Semiparametric Efficiency for Partially Linear Single-Index Regression Models

Tao Chen[a], Thomas Parker[a,*]

[a] *Department of Economics, University of Waterloo, 200 University Avenue, Ontario, Canada N2L 3G1*

Abstract

We calculate semiparametric efficiency bounds for a partially linear single-index model using a simple method developed by [1]. We show that this model can be used to evaluate the efficiency of several existing estimators.

Keywords: Semiparametric efficiency, Partially linear single-index model, Single-index model, Conditional mean restriction, Conditional quantile restriction

1. Introduction

A straightforward and convenient method for the calculation of semiparametric efficiency bounds was proposed in [1] and illustrated using several example models. In this paper, we extend their method to investigate semiparametric efficiency for the finite-dimensional parameters of the model

$$y = g_0(X'\theta_0) + Z'\beta_0 + \varepsilon, \tag{1.1}$$

where θ_0 and β_0 are unknown, $g_0(\cdot)$ is an unknown function and ε has a distribution conditional on covariates X and Z. This model encompasses several interesting special cases — the linear-, partially linear-, single-index- and partially linear single-index models — and the method used here makes it simple to account for various identification conditions. Our method is that illustrated by [1] (explained in greater detail in the recent survey [2]), and we extend their results to to semiparametric models of conditional quantiles, which were not considered by those authors.

*Corresponding author.

Email addresses: t66chen@uwaterloo.ca (Tao Chen), tmparker@uwaterloo.ca (Thomas Parker)

A number of estimators of model (1.1) and related models have been proposed. Semiparametric efficient locally-linear quasi-likelihood estimation of this model was proposed in [3]; they also showed that their estimator reached the bound (and also verified that the estimator in [4] attained this bound). Sieve estimators of θ_0 in the single-index model have been proposed by [5] and [6]. Computationally attractive estimators were proposed in [7] and in [8] for the single-index model under conditional quantile identification conditions. [9] and [10] proposed estimators of model (1.1) for a conditional mean identification condition. In addition, [11] have proposed an estimator for a closely-related group of models of conditional quantiles. It is of interest to know whether these estