}} \frac{1}{n}\sum_{i=1}^n \{\hat{U}_i^2 - m(X_i'\eta)\}^2, \tag{3.2}$$

where $\mathcal{M}$ is the set of monotone increasing functions defined on $\mathbb{R}$. Based on this, $\hat{\eta}$ can be estimated by minimizing the square sum of a score function. For example, the simple score estimator in the spirit of BGH and Balabdaoui and Groeneboom (2021) is given by

$$\hat{\eta} = \arg\min_\eta \left\| \frac{1}{n}\sum_{i=1}^n X_i\{\hat{U}_i^2 - \hat{\sigma}^2_\eta(X_i'\eta)\} \right\|^2. \tag{3.3}$$

Letting $\hat{\sigma}_i^2 = \hat{\sigma}^2_{\hat{\eta}}(X_i'\hat{\eta})$ and $W = (1, X', Z')'$, we propose the following GLS estimator

$$\hat{\theta} = \left(\sum_{i=1}^n \mathbb{I}\{X_i'\hat{\eta} \geq q_n\}\hat{\sigma}_i^{-2} W_i W_i'\right)^{-1} \left(\sum_{i=1}^n \mathbb{I}\{X_i'\hat{\eta} \geq q_n\}\hat{\sigma}_i^{-2} W_i Y_i\right), \tag{3.4}$$

where q_n is the $(n^{-1/3})$-th sample quantile of $\{X_i'\hat{\eta}\}_{i=1}^n$.

Assumption.

- **M1:** *$\{Y_i, X_i, Z_i\}_{i=1}^n$ is an iid sample of (Y, X, Z). The support of (X, Z) is convex with non-empty interiors and is a subset of $\mathcal{B}(0, R)$ for some $R > 0$.*
- **M2:** *(i) There exists $\delta_0 > 0$ such that the function $a \mapsto \sigma^2_\eta(a)$ defined in (3.1) is monotone increasing on $I_\eta = \{x'\eta, x \in \mathcal{X}\}$ for each $\eta \in \mathcal{B}(\eta_0, \delta_0)$. (ii) $0 < \inf_{a\in I_\eta} \sigma^2_\eta(a) < \sup_{u\in I_\eta} \sigma^2_\eta(a) < \infty$ for each $\eta \in \mathcal{B}(\eta_0, \delta_0)$. (iii) There exist positive constants a_0 and M such that $E[|U|^{2s}|X = x] \leq a_0 s! M^{s-2}$ for all integers $s \geq 2$ and $x \in \mathcal{X}$. (iv) $\sigma^2_\eta(a)$ is continuously differentiable on I_η for each $\eta \in \mathcal{B}(\eta_0, \delta_0)$.*
- **M3:** *The random vector X has a density function $f(x)$ that is continuous on $\mathcal{X}$, and $0 < \underline{b} < f(x) < \overline{b} < \infty$, for all $x \in \mathcal{X}$ and the real positive numbers $\underline{b}$ and $\overline{b}$.*
- **M4:** *For each $\eta \in \mathcal{B}(\eta_0, \delta_0)$, the mapping $a \mapsto E[X|X'\eta = a]$ defined on I_η is bounded and has a finite total variation.*
- **M5:** $\mathrm{Cov}[X'(\eta_0 - \eta),\ \sigma^2(X'\eta)|X'\eta] \neq 0$ *almost surely for each $\eta \neq \eta_0$.*
- **M6:** *$B := \int (x - E[X|x'\eta_0])(x - E[X|x'\eta_0])' \left.\frac{d\sigma^2(a)}{da}\right|_{a=x'\eta_0} dP_0(x)$ has full rank.*

Assumptions M1-M3 are analogs of Assumptions A1-A3, respectively. The main assumption is monotonicity of $\sigma^2_\eta(\cdot)$. Assumptions M4-M5 are additional regularity conditions for the monotone index model.

To avoid unnecessarily heavy notations, in the multivariate case, we redefine some notations, which have similar meanings to those used in Section 2. Let q_n^* be the $(n^{-1/3})$-th population quantile of $X'\eta_0$, q_n be the $(n^{-1/3})$-th sample quantile of $\{X_i'\hat{\eta}\}_{i=1}^n$, and $c^* = \lim_{n\to\infty} n^{1/3}(q_n^* - x_L)$, $D_A^L[f](a)$ be the left derivative of the greatest convex minorant of function $f(\cdot)$ evaluated at $a \in A$, $x_L = \inf_{x\in\mathcal{X}}(x'\eta_0)$, and $x_U = \sup_{x\in\mathcal{X}}(x'\eta_0)$. Also $\sigma_\varepsilon^2(x)$ is defined in Appendix B. By Assumption M1, we have $-\infty < x_L < x_U < \infty$. Then similar to Lemma 1, we obtain the following lemma for the behavior of $\hat{\sigma}^2_{\hat{\eta}}(\cdot)$ around x_L.

Lemma 2. *Under Assumptions M1-M6, it holds*

$$n^{1/3}\{\hat{\sigma}^2_{\hat{\eta}}(q_n) - \sigma^2(q_n)\} \xrightarrow{d} D^L_{[0,\infty)}\left[\sqrt{\frac{\sigma_\varepsilon^2(x_L)}{c^* f(x_L)}}\mathcal{W}_t + \left(\lim_{a\downarrow x_L}\frac{d\sigma^2(a)}{da}\right)c^*\left(\frac{1}{2}t^2 - t\right)\right](1).$$

Based on this lemma, the asymptotic distribution of the GLS estimator $\hat{\theta}$ in (3.4) is obtained as follows. Let $\sigma_i^2 = \sigma^2(X_i'\eta_0)$.

Theorem 2. *Under Assumptions M1-M6, it holds*

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, E[\sigma_i^{-2}W_iW_i']^{-1}),$$

and the asymptotic variance matrix is consistently estimated by $\left(\frac{1}{n}\sum_{i=1}^n \hat{\sigma}_i^{-2}W_iW_i'\right)^{-1}$.

Similar comments to Theorem 1 apply here. Our estimator $\hat{\theta}$ is asymptotically equivalent to the infeasible GLS estimator $\hat{\theta}_{\mathrm{IGLS}}$. In terms of technical contribution, our theoretical analysis generalizes existing ones in e.g., Babbi and Kumar (2021), BGH, and Balabdaoui and Groeneboom (2021) to accommodate generated variables.

4. Numerical illustrations

4.1. **Simulation.** We now investigate the finite sample properties of the proposed GLS estimator by a Monte Carlo experiment. We follow the simulation design by Cragg (1983) and Newey (1993). The first data generating process, denoted by DGP1, is the heteroskedastic linear model with a univariate covariate and normally distributed disturbance:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_i + u_i, \qquad u_i = \sigma_i\varepsilon_i, \qquad \varepsilon_i \sim N(0,1), \\ \beta_1 &= \beta_2 = 1, \qquad \log(X_i) \sim N(0,1), \qquad X_i \text{ and } \varepsilon_i \text{ are independent}, \\ \sigma_i^2 &= .1 + .2X_i + .3X_i^2. \end{aligned} \tag{4.1}$$

We consider two sample sizes, n =100 and 500. The number of replications is set to 1,000.

In addition to the feasible GLS estimator with monotone heteroskedasticity (MGLS), we consider the ordinary least squares (OLS), infeasible generalized least squares (GLS), feasible GLS (FGLS), and nearest neighbor estimators (k-NN). GLS requires knowledge of the heteroskedastic function (4.1), including the values of the coefficients. In contrast, FGLS proceeds with the known functional form, but the coefficients are estimated. The "k-NN automatic" chooses the number of neighbors by a cross-validation procedure suggested in Newey (1993). All the estimators except OLS are the weighted least squares estimators, and their differences come from how the weights are calculated. Following Newey (1993), we calculate the weights for each method by taking a ratio of the predicted squared residual to the estimated variance of the disturbance, censoring the result below 0.04.

Table 4.1 presents the simulation results for estimation. The first column shows the estimation methods, and the following two columns show the RMSE and mean absolute error (MAE) for DGP1 with $n = 100$. The results for GLS report the levels of the RMSE and MAE, and those for others are their ratios relative to GLS. The next two columns give the corresponding results with $n = 500$. Two rows for each estimator show the results for β_0 and β_1, respectively. The inefficiency and inaccuracy of OLS are apparent. FGLS performs quite well, and this is natural when heteroskedastic functions are correctly specified. The performance of k-NN varies with the choice of k and is in between OLS and FGLS. We observe that the performance of MGLS is better than k-NN in every choice of tuning parameters. The result of MGLS is comparable to that of FGLS if not better. MGLS's independence of a smoothing parameter is clearly desirable.

The last four columns of Table 4.1 present the results for DGP2 with a homoskedastic error:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_i + u_i, \qquad u_i \sim N(0,1), \\ \beta_1 &= \beta_2 = 1, \qquad \log(X_i) \sim N(0,1), \qquad X_i \text{ and } u_i \text{ are independent.} \end{aligned}$$

For DGP2, all estimators work reasonably well although the performance of k-NN with $k = 6$ is worse than others.

Table 4.1. Simulation: Estimation with univariate covariate

	DGP1				DGP2			
	$n = 100$		$n = 500$		$n = 100$		$n = 500$	
Estimator	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
GLS (infeasible)	0.093	0.059	0.041	0.028	0.133	0.088	0.057	0.039
	0.108	0.073	0.048	0.032	0.055	0.034	0.021	0.014
OLS	3.831	3.479	5.574	4.495	1.000	1.000	1.000	1.000
	2.543	2.370	3.377	2.971	1.000	1.000	1.000	1.000
FGLS	1.245	1.233	1.598	1.152	1.026	1.033	1.024	1.067
	1.406	1.280	1.271	1.242	1.075	1.036	1.088	1.090
k-NN Automatic	1.633	1.511	1.355	1.167	1.130	1.081	1.181	1.138
	1.606	1.431	1.424	1.267	1.065	1.006	1.197	1.097
k-NN ($k = 6$)	1.525	1.498	1.474	1.417	1.253	1.155	1.359	1.276
	1.472	1.462	1.454	1.459	1.177	1.114	1.350	1.344
k-NN ($k = 15$)	1.566	1.365	1.251	1.108	1.079	1.059	1.081	1.140
	1.546	1.408	1.247	1.197	1.037	1.012	1.066	1.053
k-NN ($k = 24$)	1.685	1.457	1.291	1.160	1.039	0.980	1.044	1.098
	1.673	1.471	1.312	1.246	1.015	0.994	1.038	1.025
MGLS	1.326	1.279	1.113	1.129	1.049	1.075	1.027	1.075
	1.332	1.249	1.113	1.144	1.051	1.058	1.055	1.066

Next, we consider the heteroskedastic linear models with multivariate covariates, denoted by DGP3:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i, \qquad u_i = \sigma_i \varepsilon_i, \qquad \varepsilon_i \sim N(0,1), \\ \beta_1 &= \beta_2 = 1, \qquad \log(X_{1i}),\ \log(X_{2i}) \sim N(0,1), \qquad X_{1i},\ X_{2i} \text{ and } \varepsilon_i \text{ are independent}, \\ \sigma_i^2 &= (.2X_{1i} + .2X_{2i})^2. \end{aligned} \tag{4.2}$$

The heteroskedastic function of DPG3 is of a monotone single-index structure. Using the notation in (3.1), DGP3 corresponds to the structure with $\sigma^2(a) = a^2$, $X' = (X_1, X_2)$, and $\eta = (.2, .2)'$. The left panel of Table 4.2 shows the results of DGP3 in the same manner as Table 4.1. For each estimation method, two rows show the results for β_0 and β_1, and those for β_2 are omitted to avoid redundancy. k-NNs and MGLS perform better than FGLS, and this is in contrast to the performance of DGP1. In general, MGLS works better than k-NNs except for a few cases.

To see the potential applicability of MGLS to a non-single-index structure, we consider another heteroskedastic linear model denoted by DGP4:

$$\begin{aligned} &Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i, \quad u_i = \sigma_i \varepsilon_i, \quad \varepsilon_i \sim N(0,1), \\ &\beta_1 = \beta_2 = 1, \quad \log(X_{1i}),\ \log(X_{2i}) \sim N(0,1), \qquad X_{1i},\ X_{2i} \text{ and } \varepsilon_i \text{ are independent}, \\ &\sigma_i^2 = .1 + .2\tilde{X}_i + .3\tilde{X}_i^2, \quad \log(\tilde{X}_i) = \frac{\log(X_{1i}) + \log(X_{2i})}{\sqrt{2}}. \end{aligned} \tag{4.3}$$

The right panel of Table 4.2 shows the results. The results for DGP 4 are similar to those of DGP3. MGLS works remarkably well for the heteroskedasticity of a non-single index structure.

Table 4.2. Simulation: Estimation with multivariate covariates

	DGP 3				DGP 4			
	$n = 100$		$n = 500$		$n = 100$		$n = 500$	
Estimator	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
GLS (infeasible)	0.110	0.071	0.045	0.028	0.115	0.076	0.049	0.033
	0.109	0.072	0.048	0.033	0.071	0.046	0.029	0.020
OLS	4.255	4.168	6.650	6.695	3.792	2.980	4.897	4.186
	2.317	2.260	3.051	2.610	2.914	2.338	3.809	3.198
FGLS	2.516	2.141	2.731	2.037	1.427	1.233	1.699	1.219
	1.709	1.486	1.732	1.358	1.344	1.281	1.326	1.271
k-NN Automatic	2.108	1.709	1.786	1.537	1.778	1.488	1.709	1.390
	1.680	1.489	1.516	1.318	1.865	1.763	2.009	1.626
k-NN ($k = 6$)	1.787	1.587	1.670	1.666	1.674	1.521	1.594	1.537
	1.514	1.486	1.497	1.362	1.769	1.764	1.813	1.654
k-NN ($k = 15$)	1.850	1.669	1.490	1.385	1.669	1.491	1.373	1.246
	1.511	1.428	1.313	1.248	1.769	1.639	1.588	1.517
k-NN ($k = 24$)	2.008	1.816	1.570	1.510	1.799	1.581	1.392	1.254
	1.611	1.571	1.371	1.267	1.890	1.729	1.626	1.511
MGLS	1.993	1.659	1.667	1.481	1.647	1.422	1.320	1.251
	1.477	1.401	1.238	1.186	1.604	1.451	1.448	1.360

Next, we turn to the simulation results on inference. Tables 4.3 and 4.4 show empirical coverages (EC) and average lengths (AL) for the 95% confidence intervals under DGPs 1-4.

Again we consider GLS, OLS, FGLS, k-NN, and MGLS. For OLS, two types of confidence intervals are considered, one with the usual OLS standard error (OLS-U) and another with the heteroskedasticity-robust standard error (OLS-R). We observe that the empirical coverage is smaller than the nominal coverage 0.95 for all DGPs and all methods except GLS. It is natural that OLS-U performs poorly since it is invalid except for DGP2. The performance of k-NN is worse than others for all DGPSs in terms of empirical coverage. OLS-R, FGLS, and MGLS work similarly in terms of empirical coverage, however, we note that the average length of OLS-R is much larger than those of FGLS and MGLS except for DGP2. MGLS works quite well, for all DGPs especially for $n = 500$.

TABLE 4.3. Simulation: Inference with univariate covariate

	DGP 1				DGP 2			
	$n = 100$		$n = 500$		$n = 100$		$n = 500$	
Estimator	EC	AL	EC	AL	EC	AL	EC	AL
GLS (infeasible)	0.955	0.370	0.956	0.164	0.947	0.516	0.955	0.224
	0.962	0.441	0.960	0.196	0.952	0.207	0.970	0.085
OLS-U	0.742	0.740	0.636	0.349	0.947	0.516	0.955	0.224
	0.421	0.290	0.348	0.131	0.952	0.207	0.970	0.085
OLS-R	0.805	0.862	0.884	0.648	0.941	0.507	0.949	0.222
	0.772	0.689	0.880	0.488	0.881	0.185	0.935	0.081
FGLS	0.847	0.328	0.872	0.162	0.925	0.493	0.935	0.216
	0.812	0.395	0.885	0.195	0.758	0.159	0.844	0.072
kNN Automatic	0.659	0.258	0.701	0.102	0.884	0.483	0.883	0.205
	0.574	0.251	0.650	0.115	0.917	0.197	0.881	0.079
k-NN ($k = 6$)	0.666	0.258	0.621	0.102	0.845	0.483	0.819	0.205
	0.574	0.251	0.650	0.115	0.917	0.197	0.881	0.079
k-NN ($k = 15$)	0.704	0.266	0.717	0.105	0.907	0.492	0.914	0.210
	0.592	0.258	0.677	0.118	0.931	0.200	0.919	0.081
k-NN ($k = 24$)	0.688	0.269	0.721	0.107	0.921	0.504	0.935	0.216
	0.537	0.244	0.668	0.118	0.942	0.204	0.939	0.083
MGLS	0.812	0.363	0.905	0.165	0.907	0.468	0.937	0.219
	0.744	0.392	0.888	0.188	0.968	0.222	0.972	0.092

4.2. **Empirical example.** We illustrate how the proposed method in this paper can improve the precision of the traditional OLS approach. In doing so, we revisit Acemoglu and Restrepo (2017) that investigate the relationship between an aging population and economic growth. After Hansen (1939), a popular perspective is that countries undergoing faster aging suffer more economically partly because of excessive savings by an aging population. In contrast to the perspective, Acemoglu and Restrepo (2017) find no evidence of a negative relationship between aging and GDP per capita after controlling for initial GDP per capita, initial demographic composition, and trends by region.

Acemoglu and Restrepo (2017) estimated six specifications for the regression of the change in (log) GDP per capita from 1990 to 2015 (denoted by GDP) on the population aging measured by the change in the ratio of the population above 50 to those between the ages of 20 and 49 (denoted by Aging). The results are reproduced in panel A of Table 4.5. Those in columns 1-4 are based on the sample including 169 countries. Column 1 shows the result of the simple

Table 4.4. Simulation: Inference with multivariate covariates

	DGP 3				DGP 4			
	$n = 100$		$n = 500$		$n = 100$		$n = 500$	
Estimator	EC	AL	EC	AL	EC	AL	EC	AL
GLS	0.951	0.413	0.946	0.175	0.943	0.440	0.956	0.192
	0.961	0.439	0.960	0.194	0.950	0.282	0.968	0.123
OLS-U	0.786	1.108	0.632	0.511	0.780	0.893	0.675	0.412
	0.549	0.369	0.491	0.164	0.611	0.298	0.521	0.133
OLS-R	0.815	1.232	0.869	0.873	0.843	1.068	0.906	0.718
	0.782	0.660	0.891	0.441	0.810	0.517	0.914	0.334
FGLS	0.845	0.759	0.897	0.336	0.823	0.424	0.834	0.191
	0.722	0.395	0.826	0.198	0.862	0.262	0.855	0.112
k-NN Automatic	0.557	0.289	0.587	0.105	0.646	0.323	0.609	0.122
	0.472	0.205	0.534	0.091	0.599	0.200	0.605	0.082
k-NN ($k = 6$)	0.574	0.289	0.549	0.105	0.639	0.323	0.571	0.122
	0.472	0.205	0.534	0.091	0.599	0.200	0.605	0.082
k-NN ($k = 15$)	0.590	0.299	0.639	0.108	0.675	0.338	0.689	0.129
	0.484	0.212	0.570	0.096	0.629	0.210	0.668	0.088
k-NN ($k = 24$)	0.561	0.313	0.618	0.111	0.664	0.347	0.702	0.132
	0.450	0.204	0.562	0.096	0.608	0.207	0.662	0.091
MGLS	0.863	0.623	0.938	0.248	0.844	0.526	0.902	0.216
	0.756	0.401	0.897	0.198	0.776	0.305	0.908	0.154

regression. Standard errors robust to heteroskedasticity are reported in square brackets. Column 2 shows the result with an additional regressor, the initial log GDP per worker in 1990. Column 3 in addition includes the initial demographic information, the ratio of the population above 50 to those between 20 and 49 in 1990 (denoted by Initial Ratio), and the population in 1990. Column 4 additionally uses dummies for seven regions, Latin America, East Asia, South Asia, Africa, North Africa and Middle East, Eastern Europe and Central Asia, and Developed Countries. Columns (5) and (6) report the result for OECD countries using specifications (1) and (3), respectively. The number of observations for the first four columns is 169, and that for the last two columns is 35. Five out of six OLS estimates indicate positive relationships and three of them are statistically significant at the 5 percent level. Acemoglu and Restrepo (2017) discuss that these findings can be explained by the adoption of automation technologies based on a theoretical model.

We estimate the same specifications by MGLS proposed in this paper. Acemoglu and Restrepo (2017) show that the negative effect of aging can be mitigated or reversed by adopting new automation technologies given abundant capital. This also implies that the effect of aging can be negative without sufficient capital. Hence it would be reasonable to consider Aging as a source of heteroskedasticity. Figure 4.1 shows the relationship between the residual from the simple regression of column 1 in Panel A and Aging. Heteroskedasticity due to Aging is not easily confirmed visually. We consider Initial Ratio as another source of heteroskedasticity since the low ratio of old to young in 1990 is likely correlated with more aging in 2015, leading to larger variability in GDP per capita by the same reasoning discussed above. Figure 4.2 presents the relationship between the residual from the simple regression of column 1 in Panel A and Initial Ratio, and we see that the variability decreases with the growing ratio.

Panels B, C, and D of Table 4.5 show the results by MGLS. Panels B and C present the results for a case where heteroskedasticity is a function of Aging and Initial Ratio, respectively. Panel D reports the results from the estimation strategy described in Section 3, and here heteroskedasticity is modeled by a monotone single index structure with a monotone link function.

Standard errors based on MGLS are reported in parentheses. First, we observe reductions in standard errors for all MGLSs relative to OLS. Second, variations in estimates look smaller for statistically significant positive relationships than those for insignificant relationships. Third, we note that the results in columns 5 and 6 in panel D show statistically significant negative relationships. This indicates that the positive and statistically insignificant relationships for OECD countries in panels A, B, and C might be due to the small number of observations, and OECD countries reacted to aging differently from other countries. Further investigations on OECD countries can lead to new insights.

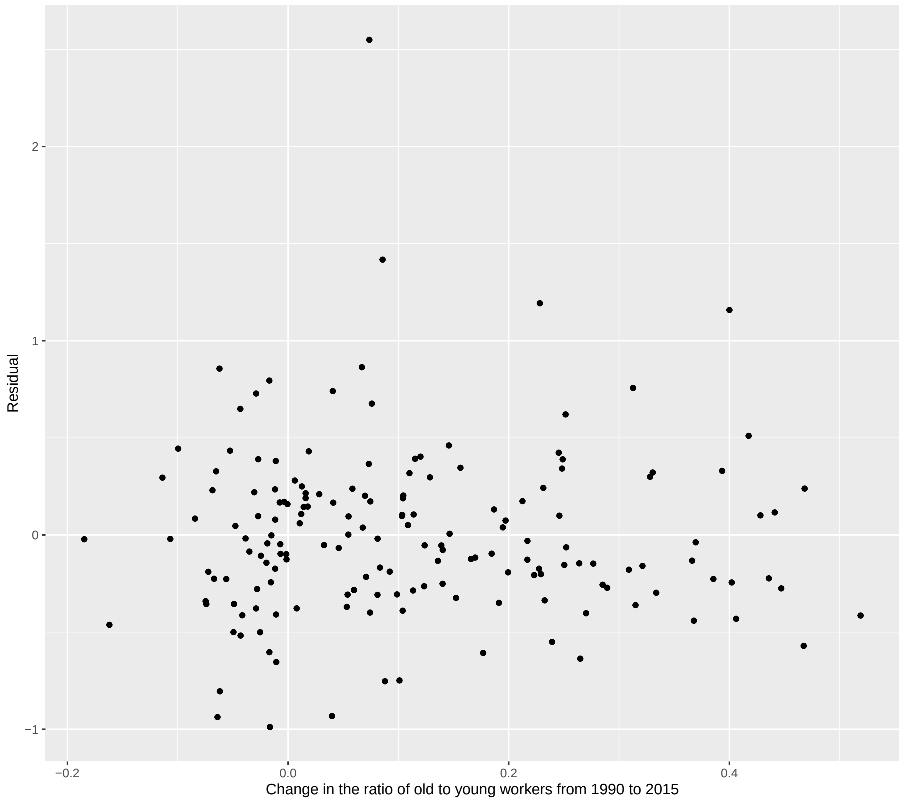

FIGURE 4.1. Residual and aging

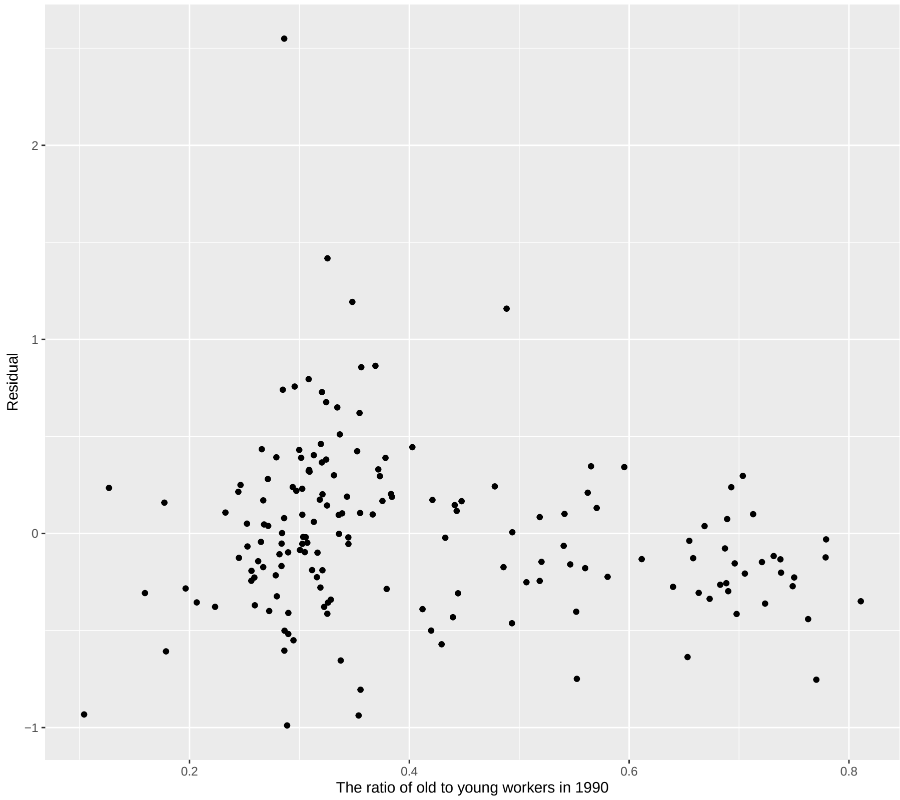

FIGURE 4.2. Residual and ratio of old to young workers in 1990

TABLE 4.5. Estimation of the Aging on GDP by OLS and MGLS

	Sample of all countries ($n = 169$)				OECD countries ($n = 35$)	
Specification	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: OLS						
Aging	0.335	1.036	1.162	0.773	-0.262	0.042
	[0.210]	[0.257]	[0.276]	[0.322]	[0.352]	[0.346]
Initial GDP		-0.153	-0.138	-0.156		-0.205
		[0.039]	[0.042]	[0.046]		[0.072]
Panel B: MGLS (Aging)						
Aging	0.387	1.098	1.191	0.751	-0.391	-0.029
	(0.189)	(0.186)	(0.203)	(0.260)	(0.247)	(0.271)
Initial GDP		-0.164	-0.155	-0.168		-0.19
		(0.027)	(0.029)	(0.029)		(0.066)
Panel C: MGLS (Initial Ratio)						
Aging	0.065	0.771	0.894	0.729	-0.266	0.034
	(0.196)	(0.222)	(0.229)	(0.292)	(0.339)	(0.262)
Initial GDP		-0.164	-0.141	-0.144		-0.208
		(0.031)	(0.034)	(0.037)		(0.069)
Panel D: MGLS (Both Aging and Initial Ratio)						
Aging	0.048	0.734	0.861	0.519	-0.446	-0.426
	(0.196)	(0.227)	(0.233)	(0.235)	(0.226)	(0.191)
Initial GDP		-0.15	-0.13	-0.162		-0.169
		(0.031)	(0.035)	(0.041)		(0.057)

Appendix A. Proof of lemma and theorem in Section 2

Notation. In this section, we use the following notation. Let $\|f\|_\infty = \sup_{x\in\mathcal{X}} |f(x)|$ be the sup-norm, $\|f\|_{2,P} = \sqrt{\int |f(x)|^2 dP}$ be the $L_2(P)$ norm, $D_A^L[f](a)$ be the left derivative of the greatest convex minorant of a function f evaluated at $a \in A$, $\mathbb{P}_n$ be the empirical measure of $\{Y_i, X_i, Z_i\}_{i=1}^n$, $\mathbb{G}_n$ be the empirical process, i.e., $\mathbb{G}_n f = \frac{1}{\sqrt{n}} \sum_{i=1}^n \{f(X_i) - E[f(X_i)]\}$, $\|\mathbb{G}_n\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}} |\mathbb{G}_n f|$, and $\mathbb{I}_A(x) = \mathbb{I}\{x \in A\}$. Let $\tau_0(x) = \sigma^2(x)$, $\tau_0'(x_L)$ be the right derivative of τ_0 at x_L, $\hat{\tau}(x) = \hat{\sigma}^2(x)$, $\mathscr{W}$ be the support of W, $F(x)$ be the distribution function of X, $F_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{I}\{X_i \le x\}$, and $M_n(x) = \frac{1}{n}\sum_{i=1}^n \hat{U}_i^2 \mathbb{I}\{X_i \le x\}$.

A.1. **Proof of Lemma 1.** Since $\hat{U}_j = Y_j - W_j'\hat{\theta}_{\text{OLS}}$ is the OLS residual, Assumptions A1-A2 and $\hat{\theta}_{\text{OLS}} - \theta = O_p(n^{-1/2})$ imply $\hat{U}_j^2 - U_j^2 = O_p(n^{-1/2}\log n) = o_p(n^{-1/3})$ uniformly over $j = 1, \ldots, n$. Assumption A3 guarantees $q_n^* - x_L = O(n^{-1/3})$ (by an expansion of $q_n^* = F^{-1}(n^{-1/3})$ for the quantile function $F^{-1}(\cdot)$ of X), and we can define $c^* = \lim_{n\to\infty} n^{1/3}(q_n^* - x_L) = \left.\frac{dF^{-1}(q)}{dq}\right|_{q\downarrow 0} \in (0,\infty)$.

We first analyze $n^{1/3}\{\hat{\tau}(q_n^*) - \tau_0(x_L)\}$. Pick any $m > 0$. Let

$$Z_{n1}(t) = n^{2/3}[\{n^{-1/3}m + \tau_0(x_L)\}F_n(x_L + t(q_n^* - x_L)) - M_n(x_L + t(q_n^* - x_L))].$$

Observe that

$$\begin{aligned} P\left(n^{1/3}\{\hat{\tau}(q_n^*) - \tau_0(x_L)\} \le m\right) &= P\left(\arg\max_{s\in[x_L,x_U]} [\{n^{-1/3}m + \tau_0(x_L)\}F_n(s) - M_n(s)] \ge q_n^*\right) \\ &= P\left(\arg\max_{t\in[0,(x_U-x_L)/(q_n^*-x_L)]} n^{-2/3}Z_{n1}(t) \ge 1\right), \end{aligned} \tag{A.1}$$

where the first equality follows from the switch relation (see, Groeneboom and Jongbloed, 2014), and the second equality follows from a change of variables $s = x_L + t(q_n^* - x_L)$. Let $\hat{U}(y,w) = y - w'\hat{\theta}_{\text{OLS}}$ and

$$g_{n,t}(y,w) = n^{1/6}\{\tau_0(x_L) - \hat{U}(y,w)^2\}\mathbb{I}_{[x_L, x_L+t(q_n^*-x_L)]}(x).$$

We decompose

$$\begin{aligned} Z_{n1}(t) &= \sqrt{n}(\mathbb{P}_n - P)g_{n,t} + n^{2/3}E[\{\tau_0(x_L) - \hat{U}(Y,W)^2\}\mathbb{I}_{[x_L,x_L+t(q_n^*-x_L)]}(X)] \\ &\quad + n^{1/3}m\{F_n(x_L + t(q_n^* - x_L)) - F(x_L + t(q_n^* - x_L))\} + n^{1/3}mF(x_L + t(q_n^* - x_L)) \\ &=: Z_{n1}^a(t) + Z_{n1}^b(t) + Z_{n1}^c(t) + Z_{n1}^d(t). \end{aligned}$$

Analysis of $Z_{n1}^a(t)$. We verify the conditions of van der Vaart (2000, Theorem 19.28). Define the class of random functions (depending on $\hat{\theta}_{\text{OLS}}$):

$$\mathcal{G}_{n1} = \{g_{n,t}(y,w) = n^{1/6}(\tau_0(x_L) - \hat{U}(y,w)^2)\mathbb{I}_{[x_L,x_L+t(q_n^*-x_L)]}(x) : t \in [0,K]\},$$

for $K \in (0,\infty)$, where n in the subscript indicates both the scaling parameter $n^{1/6}$ and $\hat{\theta}_{\text{OLS}}$. By van der Vaart (2000, Example 19.6) we know that for a bracket size ϵ, $\mathcal{G}_{n1}$ has the entropy with bracketing of order $\log(1/\epsilon)$. Thus, $\mathcal{G}_{n1}$ satisfies the entropy condition for van der Vaart (2000, Theorem 19.28).

For each $t, s \in [0, K]$, note that

$$\begin{aligned}
\mathrm{Cov}(g_{n,t}, g_{n,s}) &= n^{1/3} E[\{\hat{U}(Y,W)^2 - \tau_0(x_L)\}^2 \mathbb{I}_{[x_L, x_L + (t\wedge s)(q_n^* - x_L)]}(X)] + o_p(1) \\
&= n^{1/3} E[\{U^2 - \tau_0(x_L)\}^2 \mathbb{I}_{[x_L, x_L + (t\wedge s)(q_n^* - x_L)]}(X)] + o_p(1) \\
&= n^{1/3} E[[\varepsilon^2 + \{\tau_0(X) - \tau_0(x_L)\}^2] \mathbb{I}_{[x_L, x_L + (t\wedge s)(q_n^* - x_L)]}(X)] + o_p(1) \\
&= n^{1/3} \int_{x_L}^{x_L + (t\wedge s)(q_n^* - x_L)} [\sigma_\varepsilon^2(x) + \{\tau_0(x) - \tau_0(x_L)\}^2] f(x) dx + o_p(1) \\
&= [\sigma_\varepsilon^2(\xi_n) + \{\tau_0(\xi_n) - \tau_0(x_L)\}^2] f(\xi_n) c^*(t\wedge s) + o_p(1) \\
&= \sigma_\varepsilon^2(x_L) f(x_L) c^*(t\wedge s) + o_p(1), \qquad \text{(A.2)}
\end{aligned}$$

for $\xi_n \in (x_L, x_L + (t\wedge s)q_n^*)$, where the first equality follows from $q_n^* - x_L = O(n^{-1/3})$, the second equality follows from $\hat{\theta}_{\mathrm{OLS}} - \theta = O_p(n^{-1/2})$, the third equality follows from the definition $\varepsilon = U^2 - \tau(X)$ and $E[\varepsilon|X] = 0$, the fourth equality follows from the law of iterated expectations, the fifth equality follows from a Taylor expansion, and the last equality follows from continuity of $\sigma_\varepsilon^2(\cdot)$ and $\tau_0(\cdot)$ at x_L from right. Similarly, we have $\mathrm{Var}(g_{n,t}) = \sigma_\varepsilon^2(x_L) f(x_L) c^* t + o_p(1)$.

We next consider the envelop function of the class $\mathcal{G}_{n1}$, that is

$$G_{n1}(y,w) = n^{1/6} |\tau_0(x_L) - \hat{U}(y,w)^2| \cdot \mathbb{I}_{[x_L, x_L + K(q_n^* - x_L)]}(x).$$

We can see that G_{n1} is square integrable since a similar argument to (A.2) yields

$$\begin{aligned}
E[G_{n1}^2(Y,W)] &= n^{1/3} E[|\tau_0(x_L) - \hat{U}(Y,W)^2| \cdot \mathbb{I}_{[x_L, x_L + K(q_n^* - x_L)]}(X)] \\
&= n^{1/3} E[|\tau_0(x_L) - U^2| \cdot \mathbb{I}_{[x_L, x_L + K(q_n^* - x_L)]}(X)] + o_p(1) \\
&= n^{1/3} E[[\varepsilon^2 + \{\tau_0(X) - \tau_0(x_L)\}^2] \cdot \mathbb{I}_{[x_L, x_L + K(q_n^* - x_L)]}(X)] + o_p(1) \\
&= n^{1/3} \int_{x_L}^{x_L + K(q_n^* - x_L)} [\sigma_\varepsilon^2(x) + \{\tau_0(x) - \tau_0(x_L)\}^2] f(x) dx + o_p(1) = O_p(1) \qquad \text{(A.3)}
\end{aligned}$$

and thus the Lindeberg condition can be verified by Assumption A2: for any $\zeta > 0$ and some $\delta > 0$,

$$\begin{aligned}
E[G_{n1}^2 \mathbb{I}\{G_{n1} > \zeta\sqrt{n}\}] &\leq \frac{n^{(2+\delta)1/6}}{\zeta^\delta n^{\delta/2}} E[|\tau_0(x_L) - \hat{U}(Y,W)^2|^{2+\delta} \cdot \mathbb{I}_{[x_L, x_L + K(q_n^* - x_L)]}(X)] \\
&= \frac{n^{(2+\delta)1/6}}{\zeta^\delta n^{\delta/2}} E[|\tau_0(x_L) - U^2|^{2+\delta} \cdot \mathbb{I}_{[x_L, x_L + K(q_n^* - x_L)]}(X)] + o_p(1) \\
&= O(n^{-\delta/3}) + o_p(1) = o_p(1), \qquad \text{(A.4)}
\end{aligned}$$

where the inequality follows from Markov's inequality, the first equality follows from $\hat{\theta}_{\mathrm{OLS}} - \theta = O_p(n^{-1/2})$, and the second equality follows from a similar argument to (A.3). Similarly, as $\delta_n \to 0$, we obtain

$$\begin{aligned}
\sup_{|t-s|\leq\delta_n} E|g_{n,t} - g_{n,s}|^2 &= n^{1/3} \sup_{|t-s|\leq\delta_n} E[\{\hat{U}(Y,W)^2 - \tau_0(x_L)\}^2 \mathbb{I}_{[x_L, x_L + |t-s|q_n^*]}(X)] \\
&= n^{1/3} \sup_{|t-s|\leq\delta_n} E[[\varepsilon^2 + \{\tau_0(X) - \tau_0(x_L)\}^2] \cdot \mathbb{I}_{[x_L, x_L + |t-s|q_n^*]}(X)] + o_p(\delta_n) \\
&= O_p(\delta_n) = o_p(1). \qquad \text{(A.5)}
\end{aligned}$$

By (A.2)-(A.5), we can apply van der Vaart (2000, Theorem 19.28), which implies for each $K \in (0, \infty)$,

$$Z^a_{n1}(t) \xrightarrow{d} \sqrt{\sigma^2_\varepsilon(x_L) f(x_L)} \mathcal{W}_t \text{ in } l^\infty[0, K]. \tag{A.6}$$

Analysis of $Z^b_{n1}(t)$. Observe that

$$\begin{aligned}
Z^b_{n1}(t) &= n^{2/3} E[\{\tau_0(x_L) - U^2\} \mathbb{I}_{[x_L, x_L + t(q^*_n - x_L)]}(X)] + n^{2/3} E[(U^2 - \hat{U}(Y, W)^2) \mathbb{I}_{[x_L, x_L + t(q^*_n - x_L)]}(X)] \\
&= n^{2/3} \int_{x_L}^{x_L + t(q^*_n - x_L)} \{\tau_0(x_L) - \tau_0(F^{-1}(F(x)))\} dF(x) + o_p(1) \\
&= n^{2/3} \int_{F(x_L)}^{F(x_L + t(q^*_n - x_L))} \{\tau_0(x_L) - \tau_0(F^{-1}(v))\} dv + o_p(1) \\
&= -n^{2/3} \int_{F(x_L)}^{F(x_L + t(q^*_n - x_L))} \tau'_0(x_L) \{F^{-1}(v) - F^{-1}(F(x_L))\} dv + o_p(1) \\
&= -n^{2/3} \int_{F(x_L)}^{F(x_L + t(q^*_n - x_L))} \tau'_0(x_L) \frac{v - F(x_L)}{f(x_L)} dv + o_p(1) \\
&= -n^{2/3} \tau'_0(x_L) \frac{\{F(x_L + t(q^*_n - x_L)) - F(x_L)\}^2}{2 f(x_L)} + o_p(1) \\
&= -\tau'_0(x_L) \frac{t^2 (c^*)^2}{2} f(x_L) + o_p(1),
\end{aligned} \tag{A.7}$$

uniformly over $t \in [0, K]$, where the second equality follows from $E[\{U^2 - \hat{U}(Y, W)^2\} \mathbb{I}_{[x_L, x_L + t(q^*_n - x_L)]}(X)] = O_p(n^{-5/6}) = o_p(n^{-2/3})$, the third equality follows from a change of variables $v = F(x)$, the fourth equality follows from a Taylor expansion, the fifth equality follows from $F^{-1}(v) - x_L = \frac{1}{f(x_L)}(v - F(x_L)) + o(v - F(x_L))$, the sixth equality follows from evaluating the integral, and the last equality follows from a Taylor expansion and $q^*_n - x_L = O(n^{-1/3})$.

Analysis of $Z^c_{n1}(t)$. By Kim and Pollard (1990, Maximal inequality 3.1), we have

$$E\left[\sup_{t \in [0, K]} |F_n(x_L + t(q^*_n - x_L)) - F(x_L + t(q^*_n - x_L))|\right] \leq n^{-1/2} J \sqrt{PG^2_n},$$

for some constant $J \in (0, \infty)$, where $PG^2_n = F(n^{-1/3} t) = f(x_L)(1 + o(1)) c^* n^{-1/3}$. This implies

$$Z^c_{n1}(t) \leq n^{1/3} n^{-1/2} m J \sqrt{PG^2_n} = o(1), \tag{A.8}$$

uniformly over $t \in [-0, K]$.

Analysis of $Z^d_{n1}(t)$. A Taylor expansion yields

$$Z^d_{n1}(t) = n^{1/3} m F(x_L + t(q^*_n - x_L)) = m t f(x_L) c^* + o(1), \tag{A.9}$$

uniformly over $t \in [0, K]$, for every $K < \infty$.

Combining (A.6)-(A.9), it holds that for each $0 < K < \infty$,

$$Z_{1n}(t) \xrightarrow{d} Z_1(t) := \sqrt{\sigma^2_\varepsilon(x_L) f(x_L) c^*} \mathcal{W}_t - \tau'_0(x_L) \frac{t^2 (c^*)^2}{2} f(x_L) + m f(x_L) t c^* \text{ in } l^\infty[0, K]. \tag{A.10}$$

We now verify the conditions of the argmax continuous mapping theorem (Kim and Pollard, 1990). Note that for each $t \neq s$,

$$\text{Var}(Z_1(s) - Z_1(t)) = \sigma^2_\varepsilon(x_L) f(x_L) |t - s| \neq 0.$$

By Kim and Pollard (1990), the process $t \to Z_1(t)$ achieves its maximum a.s. at a unique point. Consider extended versions of Z_{1n} and Z_1 to the real line:

$$\tilde{Z}_{1n}(t) = \begin{cases} Z_{1n}(t), & t \geq 0 \\ t & t < 0 \end{cases}, \qquad \tilde{Z}_1(t) = \begin{cases} Z_1(t), & t \geq 0 \\ t & t < 0 \end{cases}.$$

It holds $\tilde{Z}_{1n}(t) \xrightarrow{d} \tilde{Z}_1(t)$, and the similar argument to Lemma SM.2.1 (ii) in Babbi and Kumar (2021) yields that the maximum of $\tilde{Z}_{1n}(t)$ is uniformly tight. Therefore, by Kim and Pollard (1990, Theorem 2.7),

$$\begin{aligned} & P\left(n^{1/3}\{\hat{\tau}(q_n^*) - \tau_0(x_L)\} \leq m\right) \to P\left(\left[\arg\max_{t \geq 0} Z_1(t)\right] \geq 1\right) \\ = \; & P\left(\left[\arg\max_{t \geq 0} \sqrt{\frac{\sigma_\varepsilon^2(x_L)}{c^* f(x_L)}} \mathcal{W}_t - \tau_0'(x_L)\frac{t^2 c^*}{2} + mt\right] \geq 1\right) \\ = \; & P\left(\left[D^L_{[0,\infty)}\left(\sqrt{\frac{\sigma_\varepsilon^2(x_L)}{c^* f(x_L)}} \mathcal{W}_t + \tau_0'(x_L)\frac{t^2 c^*}{2}\right)(1)\right] \leq m\right), \end{aligned}$$

where the second equality follows from the switch relation and symmetry of the process $\mathcal{W}_t$. Thus, we have

$$n^{1/3}\{\hat{\tau}(q_n^*) - \tau_0(x_L)\} \xrightarrow{d} D^L_{[0,\infty)}\left(\sqrt{\frac{\sigma_\varepsilon^2(x_L)}{c^* f(x_L)}} \mathcal{W}_t + \tau_0'(x_L)\frac{t^2 c^*}{2}\right)(1), \tag{A.11}$$

which also implies

$$\begin{aligned} & n^{1/3}\{\hat{\tau}(q_n^*) - \tau_0(q_n^*)\} \\ \xrightarrow{d} \; & D^L_{[0,\infty)}\left(\sqrt{\frac{\sigma_\varepsilon^2(x_L)}{c^* f(x_L)}} \mathcal{W}_t + \tau_0'(x_L)\frac{t^2 c^*}{2}\right)(1) - \lim_{n\to\infty} n^{1/3}\{\tau_0(q_n^*) - \tau_0(x_L)\} \\ \overset{d}{\sim} \; & D^L_{[0,\infty)}\left(\sqrt{\frac{\sigma_\varepsilon^2(x_L)}{c^* f(x_L)}} \mathcal{W}_t + \tau_0'(x_L)\frac{t^2 c^*}{2} - \tau_0'(x_L)c^* t\right)(1), \end{aligned} \tag{A.12}$$

where the distribution relation follows from the fact that the $D^L_{[0,\infty)}$ is a linear operator for a linear function of t.

Finally, we analyze $n^{1/3}\{\hat{\tau}(q_n) - \tau_0(q_n)\}$. Recall q_n is the $(n^{-1/3})$-th sample quantile of X. Assumption A3 guarantees $q_n - q_n^* = O_p(n^{-1/2}) = o_P(n^{-1/3})$, which also implies $\text{plim}_{n\to\infty} n^{1/3}(q_n - x_L) = \lim_{n\to\infty} n^{1/3}(q_n^* - x_L) = c^*$. Thus, the same argument for (A.11) can be applied to show that the result in (A.11) holds true even if we replace q_n^* with q_n. Therefore, the conclusion follows.

A.2. **Proof of Theorem 1.** By the definitions of the estimators, it holds

$$\begin{aligned} \sqrt{n}(\hat{\theta} - \theta) \; &= \; \left(\frac{1}{n}\sum_{i: x_i > q_n} \hat{\sigma}_i^{-2} W_i W_i'\right)^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i: x_i > q_n} \hat{\sigma}_i^{-2} W_i U_i\right), \\ \sqrt{n}(\hat{\theta}_{\text{IGLS}} - \theta) \; &= \; \left(\frac{1}{n}\sum_{i=1}^n \sigma_i^{-2} W_i W_i'\right)^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \sigma_i^{-2} W_i U_i\right). \end{aligned}$$

Thus it is sufficient for the conclusion to show

$$T_1 := \frac{1}{\sqrt{n}} \sum_{i:x_i > q_n} \hat{\sigma}_i^{-2} W_i U_i - \frac{1}{\sqrt{n}} \sum_{i=1}^n \sigma_i^{-2} W_i U_i \xrightarrow{p} 0,$$

$$T_2 := \frac{1}{n} \sum_{i:x_i > q_n} \hat{\sigma}_i^{-2} W_i W_i' - \frac{1}{n} \sum_{i=1}^n \sigma_i^{-2} W_i W_i' \xrightarrow{p} 0.$$

By the consistency of $\hat{\sigma}^2(\cdot)$, $\frac{1}{n}\sum_{i=1}^n \sigma_i^{-2} W_i W_i' \xrightarrow{p} E[\sigma_i^{-2} W_i W_i']$, and $q_n \xrightarrow{p} x_L$, it holds $T_2 \xrightarrow{p} 0$. Therefore, we focus on the proof of $T_1 \xrightarrow{p} 0$. Decompose

$$T_1 = \frac{1}{\sqrt{n}} \sum_{i:x_i > q_n} (\hat{\sigma}_i^{-2} - \sigma_i^{-2}) W_i U_i - \frac{1}{\sqrt{n}} \sum_{i:x_i \le q_n} \sigma_i^{-2} W_i U_i =: T_{11} - T_{12}.$$

We first consider T_{12}. Let W_i^h and T_{12}^h be the h-th element of W_i and T_{12}, respectively. Note that $E[T_{12}^h | q_n] = 0$ by $E[U|W] = 0$. Also we have

$$\mathrm{Var}(T_{12}^h | q_n) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{X_i \le q_n\} E\left[\frac{W_i^h U_i}{\sigma^2(X_i)}\right]^2 \xrightarrow{p} 0,$$

where the convergence follows by the facts that (i) both W_i and $\frac{1}{\sigma^2(X_i)}$ are bounded (by Assumptions A1 and A2), so $E\left[\frac{W_i^h U_i}{\sigma^2(X_i)}\right]^2 \le \max_i \left(\frac{W_i^h}{\sigma^2(X_i)}\right)^2 E[U_i^2] < \infty$, and (ii) $\frac{1}{n}\sum_{i=1}^n \mathbb{I}\{X_i \le q_n\} = O_p(n^{-1/3})$ by the definition of q_n. Thus, we obtain $T_{12} \xrightarrow{p} 0$.

Now we focus on T_{11}. Since the proof is similar, we only present the proof for one element of T_{11}, i.e., for any constant $A > 0$,

$$P\{|\mathbb{G}_n \hat{f}| \ge A\} \to 0, \tag{A.13}$$

where $\hat{f}(w,u) = \mathbb{I}\{x > q_n\}\left(\frac{1}{\hat{\tau}(x)} - \frac{1}{\tau_0(x)}\right) w_h u$ for the h-th element of w. We set $\tau_0(x_L) = C_0 = 2K_0 > 0$. Pick any $A > 0$ and $\nu > 0$. Let $r_n = (\log n)^2 n^{-2/3}$. By Lemma 3 below, there exists a positive constant C such that

$$\begin{aligned} P\{|\mathbb{G}_n \hat{f}| \ge A\} &\le P\left\{|\mathbb{G}_n \hat{f}| \ge A,\ \|\hat{\tau}\|_\infty \le C\log n,\ \|\hat{\tau} - \tau_0\|_{2,P}^2 \le C r_n,\ \frac{\mathbb{I}\{x > q_n\}}{\hat{\tau}(x)} \le \frac{1}{K_0}\right\} + \frac{\nu}{2} \\ &\le \frac{E[\|\mathbb{G}_n\|_{\mathcal{F}}]}{A} + \frac{\nu}{2} \le \nu, \end{aligned} \tag{A.14}$$

for all n large enough, where the first inequality follows from Lemmas 1 and 3 (i)-(ii), and the fact that $\hat{\tau}$ is monotone increasing (so that the lower bound at the truncation point is the uniform lower bound), and the second inequality follows from Markov's inequality with

$$\mathcal{F}_n := \left\{ f_n(w,u) = \mathbb{I}\{x > q_n\}\left(\frac{1}{\tau(x)} - \frac{1}{\tau_0(x)}\right) w_h u : \begin{array}{c} \tau \ge 0 \text{ is monotone increasing on } \mathcal{X}, \\ \|\tau\|_\infty \le C\log n,\ \|\tau - \tau_0\|_{2,P}^2 \le C r_n, \\ \mathbb{I}\{x > q_n\}/\tau(x) \le 1/K_0 \end{array} \right\}.$$

and the last inequality follows from Lemma 3 (iii). Since ν can be arbitrarily small, we obtain (A.13) and the conclusion follows.

Lemma 3. *Under Assumptions A1-A3, it holds*

(i): $\|\hat{\tau}\|_\infty = O_p(\log n)$,

(ii): $\|\hat{\tau} - \tau_0\|_{2,P}^2 = O_p((\log n)^2 n^{-2/3})$,

(iii): $E[\|\mathbb{G}_n\|_{\mathcal{F}}] \leq \frac{A\nu}{2}$.

Proof of Lemma 3 (i). The proof is based on Balabdaoui, Durot and Jankowski (2019, Lemma 7.1) (BDJ hereafter). The min-max formula of the isotonic regression says

$$\min_{1\leq k\leq n} \frac{\sum_{j=1}^{k} \hat{U}_j^2}{k} \leq \hat{\tau}(x) \leq \max_{1\leq k\leq n} \frac{\sum_{j=k}^{n} \hat{U}_j^2}{n-k+1},$$

for each $x \in \mathcal{X}$, which implies $\min_{1\leq j\leq n} \hat{U}_j^2 \leq \hat{\tau}(x) \leq \max_{1\leq j\leq n} \hat{U}_j^2$ for each $x \in \mathcal{X}$. Thus, it is sufficient for the conclusion to show that

$$\max_{1\leq j\leq n} \hat{U}_j^2 = O_p(\log n). \tag{A.15}$$

Observe that

$$\max_{1\leq j\leq n} \hat{U}_j^2 \leq \max_{1\leq j\leq n} U_j^2 + 2Rk||\hat{\theta}_{\text{OLS}} - \theta||_\infty \max_{1\leq j\leq n} |U_j| + R^2k^2||\hat{\theta}_{\text{OLS}} - \theta||_\infty^2.$$

From BDJ (2019, eq. (7.11)), Assumption A2 guarantees $\max_{1\leq j\leq n} U_j^2 = O_p(\log n)$. Since $\hat{\theta}_{\text{OLS}}$ is the OLS estimator, $||\hat{\theta}_{\text{OLS}} - \theta||_\infty = O_p(n^{-1/2})$. We also have $\max_{1\leq j\leq n} |U_j| = O_p(\log n)$. Combining these results with Assumption A1, we have (A.15).

Proof of Lemma 3 (ii). The proof is based on that of Proposition 4 of BGH (p.8 of BGH-supp). Recall that $\hat{\tau}(\cdot)$ is the solution of $\min_{\tau\in\{\text{all monotone funcitons}\}} \sum_{j=1}^{n} \{\hat{U}_j^2 - \tau(X_j)\}^2$, or equivalently

$$\max_{\tau\in\{\text{all monotone funcitons}\}} \sum_{j=1}^{n} \{2\hat{U}_j^2 \tau(X_j) - \tau(X_j)\}. \tag{A.16}$$

On the other hand, $\tau_0(\cdot)$ is the solution of $\min_{\tau\in\{\text{all monotone funcitons}\}} E[\{U^2 - \tau(X)\}^2]$, or equivalently

$$\max_{\tau\in\{\text{all monotone funcitons}\}} E[2U^2\tau(X) - \tau(X)^2]. \tag{A.17}$$

By (A.16), it holds

$$\sum_{j=1}^{n} \{2\hat{U}_j^2 \hat{\tau}(X_j) - \hat{\tau}(X_j)^2\} \geq \sum_{j=1}^{n} \{2\hat{U}_j^2 \tau_0(X_j) - \tau_0(X_j)^2\},$$

or equivalently (by plugging in $\hat{U}_j = U_j - W_j'(\hat{\theta}_{\text{OLS}} - \theta)$),

$$\begin{aligned} &\sum_{j=1}^{n} \{2U_j^2 \hat{\tau}(X_j) - \hat{\tau}(X_j)^2\} + 2\sum_{j=1}^{n} \left(-2U_j W_j'(\hat{\theta}_{\text{OLS}} - \theta) + \{W_j'(\hat{\theta}_{\text{OLS}} - \theta)\}^2\right) \{\hat{\tau}(X_j) - \tau_0(X_j)\} \\ \geq\ & \sum_{j=1}^{n} \{2U_j^2 \tau_0(X_j) - \tau_0(X_j)^2\}. \end{aligned} \tag{A.18}$$

Note that for any monotone function τ,

$$\begin{aligned} & E[2U^2\tau(X) - \tau(X)^2] - E[2U^2\tau_0(X) - \tau_0(X)^2] \\ =\ & E[2E[U^2|X]\tau(X) - \tau(X)^2 - 2E[U^2|X]\tau_0(X) + \tau_0(X)^2] \\ =\ & E[2\tau_0(X)\tau(X) - \tau(X)^2 - \tau_0(X)^2] = -d_2^2(\tau, \tau_0), \end{aligned} \tag{A.19}$$

where the first equality follows from the law of iterated expectation, the second equality follows from the definition $\tau_0(x) = E[U^2|X = x]$, and the last equality follows from the definition of $d_2^2(\cdot,\cdot)$.

Define

$$\begin{aligned} g_\tau(u,x) &= \{2u^2\tau(x) - \tau(x)^2\} - \{2u^2\tau_0(x) - \tau_0(x)^2\}, \\ R_n &= \frac{2}{n}\sum_{j=1}^{n}\left(-2U_jW_j'(\hat\theta_{\mathrm{OLS}} - \theta) + \{W_j'(\hat\theta_{\mathrm{OLS}} - \theta)\}^2\right)\{\hat\tau(X_j) - \tau_0(X_j)\}. \end{aligned}$$

From (A.18) and (A.19), it holds

$$\int g_{\hat\tau}(u,x)d(\mathbb{P}_n - P)(u,x) + R_n \geq d_2^2(\hat\tau,\tau_0). \tag{A.20}$$

Note that R_n is bounded as

$$|R_n| \leq -(\hat\theta_{\mathrm{OLS}} - \theta)'\frac{4}{n}\sum_{j=1}^{n}W_jU_j\{\hat\tau(X_j) - \tau_0(X_j)\} + \frac{2}{n}\sum_{j=1}^{n}\{W_j'(\hat\theta_{\mathrm{OLS}} - \theta)\}^2\{\hat\tau(X_j) - \tau_0(X_j)\}.$$

The second term is of order $O_p(n^{-1}\log n)$ (because $\hat\theta_{\mathrm{OLS}} - \theta = O_p(n^{-1/2})$ and Lemma 3 (i)). By similar arguments in p.22 of BGH-supp and in the proof of Lemma 3 (i), the first term is of order $O_p\left((\log n)^2 n^{-\frac{5}{6}}\right)$.

Then

$$R_n = O_p\left((\log n)^2 n^{-\frac{5}{6}}\right). \tag{A.21}$$

Thus, for some constants $C, K > 0$ and a shrinking sequence ϵ_n, set inclusion relationships yield

$$\begin{aligned} P(d_2^2(\hat\tau,\tau_0) \geq \epsilon_n^2) &= P\left(d_2(\hat\tau,\tau_0) \geq \epsilon_n,\ \int g_{\hat\tau}(u,x)d(\mathbb{P}_n - P)(u,x) + R_n \geq d_2^2(\hat\tau,\tau_0)\right) \\ &\leq P\begin{pmatrix} d_2(\hat\tau,\tau_0) \geq \epsilon_n,\ |R_n| \leq C(\log n)^2 n^{-\frac{5}{6}},\ \|\hat\tau\|_\infty \leq K\log n \\ \int g_{\hat\tau}(u,x)d(\mathbb{P}_n - P)(u,x) + R_n - d_2^2(\hat\tau,\tau_0) \geq 0 \end{pmatrix} \\ &\quad + P(|R_n| > C(\log n)^2 n^{-\frac{5}{6}}) + P(\|\hat\tau\|_\infty > K\log n) \\ &=: P_1 + P_2 + P_3, \end{aligned}$$

where the first equality follows from (A.20). For P_2 and P_3, (A.21) and Lemma 3 (i) imply that we can choose C and K to make these terms arbitrarily small. Thus, we focus on the first term P_1.

Now let

$$\begin{aligned} \mathcal{T} &= \{\tau : \tau \text{ is positive and monotne increasing on } \mathcal{X},\ \|\tau\|_\infty \leq K\log n\}, \\ \mathcal{G} &= \{g_\tau(u,x) = \{2u^2\tau(x) - \tau(x)^2\} - \{2u^2\tau_0(x) - \tau_0(x)^2\} :\ \tau \in \mathcal{T}\}, \\ \mathcal{G}_v &= \{g \in \mathcal{G} : d_2(\tau,\tau_0) \leq v\}. \end{aligned}$$

Set inclusion relationships and Markov's inequality yield

$$\begin{aligned}
P_1 &\leq P\left(\sup_{\tau\in\mathcal{T},d_2(\tau,\tau_0)\geq\epsilon_n}\{\int g_\tau(u,x)d(\mathbb{P}_n-P)(u,x)-d_2^2(\tau,\tau_0)\}\geq -C(\log n)^2 n^{-\frac{5}{6}}\right)\\
&\leq \sum_{s=0}^{\infty}P\left(\sup_{\tau\in\mathcal{T},2^s\epsilon_n\leq d_2(\tau,\tau_0)\leq 2^{s+1}\epsilon_n}\sqrt{n}\{\int g_\tau(u,x)d(\mathbb{P}_n-P)(u,x)\}\geq\sqrt{n}\left(2^{2s}\epsilon_n^2-C(\log n)^2n^{-\frac{5}{6}}\right)\right)\\
&\leq \sum_{s=0}^{\infty}P\left(\|\mathbb{G}_n g\|_{\mathcal{G}_{2^{s+1}\epsilon_n}}\geq\sqrt{n}\left(2^{2s}\epsilon_n^2-C(\log n)^2n^{-\frac{5}{6}}\right)\right)\\
&\leq \sum_{s=0}^{\infty}E[\|\mathbb{G}_n g\|_{\mathcal{G}_{2^{s+1}\epsilon_n}}]/\{\sqrt{n}\left(2^{2s}\epsilon_n^2-C(\log n)^2n^{-\frac{5}{6}}\right)\},
\end{aligned}$$

where $\mathcal{G}_\delta=\{g\in\mathcal{G}:\|g\|_P\leq\delta\}$. For any constant $\tilde{C}>0$, the sequence $\epsilon_n^2=\tilde{C}(\log n)^2n^{-\frac{2}{3}}$ will dominate $C(\log n)^2n^{-\frac{5}{6}}$, so it holds $\sqrt{n}\left(2^{2s}\epsilon_n^2-Cn^{-\frac{5}{6}}\log^2 n\right)=\sqrt{n}2^{2s}\epsilon_n^2(1+o(1))$. Therefore, the standard result for the L^2-convergence of the isotonic estimator under Assumption A2 (e.g., pp. 8-11 in BGH-supp) implies the last term converges to zero.

A.2.1. *Proof of Lemma 3 (iii).* We show $E[\|\mathbb{G}_n\|_{\mathcal{F}}]\leq\frac{A\nu}{2}$ by using van der Vaart and Wellner (1996, Lemma 3.4.3). First we introduce some notation for this part. Let $N_{[]}(\varepsilon,\mathcal{F},||\cdot||)$ be the ε-bracketing number of the function class $\mathcal{F}$ under the norm $||\cdot||$, $H_B(\varepsilon,\mathcal{F},||\cdot||)=\log N_{[]}(\varepsilon,\mathcal{F},||\cdot||)$ be the entropy, $J_n(\delta,\mathcal{F},||\cdot||)=\int_0^\delta\sqrt{1+H_B(\varepsilon,\mathcal{F},||\cdot||)}d\varepsilon$, and $\|f\|_{B,P}=(2E[e^{|f|}-|f|-1])^{1/2}$ be the Bernstein norm.

Lemma 3.4.3 in van der Vaart and Wellner (1996): Let $\mathcal{F}$ be a class of measurable functions such that $\|f\|_{B,P}^2\leq\delta$ for every f in $\mathcal{F}$. Then

$$E[\|\mathbb{G}_n\|_{\mathcal{F}}]\lesssim J_n(\delta,\mathcal{F},||\cdot||_{B,P})\{1+J_n(\delta,\mathcal{F},||\cdot||_{B,P})/(\sqrt{n}\delta^2)\}.$$

To apply this lemma, we need to compute $H_B(\epsilon,\tilde{\mathcal{F}},\|\cdot\|_{B,P})$ and $\|\tilde{f}\|_{B,P}^2$, where $\tilde{\mathcal{F}}=\{\tilde{f}=D^{-1}f:f\in\mathcal{F}\}$, and the constant $D>0$ will be defined later to guarantee that the Bernstein norm of $\tilde{f}$ is finite. Let us define

$$\mathcal{T}_{\mathcal{I},K_1}=\{\tau\text{ monotone non-decreasing on some interval }\mathcal{I}\text{ and }0<\tau<K_1\}.$$

Assumption A2 implies $0<\underline{C}<\tau_0<\overline{C}<\infty$. Also let

$$\mathcal{F}_n=\left\{f_n(w,u)=\mathbb{I}\{x>q_n\}\left(\frac{1}{\tau(x)}-\frac{1}{\tau_0(x)}\right)w_h u:\begin{array}{c}\tau\in\mathcal{T}_{\mathcal{X},K_1},\ \|\tau-\tau_0\|_{2,P}^2\leq v,\\ \mathbb{I}\{x>q_n\}/\tau(x)\leq 1/K_0.\end{array}\right\},\quad\text{(A.22)}$$

where w_h is the h-th component of w. We set $2K_0=\underline{C}$, $K_1=K_2\log n$, and $v=K_3(\log n)n^{-1/3}$ for some constants $K_2,K_3>0$.

Consider ϵ-brackets (τ^L,τ^U) under the $L_2(P)$-norm of the functions $x\mapsto\tau(x)$. By van der Vaart and Wellner (1996, Theorem 2.7.5), there exists some constant $C'>0$ such that

$$H_B(\epsilon,\mathcal{T}_{\mathcal{X},K_1},\|\cdot\|_P)\leq\frac{C'K_1}{\epsilon},\quad\text{for each }\epsilon\in(0,K_1).\qquad\text{(A.23)}$$

Without loss of generality, we can choose those bracket functions which satisfy $\mathbb{I}\{x > q_n\}/\tau^L(x) \leq 1/K_0$.[1] Define

$$f^L(w,u) = \begin{cases} \mathbb{I}\{x > q_n\}\left(\frac{1}{\tau^U(x)} - \frac{1}{\tau_0(x)}\right) w_h u & \text{if } w_h u \geq 0, \\ \mathbb{I}\{x > q_n\}\left(\frac{1}{\tau^L(x)} - \frac{1}{\tau_0(x)}\right) w_h u & \text{if } w_h u < 0, \end{cases}$$

$$f^U(w,u) = \begin{cases} \mathbb{I}\{x > q_n\}\left(\frac{1}{\tau^L(x)} - \frac{1}{\tau_0(x)}\right) w_i u & \text{if } w_i u \geq 0, \\ \mathbb{I}\{x > q_n\}\left(\frac{1}{\tau^U(x)} - \frac{1}{\tau_0(x)}\right) w_i u & \text{if } w_i u < 0. \end{cases}$$

Note that (f^L, f^U) is a bracket of $f \in \mathcal{F}$ for every $q_n \in [x_L, x_U]$.

Now we compute the bracket size of $(\tilde{f}^L, \tilde{f}^U) := (D^{-1}f^L, D^{-1}f^U)$ with respect to the Bernstein norm. Note that

$$\begin{aligned} &\|\tilde{f}^U - \tilde{f}^L\|_{B,P}^2 = \|D^{-1}f^U - D^{-1}f^L\|_{B,P}^2 \\ \leq\ & 2\sum_{k=2}^{\infty} \frac{1}{k!D^k} \int_{\mathscr{W}\times\mathbb{R}} \left| \frac{\tau^U(x) - \tau^L(x)}{\tau^L(x)\tau^U(x)} w_h u \right|^k dP(w,u) \\ \leq\ & 2\sum_{k=2}^{\infty} \frac{1}{k!D^k} \left\{ \frac{R^k k! M_0^{k-2} a_0 (2K_1)^{k-2}}{K_0^{2k}} \|\tau^U - \tau^L\|_P^2 \right\} \leq 2a_0 \left(\frac{R}{DK_0^2}\right)^2 \sum_{k=0}^{\infty} \left(\frac{2RM_0K_1}{DK_0^2}\right)^k \epsilon^2, \end{aligned}$$

where the first inequality follows from the definition of $\|\cdot\|_{B,P}^2$ and $\mathbb{I}\{x > q_n\} \leq 1$, the second inequality follows from Assumption A2 (where we can choose $a_0, M_0 > 1$) and $\frac{\mathbb{I}\{x>q_n\}}{\tau^L(x)} \leq \frac{1}{K_0}$. Thus, by setting $D = 4M_0RK_1/K_0^2$, we obtain$\|\tilde{f}^U - \tilde{f}^L\|_{B,P}^2 \leq \frac{a_0}{4M_0^2K_1^2}\epsilon^2$, which implies

$$\|\tilde{f}^U - \tilde{f}^L\|_{B,P} \leq \tilde{K}\epsilon,$$

for $\tilde{K} = \frac{a_0^{1/2}}{2M_0K_1}$. Now $(\tilde{f}^L, \tilde{f}^U)$ is: (a) a set of brackets in $\tilde{\mathcal{F}}$, (b) one-to-one induced by (τ^L, τ^U), an ϵ-bracket in $\mathcal{T}_{\mathcal{X},K_1}$ with the bracket number $H_B(\epsilon, \mathcal{T}_{\mathcal{X},K_1}, \|\cdot\|_P)$, and (c) $\|\tilde{f}^U - \tilde{f}^L\|_{B,P} \leq \tilde{K}\epsilon$. Based on these facts, (A.23) yields

$$H_B(\tilde{K}\epsilon, \tilde{\mathcal{F}}, \|\cdot\|_{B,P}) \leq H_B(\epsilon, \mathcal{T}_{\mathcal{X},K_1}, \|\cdot\|_P) \leq \frac{C'K_1}{\epsilon},$$

which implies (by a change of variable argument)

$$H_B(\epsilon, \tilde{\mathcal{F}}, \|\cdot\|_{B,P}) \leq \frac{\tilde{K}C'K_1}{\epsilon} = \frac{\tilde{B}}{\epsilon}, \quad \text{for } \tilde{B} = \frac{C'a_0^{1/2}}{2M_0}. \tag{A.24}$$

[1] By definition (A.22), the $\tau(\cdot)$ in $\mathcal{F}$ must satisfy $\mathbb{I}\{x > q_n\}/\tau(x) \leq 1/K_0$. Since $\mathcal{T}_{\mathcal{X},K_1}$ is a class of monotone increasing function, any ϵ-brackets of $\mathcal{T}_{\mathcal{X},K_1}$ can be modified to be a ϵ-bracket of the "$\mathcal{F}$-subset" of $\mathcal{T}_{\mathcal{X},K_1}$ satisfying $\mathbb{I}\{x > q_n\}/\tau(x) \leq 1/K_0$ by leveling-up certain part of lower bounds functions τ^L, without changing the bracket numbers, and the size of each bracket can only be smaller.

We now characterize the Bernstein norm of $\tilde{f}$, that is

$$\begin{aligned}\|\tilde{f}\|_{B,P}^2 &\leq 2\sum_{k=2}^{\infty}\frac{1}{k!D^k}\int_{\mathscr{W}\times\mathbb{R}}\left|\frac{\tau(x)-\tau_0(x)}{\tau(x)\tau_0(x)}w_h u\right|^k dP(w,u)\\ &\leq 2\sum_{k=2}^{\infty}\frac{1}{k!D^k}\left\{\frac{R^k k! M_0^{k-2}a_0(2K_1)^{k-2}}{K_0^{2k}}\|\tau-\tau_0\|_P^2\right\}\\ &\leq 2a_0\left(\frac{R}{DK_0^2}\right)^2\sum_{k=0}^{\infty}\left(\frac{2RM_0K_1}{DK_0^2}\right)^k v^2 \leq \frac{a_0}{4M_0^2}\frac{1}{K_1^2}v^2,\end{aligned}$$

where the second inequality follows from $\frac{\mathbb{I}\{x>q_n\}}{\tau(x)} \leq \frac{1}{K_0}$. Therefore, we have

$$\|\tilde{f}\|_{B,P} \leq \frac{Bv}{K_1}, \quad \text{for } B = \frac{a_0^{1/2}}{2M_0}. \tag{A.25}$$

Combining (A.24) and (A.25), van der Vaart and Wellner (1996, Lemma 3.4.3) implies

$$E[\|\mathbb{G}_n\|_{\tilde{\mathcal{F}}}] \lesssim J_n(BK_1^{-1}v)\left(1+\frac{J_n(BK_1^{-1}v)}{\sqrt{n}B^2v^2/K_1^2}\right),$$

where $J_n(\cdot)$ is the abbreviation of $J_n(\cdot,\tilde{\mathcal{F}},\|\cdot\|_{B,P})$. By the arguments used in the proof of Proposition 7.9 of BDJ, it holds

$$J_n(BK_1^{-1}v) \leq BK_1^{-1}v + 2\tilde{B}^{1/2}B^{1/2}K_1^{-1/2}v^{1/2} \lesssim B_1K_1^{-1/2}v^{1/2},$$

for some $B_1 > 0$ and sufficiently small v, which implies

$$E[\|\mathbb{G}_n\|_{\tilde{\mathcal{F}}}] \lesssim B_1K_1^{-1/2}v^{1/2}\left(1+K_1^2\frac{B_1K_1^{-1/2}v^{1/2}}{\sqrt{n}B^2v^2}\right) \lesssim B_1K_1^{-1/2}v^{1/2}\left(1+\frac{B_2K_1^{3/2}}{\sqrt{n}v^{3/2}}\right),$$

for some $B_2 > 0$. By the definition of the class $\tilde{\mathcal{F}} = \{\tilde{f} = D^{-1}f : f \in \mathcal{F}\}$, it follows that

$$E[\|\mathbb{G}_n\|_{\mathcal{F}}] = D\cdot E[\|\mathbb{G}_n\|_{\tilde{\mathcal{F}}}] \lesssim DB_1K_1^{-1/2}v^{1/2}\left(1+\frac{B_2K_1^{3/2}}{\sqrt{n}v^{3/2}}\right) \lesssim B_3K_0^{-2}K_1^{1/2}v^{1/2}\left(1+\frac{B_2K_1^{3/2}}{\sqrt{n}v^{3/2}}\right),$$

for some $B_3 > 0$. Observe that with $v = K_3(\log n)n^{-1/3}$ and $K_1 = K_2\log n$,

$$E[\|\mathbb{G}_n\|_{\mathcal{F}}] \lesssim C_3(\log n)n^{-1/6}(1+C_4) \lesssim \frac{A\nu}{2},$$

where $C_3 = B_3K_0^{-2}K_1^{1/2}K_3^{1/2}$ and $C_4 = B_2(K_2/K_3)^{3/2}$, and A and ν can be arbitrary small positive numbers. Therefore, the conclusion follows.

Appendix B. Proof of lemma and theorem in Section 3

Notation. To avoid heavy notations, some of them are used in Appendix A but redefined here. Let $\|f\|_\infty = \sup_{x\in\mathcal{X}}|f(x)|$ be the sup-norm, $\|f\|_{2,P} = \sqrt{\int |f(x)|^2dP}$ be the $L_2(P)$ norm, $\mathbb{G}_n$ be the empirical process, i.e., $\mathbb{G}_nf = \frac{1}{\sqrt{n}}\sum_{i=1}^n\{f(x_i)-E[f(x_i)]\}$, and $\|\mathbb{G}_n\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}}|\mathbb{G}_nf|$. Define $\tau_\eta(a) = E[\sigma^2(X'\eta_0)|X'\eta = a]$ and $\tau_{\eta_0}(a) = \tau_0(a)$ (note that $\tau_0(x'\eta_0) = \sigma^2(x'\eta_0)$). Let $\hat{\tau}_\eta = \hat{\tau}_\eta(x'\eta)$ be the isotonic estimator obtained by (3.2) for given η, $D_A^L[f](a)$ be the left derivative of the greatest convex minorant of function $f(\cdot)$ evaluated at $a \in A$, $F_n(t) = \frac{1}{n}\sum_1^n \mathbb{I}\{X_i'\hat{\eta} \leq t\}$, and $M_n(t) = \frac{1}{n}\sum_{i=1}^n \hat{U}_i^2 \mathbb{I}\{X_i'\hat{\eta} \le t\}$. Based on $\tau_0(x'\eta_0) = E[U^2|X = x]$, define $\varepsilon = U^2 - \tau_0(X'\eta_0)$ and $\sigma_\varepsilon^2(x) = E[\varepsilon^2|X = x]$. We also redefine $x_L = \inf_{x\in\mathcal{X}}(x'\eta_0)$ and $x_U = \sup_{x\in\mathcal{X}}(x'\eta_0)$.

B.1. **Proof of Lemma 2.** The main part of the proof is similar to that of Lemma 1. Recall that q_n^* is the $(n^{-1/3})$-th population quantile of $(X'\eta_0)$ and q_n is the $(n^{-1/3})$-th sample quantile of $\{X_i'\hat{\eta}\}_{i=1}^n$ with $\hat{\eta}$ estimated by (3.3). To proceed, we use the following lemma.

Lemma 4. *Under Assumptions M1-M3, it holds*

(i): $\hat{\eta} - \eta_0 = O_p(n^{-1/2})$,
(ii): $\tau_{\hat{\eta}}(a) - \tau_0(a) = O_p(n^{-1/2})$ *for each* $a \in I_{\hat{\eta}}$.

The proof of this lemma is in Appendix B.3. Based on Lemma 4 (i), Assumptions M2-M3, and properties of the sample quantile, we obtain $q_n - q_n^* = O_p(n^{-1/2}) = o_p(n^{-1/3})$, which implies $c^* = \lim_{n\to\infty} n^{1/3}(q_n^* - x_L) = \text{plim}_{n\to\infty} n^{1/3}(q_n - x_L) < \infty$. By Assumption M2, Lemma 4 (ii), and similar arguments in Appendix A.1, we have

$$\begin{aligned}
& n^{1/3}\{\hat{\tau}_{\hat{\eta}}(q_n) - \tau_0(q_n)\} = n^{1/3}\{\hat{\tau}_{\hat{\eta}}(q_n) - \tau_{\hat{\eta}}(q_n)\} + o_p(1) \\
= \;& n^{1/3}[\{\hat{\tau}_{\hat{\eta}}(q_n) - \tau_{\hat{\eta}}(x_L)\} - \{\tau_{\hat{\eta}}(q_n) - \tau_0(x_L)\}] + o_p(1) \\
\overset{d}{\to} \;& D^L_{[0,\infty)}\left(\sqrt{\frac{\sigma_\varepsilon^2(x_L)}{c^* f(x_L)}}\mathcal{W}_t + \tau_0'(x_L)\frac{t^2 c^*}{2}\right)(1) - \underset{n\to\infty}{\text{plim}} n^{1/3}\{\tau_0(q_n) - \tau_0(x_L)\} \\
\overset{d}{\sim} \;& D^L_{[0,\infty)}\left(\sqrt{\frac{\sigma_\varepsilon^2(x_L)}{c^* f(x_L)}}\mathcal{W}_t + \tau_0'(x_L)\frac{t^2 c^*}{2}\right)(1) - \lim_{n\to\infty} n^{1/3}\{\tau_0(q_n^*) - \tau_0(x_L)\} \\
\overset{d}{\sim} \;& D^L_{[0,\infty)}\left(\sqrt{\frac{\sigma_\varepsilon^2(x_L)}{c^* f(x_L)}}\mathcal{W}_t + \tau_0'(x_L)\frac{t^2 c^*}{2} - \tau_0'(x_L)c^* t\right)(1),
\end{aligned}$$

where the first and second equalities follow from Lemma 4 (ii), the convergence follows from a similar argument to (A.12), the first distribution relation follows from Lemma 4 (ii), Assumption M2(iv), and $q_n^* - q_n = o_p(n^{-1/3})$, and the second distribution relation follows from the fact that the $D^L_{[0,\infty)}$ is a linear operator for a linear function of t.

B.2. **Proof of Theorem 2.** Define

$$\mathcal{F}_n = \left\{ f_n(w,u) = \mathbb{I}\{x'\eta > q_n\}\left(\frac{1}{\tau(x'\eta)} - \frac{1}{\tau_\eta(x'\eta)}\right) wu : \begin{array}{c} \tau \ge 0 \text{ is monotone increasing on } [-R,R], \\ \|\tau\|_\infty \le C\log n, \;\; \|\tau - \tau_\eta\|_{2,P}^2 \le C r_n, \\ \mathbb{I}(x'\eta > q_n)/\tau(x'\eta) \le 1/K_0. \end{array} \right\},$$

for some constant $C > 0$, where R is the positive constant defined in Assumption M1. Similar to Theorem 1, it is sufficient for the conclusion to prove the following lemma.

Lemma 5. *Under Assumptions M1-M6, it holds*

(i): $\|\hat{\tau}_\eta\|_\infty = O_p(\log n)$ *uniformly over* $\eta \in \mathcal{B}(\eta_0, \delta_0)$,
(ii): $\|\hat{\tau}_{\hat{\eta}} - \tau_0\|_{2,P}^2 = O_p((\log n)^2 n^{-2/3})$,
(iii): $E\left[\|\mathbb{G}_n\|_{\mathcal{F}}\right] \le \frac{A\nu}{2}$.

B.2.1. *Proof of Lemma 5 (i).* The proof is similar to that of BDJ's Lemma 7.1. For fixed η, let $\{\hat{U}_{\eta,i}^2\}_{i=1}^n$ be a permutation of $\{\hat{U}_j^2\}_{j=1}^n$, which is arranged according to the monotonically ordered series $\{X_i'\eta\}_{i=1}^n$. The min-max formula of the isotonic regression says

$$\min_{1\leq k\leq n}\frac{\sum_{i=1}^k \hat{U}_{\eta,i}^2}{k}\leq \hat{\tau}_\eta(x'\eta)\leq \max_{1\leq k\leq n}\frac{\sum_{i=k}^n \hat{U}_{\eta,i}^2}{n-k+1},$$

for each $x\in\mathcal{X}$ and $\eta\in\mathcal{B}(\eta_0,\delta_0)$, which implies $\min_{1\leq j\leq n}\hat{U}_j^2\leq\hat{\tau}_\eta(x'\eta)\leq\max_{1\leq j\leq n}\hat{U}_j^2$ for each $x\in\mathcal{X}$. Thus, it is sufficient for the conclusion to show that

$$\max_{1\leq j\leq n}\hat{U}_j^2=O_p(\log n). \tag{B.1}$$

Observe that

$$\max_{1\leq j\leq n}\hat{U}_j^2\leq\max_{1\leq j\leq n}U_j^2+2Rk||\hat{\theta}_{\text{OLS}}-\theta||_\infty\max_{1\leq j\leq n}|U_j|+R^2k^2||\hat{\theta}_{\text{OLS}}-\theta||_\infty^2,$$

where k is the dimension of θ. From Lemma 7.1 of BDJ, Assumption M2 guarantees $\max_{1\leq j\leq n}U_j^2=O_p(\log n)$, and we also have $\max_{1\leq j\leq n}|U_j|=O_p(\log n)$ and $||\hat{\theta}_{\text{OLS}}-\theta||_\infty=O_p(n^{-1/2})$. Thus, we have $\|\hat{\tau}_\eta\|_\infty=O_p(\log n)$. Since different η only changes the permutation $\{\hat{U}_{\eta,i}^2\}_{i=1}^n$ but not $\max_{1\leq j\leq n}\hat{U}_j^2$, we have $\|\hat{\tau}_\eta\|_\infty=O_p(\log n)$ uniformly over $\eta\in\mathcal{B}(\eta_0,\delta_0)$.

B.2.2. *Proof of Lemma 5 (ii).* The main part of the proof is similar to those of Lemma 3 (ii) and Proposition 4 of BGH-supp. Define

$$\begin{aligned}
g_{\eta,\tau}(u,x) &= \{2u^2\tau(x'\eta)-\tau(x'\eta)^2\}-\{2u^2\tau_\eta(x'\eta)-\tau_\eta(x'\eta)^2\},\\
R_{n,\eta} &= \frac{2}{n}\sum_{j=1}^n\left(-2U_jW_j(\hat{\theta}_{\text{OLS}}-\theta)+\{W_j(\hat{\theta}_{\text{OLS}}-\theta)\}^2\right)\{\hat{\tau}_\eta(X_j'\eta)-\tau_\eta(X_j'\eta)\},\\
d_2^2(\tau_1,\tau_2) &= -E[2\tau_1\tau_2-\tau_1^2-\tau_2^2],
\end{aligned}$$

where $R_{n,\eta}=O_p(n^{-1}\log n)$ with the same argument to (A.21), and here $\tau_\eta(x'\eta)$ plays the similar role to $\tau_0(x)$ in Lemma 3 (ii). We will discuss the rate of $\{\tau_{\hat{\eta}}(x'\hat{\eta})-\tau_0(x'\eta_0)\}$ in the end. With similar arguments to (A.16)-(A.21), we have for some C and K,

$$\begin{aligned}
&P(\sup_{\eta\in\mathcal{B}(\eta_0,\delta_0)}d_2^2(\hat{\tau}_\eta,\tau_\eta)\geq\epsilon_n^2)\\
\leq\;& P\begin{pmatrix}\sup_{\eta\in\mathcal{B}(\eta_0,\delta_0)}d_2(\hat{\tau}_\eta,\tau_\eta)\geq\epsilon_n,\;\;\sup_{\eta\in\mathcal{B}(\eta_0,\delta_0)}\|\hat{\tau}_\eta\|_\infty\leq K\log n,\\ \sup_{\eta\in\mathcal{B}(\eta_0,\delta_0)}\int g_{\eta,\hat{\tau}}(u,x)d(\mathbb{P}_n-P)(u,x)+R_{n,\eta}-d_2^2(\hat{\tau}_\eta,\tau_\eta)\geq 0,\\ \sup_{\eta\in\mathcal{B}(\eta_0,\delta_0)}|R_{n,\eta}|\leq Cn^{-1}\log n\end{pmatrix}\\
&+P(|R_{n,\eta}|>Cn^{-1}\log n)+P\left(\sup_{\eta\in\mathcal{B}(\eta_0,\delta_0)}\|\hat{\tau}_\eta\|_\infty>K\log n\right)\\
=:\;& P_1+P_2+P_3.
\end{aligned}$$

Now let

$$
\begin{aligned}
\mathcal{T} &= \{\tau : \tau \text{ is positive and monotone increasing function on } [-R, R],\ \|\tau\|_\infty \le K \log n\}, \\
\mathcal{G} &= \{g(x,u) = \{2u^2\tau(x'\eta) - \tau(x'\eta)^2\} - \{2u^2\tau_\eta(x'\eta) - \tau_\eta(x'\eta)^2\} :\ \tau \in \mathcal{T}\}, \\
\mathcal{G}_v &= \{g \in \mathcal{G} : d_2(\tau, \tau_\eta) \le v,\ \text{for all } \eta \in \mathcal{B}(\eta_0, \delta_0)\}.
\end{aligned}
$$

By a similar argument to Lemma 3 (ii) and Proposition 4 of BGH-supp, we have

$$
P_1 \le \sum_{s=0}^{\infty} E\left[\|\mathbb{G}_n g\|_{\mathcal{G}_{2^{s+1}\epsilon_n}}\right] / \{\sqrt{n} 2^{2s} \epsilon_n^2 - C n^{-1/2} \log n\}.
$$

A similar argument to the proof of Lemma 3 (iii) guarantees that P_1 can be arbitrarily small, and the conclusion is obtained. The proof of the previous subsection implies $P_3 \to 0$.

For $P_2 \to 0$, it remains to derive the convergence rate of $\hat{\tau}_{\hat{\eta}}$. By similar arguments to the proof of Proposition 4 of BGH-supp, we obtain

$$
\sup_{\eta \in \mathcal{B}(\eta_0, \delta_0)} \int \{\hat{\tau}_\eta(x'\eta) - \tau_\eta(x'\eta)\}^2 dF(x) = O_p((\log n)^2 n^{-2/3}). \tag{B.2}
$$

Thus, Lemma 4 (ii) and the triangle inequality imply $\|\hat{\tau}_{\hat{\eta}} - \tau_0\|_{2,P}^2 = O_p((\log n)^2 n^{-2/3})$.

Proof of Lemma 5 (iii). To avoid heavy notation, we use the same notation as in the proof of Lemma 3 (iii), but some notation is redefined here. Let

$$
\mathcal{T}_{\mathcal{I},K_1} = \{\tau \text{ monotone non-decreasing on some interval } \mathcal{I} \text{ and } 0 < \tau < K_1\}.
$$

Assumption M2 guarantees $0 < \underline{C} < \tau_0 < \overline{C} < \infty$. Similar to the proof of Lemma 3 (iii), we calculate $H_B(\epsilon, \tilde{\mathcal{F}}, \|\cdot\|_{B,P})$ and $\|\tilde{f}\|_{B,P}^2$, with $\tilde{\mathcal{F}} = \{\tilde{f} = D^{-1} f : f \in \mathcal{F}\}$, where the constant $D > 0$ is determined later. Define $I_\eta^* = (a^L, a^U)$ with $a^L = \inf_{x \in \mathcal{X}, \eta \in \mathcal{B}(\eta_0, \delta_0)} x'\eta$ and $a^U = \sup_{x \in \mathcal{X}, \eta \in \mathcal{B}(\eta_0, \delta_0)} x'\eta$. Define

$$
\mathcal{F}_n = \left\{ f_n(w,u) = \mathbb{I}\{x'\eta > q_n\} \left(\frac{1}{\tau(x'\eta)} - \frac{1}{\tau_\eta(x'\eta)} \right) w_h u : \begin{array}{c} \tau \in \mathcal{T}_{I_\eta^*, K_1}, \eta \in \mathcal{B}(\eta_0, \delta_0), \\ \|\tau - \tau_0\|_{2,P}^2 \le v, \\ \mathbb{I}(x'\eta > q_n)/\tau(x) \le 1/K_0. \end{array} \right\},
$$

where w_h is the h-th component of w. We set $2K_0 = \underline{C}$, $K_1 = K_2 \log n$, and $v = K_3 (\log n) n^{-1/3}$ for some positive constants K_2 and K_3.

By van der Vaart and Wellner (1996, Theorem 2.7.5), it holds for each $\epsilon \in (0, K_1)$,

$$
H_B(\epsilon, \mathcal{T}_{I_\eta^*, K_1}, \|\cdot\|_P) \le \frac{C' K_1}{\epsilon}.
$$

Similarly to the univariate case, we can choose those bracket functions (τ^L, τ^U), which satisfy $\mathbb{I}\{x'\eta > q_n\}/\tau(x'\eta) \le 1/K_0$. Define

$$f^L(w,u) = \begin{cases} \mathbb{I}\{x'\eta > q_n\}\left(\frac{1}{\tau^U(x'\eta)} - \frac{1}{\tau_\eta(x'\eta)}\right) w_h u & \text{if } w_h u \ge 0, \\ \mathbb{I}\{x'\eta > q_n\}\left(\frac{1}{\tau^L(x'\eta)} - \frac{1}{\tau_\eta(x'\eta)}\right) w_h u & \text{if } w_h u < 0, \end{cases}$$

$$f^U(w,u) = \begin{cases} \mathbb{I}\{x'\eta > q_n\}\left(\frac{1}{\tau^L(x'\eta)} - \frac{1}{\tau_\eta(x'\eta)}\right) w_h u & \text{if } w_h u \ge 0, \\ \mathbb{I}\{x'\eta > q_n\}\left(\frac{1}{\tau^U(x'\eta)} - \frac{1}{\tau_\eta(x'\eta)}\right) w_h u & \text{if } w_h u < 0. \end{cases}$$

Note that (f^L, f^U) is a bracket for $f \in \mathcal{F}_n$ and $\eta \in \mathcal{B}(\eta_0, \delta_0)$ whose bracket size is

$$\begin{aligned} &\|\tilde{f}^U - \tilde{f}^L\|_{B,P}^2 = \|D^{-1}f^U - D^{-1}f^L\|_{B,P}^2 \\ = \;& 2\sum_{k=2}^{\infty} \frac{1}{k!D^k}\int_{\mathscr{W}\times\mathbb{R}} \mathbb{I}\{x'\eta > q_n\}\left|\left(\frac{1}{\tau^L(x'\eta)} - \frac{1}{\tau^U(x'\eta)}\right) w_h u\right|^k dP(w,u) \\ \le \;& 2\sum_{k=2}^{\infty} \frac{1}{k!D^k}\left\{\frac{R^k k! M_0^{k-2} a_0 (2K_1)^{k-2}}{K_0^{2k}}\|\tau^U - \tau^L\|_P^2\right\} \\ \le \;& 2a_0\left(\frac{R}{DK_0^2}\right)^2 \sum_{k=0}^{\infty}\left(\frac{2RM_0K_1}{DK_0^2}\right)^k \epsilon^2, \end{aligned}$$

where the inequality follows from Assumption M2 (where we can choose $a_0, M_0 > 1$) and $\frac{\mathbb{I}\{x'\eta > q_n\}}{\tau^L(x'\eta)} \le \frac{1}{K_0}$. Setting $D = 4M_0RK_1/K_0^2$ yields $\|\tilde{f}^U - \tilde{f}^L\|_{B,P} \le \tilde{K}\epsilon$ for $\tilde{K} = \frac{a_0^{1/2}}{2M_0K_1}$, and thus

$$H_B(\epsilon, \tilde{\mathcal{F}}, \|\cdot\|_{B,P}) \le \frac{\tilde{B}}{\epsilon}, \quad \text{for } \tilde{B} = \frac{a_0^{1/2}C_2}{2M_0}. \tag{B.3}$$

Now we compute the Bernstein norm of $\tilde{f}$:

$$\begin{aligned} \|\tilde{f}\|_{B,P}^2 &= 2\sum_{k=2}^{\infty} \frac{1}{k!D^k}\int_{\mathscr{W}\times\mathbb{R}} \mathbb{I}\{x'\eta > q_n\}\left|\left(\frac{1}{\tau(x'\eta)} - \frac{1}{\tau_\eta(x'\eta)}\right) w_h u\right|^k dP(w,u) \\ &\le 2\sum_{k=2}^{\infty} \frac{1}{k!D^k}\left\{\frac{R^k k! M_0^{k-2} a_0 (2K_1)^{k-2}}{K_0^{2k}}\|\tau - \tau_0\|_P^2\right\} \\ &\le 2a_0\left(\frac{R}{DK_0^2}\right)^2 \sum_{k=0}^{\infty}\left(\frac{2RM_0K_1}{DK_0^2}\right)^k v^2 \le \frac{a_0}{4M_0^2}\frac{1}{K_1^2}v^2, \end{aligned}$$

where the first inequality follows from $\frac{\mathbb{I}\{x'\eta > q_n\}}{\tau(x'\eta)} \le \frac{1}{K_0}$. This implies

$$\|\tilde{f}\|_{B,P} \le B\frac{v}{K_1}, \quad \text{for } B = \frac{a_0^{1/2}}{2M_0}. \tag{B.4}$$

Combining (B.3) and (B.4), the remaining steps are same as those in the proof of Lemma 3 (iii).

B.3. **Proof of Lemma 4.** Recall for fixed η, we first obtain $\hat{\tau}_\eta = \arg\min_{\tau\in\mathcal{M}} \frac{1}{n}\sum_{i=1}^n \{\hat{U}_i^2 - \tau(X_i'\eta)\}^2$ and then obtain $\hat{\eta}$ by $\hat{\eta} = \underset{\eta}{\operatorname{argmin}} ||\frac{1}{n}\sum_{i=1}^n X_i'\{\hat{U}_i^2 - \hat{\tau}_\eta(X_i'\eta)\}||^2$. We denote $E[X|X'\eta =$

$x'\eta]$ by $E[X|x'\eta]$. The proof is similar to the ones in BGH and Balabdaoui and Groeneboom (2020) except that we need to handle the influence of the estimated dependent variables $\hat{U}_i^2$.

The proof of consistency of $\hat{\eta}$ is similar to pp.16-17 of BGH-supp. By a similar argument in Balabdaoui and Groeneboom (2020, Lemma 3.2), under Assumptions M1-M3, we have

$$\frac{1}{n}\sum_{i=1}^{n} X_i'\{\hat{U}_i^2 - \hat{\tau}_\eta(X_i'\eta)\} = \frac{1}{n}\sum_{i=1}^{n}(X_i - E[X|X_i'\eta])\{\hat{U}_i^2 - \tau_\eta(X_i'\eta)\} + o_p(n^{-1/2}),$$

for each η, where we also use (B.2). Thus, it holds

$$\begin{aligned}\left\|\frac{1}{n}\sum_{i=1}^{n} X_i\{\hat{U}_i^2 - \hat{\tau}_{\hat{\eta}}(X_i'\hat{\eta})\}\right\| &= \min_\eta \left\|\frac{1}{n}\sum_{i=1}^{n} X_i\{\hat{U}_i^2 - \hat{\tau}_\eta(X_i'\eta)\}\right\| \\ &\leq \min_\eta \left\|\frac{1}{n}\sum_{i=1}^{n}(X_i - E[X|X_i'\eta])\{\hat{U}_i^2 - \tau_\eta(X_i'\eta)\} + o_p(n^{-1/2})\right\|.\end{aligned}$$

The leading term in the last expression does not depend on $\hat{\tau}_\eta$, and is a smooth function of η. Thus, under standard conditions for the method of moments, we have $\min_\eta \left\|\frac{1}{n}\sum_{i=1}^{n}(X_i - E[X_i|X_i'\eta])\{\hat{U}_i^2 - \tau_\eta(X_i'\eta)\}\right\| = 0$, and

$$\begin{aligned} o_p(n^{-1/2}) &= \frac{1}{n}\sum_{i=1}^{n} X_i\{\hat{U}_i^2 - \hat{\tau}_{\hat{\eta}}(X_i'\hat{\eta})\} \\ &= \frac{1}{n}\sum_{i=1}^{n}(X_i - E[X|X_i'\hat{\eta}])\{\hat{U}_i^2 - \hat{\tau}_{\hat{\eta}}(X_i'\hat{\eta})\} + o_p(n^{-1/2} + (\hat{\eta} - \eta)) \\ &= \int (x - E[X|x'\hat{\eta}])\{\hat{u}^2 - \tau_0(x'\eta_0)\}d(\mathbb{P}_n - P)(x, \hat{u}) \\ &\quad + \int (x - E[X|x'\hat{\eta}])\{\hat{u}^2 - \tau_{\hat{\eta}}(x'\hat{\eta})\}dP(x, \hat{u}) + o_p(n^{-1/2} + (\hat{\eta} - \eta)) \\ &=: I + II + o_p(n^{-1/2} + (\hat{\eta} - \eta)), \end{aligned} \tag{B.5}$$

where the second equality follows from similar arguments to pp.18-20 of BGH-supp and (B.2), and the third equality follows from a similar argument in pp.21-23 of BGH-supp.

Let $\hat{U}(w, u) = u - w'(\hat{\theta}_{\text{OLS}} - \theta)$ and

$$\hat{e}(w, u) := \hat{U}(w, u) - u^2 = -2w'(\hat{\theta}_{\text{OLS}} - \theta)u + \{w'(\hat{\theta}_{\text{OLS}} - \theta)\}^2. \tag{B.6}$$

For I, we have

$$\begin{aligned} I &= \int (x - E[X|x'\hat{\eta}])\{u^2 + \hat{e}(w, u) - \tau_0(x'\eta_0)\}d(\mathbb{P}_n - P)(w, u) \\ &= \int (x - E[X|x'\eta_0])\{u^2 - \tau_0(x'\eta_0)\}d(\mathbb{P}_n - P)(x, u) \\ &\quad + \int (x - E[X|x'\hat{\eta}])\hat{e}(w, u)d(\mathbb{P}_n - P)(w, u) + o_p(n^{-1/2}) \\ &= \int (x - E[X|x'\eta_0])\{u^2 - \tau_0(x'\eta_0)\}d(\mathbb{P}_n - P)(x, u) + o_p(n^{-1/2}), \end{aligned} \tag{B.7}$$

where the second equality follows from p.21 of BGH-supp, and the third equality follows from the facts that (a) $\hat{\theta}_{\mathrm{OLS}} - \theta = O_p(n^{-1/2})$, (b) $\hat{e}(w,u)$ is a parametric function of w and u in a changing class indexed by $\hat{\theta}_{\mathrm{OLS}}$ (see (B.6)), so its ϵ-entropy is of order $\log(1/\epsilon) \leq 1/\epsilon$ (see, e.g., Example 19.7 of van der Vaart and Wellner, 2000), and (c) similar arguments in pp.22-23 of BGH-supp. By Lemma 17 of BGH-supp we have

$$\tau_\eta(x'\eta) = \tau_0(x'\eta_0) + (\eta - \eta_0)(x - E[X|X'\eta_0 = x'\eta_0])\tau_0'(x'\eta_0) + o_p(\eta - \eta_0). \tag{B.8}$$

For II, observe that

$$\begin{aligned}
II &= \int (x - E[X|x'\hat{\eta}])\{u^2 - \tau_{\hat{\eta}}(x'\hat{\eta})\} dP(x,u) + \int (x - E[X|x'\hat{\eta}])\hat{e}(w,u) dP(w,u) \\
&= \left\{ \int (x - E[X|x'\eta_0])(x - E[X|X'\eta_0 = x'\eta_0])\tau_0'(x'\eta_0) dP(x) \right\} (\hat{\eta} - \eta_0) \\
&\quad + \int (x - E[X|x'\hat{\eta}])\hat{e}(w,u) dP(w,u) + o_p(\hat{\eta} - \eta_0) \\
&= \left\{ \int (x - E[X|x'\eta_0])(x - E[X|x'\eta_0])\tau_0'(x'\eta_0) dP(x) \right\} (\hat{\eta} - \eta_0) + O_p(n^{-1/2}) + o_p(\hat{\eta} - \eta_0) \\
&= B(\hat{\eta} - \eta_0) + O_p(n^{-1/2}) + o_p(\hat{\eta} - \eta_0),
\end{aligned} \tag{B.9}$$

where the third equality follows from (B.8) and $(E[X|x'\hat{\eta}] - E[X|x'\eta_0])(\hat{\eta} - \eta_0) = o_p(\hat{\eta} - \eta_0)$, the fourth equality follows from $\hat{\theta}_{\mathrm{OLS}} - \theta = O_p(n^{-1/2})$ and the definition of B in Assumption M6.

Combining (B.5), (B.7), and (B.9), we have

$$\hat{\eta} - \eta_0 = B^- \int (x - E[X|x'\eta_0])\{u^2 - \tau_0(x'\eta_0)\} d(\mathbb{P}_n - P)(x,u) + O_p(n^{-1/2}) + o_p(n^{-1/2} + (\hat{\eta} - \eta)),$$

where B^- is the Moore-Penrose inverse of B (see, p.17 of BGH for more details). Therefore, we have $\hat{\eta} - \eta_0 = O_p(n^{-1/2})$. This result combined with (B.8) and Assumption M2 (iv) imply $\tau_{\hat{\eta}}(u) - \tau_0(u) = O_p(n^{-1/2})$.

References

[1] Acemoglu, D. and P. Restrepo (2017) Secular stagnation? The effect of aging on economic growth in the age of automation, *American Economic Review*, 107, 174-179.

[2] Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T. and E. Silverman (1955) An empirical distribution function for sampling with incomplete information, *Annals of Mathematical Statistics*, 26, 641-647.

[3] Babii, A. and R. Kumar (2021) Isotonic regression discontinuity designs, forthcoming in *Journal of Econometrics*.

[4] Balabdaoui, F., Durot, C. and H. Jankowski (2019) Least squares estimation in the monotone single index model, *Bernoulli*, 25, 3276-3310.

[5] Balabdaoui, F., Groeneboom, P. and K. Hendrickx (2019) Score estimation in the monotone single-index model, *Scandinavian Journal of Statistics*, 46, 517-544.

[6] Balabdaoui, F. and P. Groeneboom (2021) Profile least squares estimators in the monotone single index model, in *Advances in Contemporary Statistics and Econometrics*, pp. 3-22, Springer.

[7] Barlow, R. and H. Brunk (1972) The isotonic regression problem and its dual, *Journal of the American Statistical Association*, 67, 140-147.

[8] Bickel, P. J. (1978) Using residuals robustly I: Tests for heteroscedasticity, nonlinearity, *Annals of Statistics*, 6, 266-291.

[9] Box, G. E. and W. J. Hill (1974) Correcting inhomogeneity of variance with power transformation weighting, *Technometrics*, 16, 385-389.

[10] Carroll, R. J. (1982) Adapting for heteroscedasticity in linear models, *Annals of Statistics*, 10, 1224-1233.

[11] Chetverikov, D. (2019) Testing regression monotonicity in econometric models, *Econometric Theory*, 35, 729-776.

[12] Cragg, J. G. (1983) More efficient estimation in the presence of heteroscedasticity of unknown form, *Econometrica*, 51, 751-764.

[13] Dumbgen, L. and V. G. Spokoiny (2001) Multiscale testing of qualitative hypotheses, *Annals of Statistics*, 29, 124-152.

[14] Ghosal, S., Sen, A. and A. van der Vaart (2000) Testing monotonicity of regression, *Annals of Statistics*, 28, 1054-1082.

[15] Grenander, U. (1956) On the theory of mortality measurement. II., *Skand. Aktuarietidskr*, 39, 125-153.

[16] Groeneboom, P. and G. Jongbloed (2014) *Nonparametric Estimation under Shape Constraints*, Cambridge University Press.

[17] Hall, P. and N. E. Heckman (2000). Testing for monotonicity of a regression mean by calibrating for linear functions, *Annals of Statistics*, 28, 20-39.

[18] Hansen, A. H. (1939) Economic progress and declining population growth, *American Economic Review*, 29, 1-15.

[19] Hsu, Y. C., Liu, C. A. and X. Shi (2019) Testing generalized regression monotonicity, *Econometric Theory*, 35, 1146-1200.

[20] Jobson, J. D. and W. A. Fuller (1980) Least squares estimation when the covariance matrix and parameter vector are functionally related, *Journal of the American Statistical Association*, 75, 176-181.

[21] Kim, J. and D. Pollard (1990) Cube root asymptotics, *Annals of Statistics*, 18, 191-219.

[22] Matzkin, R. L. (1994) Restrictions of economic theory in nonparametric methods, in Engle R. F. and D. L. McFadden (eds.), *Handbook of Econometrics*, vol. IV, pp. 2523-2558, Elsevier.

[23] Newey, W. K. (1993) Efficient estimation of models with conditional moment restrictions, in Maddala, G. S., Rao, C. R. and H. D. Vinod (eds.), *Handbook of Statistics*, vol. 11, pp. 419-454, Elsevier.

[24] Rao, B. P. (1969) Estimation of a unimodal density, *Sankhyā*, A 31, 23-36.

[25] Rao, B. P. (1970) Estimation for distributions with monotone failure rate, *Annals of Mathematical Statistics*, 41, 507-519.

[26] Robinson, P. M. (1987) Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form, *Econometrica*, 55, 875-891.

[27] van der Vaart, A. W. (2000) *Asymptotic Statistics*, Cambridge university press.

[28] van der Vaart, A. W. and J. A. Wellner (1996) *Weak Convergence and Empirical Processes*, Springer.

[29] Wooldridge, J. M. (2010) *Econometric Analysis of Cross Section and Panel Data*, 2nd ed., MIT Press.

School of Social Sciences, Waseda University, 1-6-1 Nishiwaseda, Shinjuku-ku, Tokyo 169-8050, Japan.

Email address: `yarai@waseda.jp`

Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK.

Email address: `t.otsu@lse.ac.uk`

Department of Economics, University of Mannheim, L7 3-5, 68161, Mannheim, Germany.

Email address: `mengshan.xu@uni-mannheim.de`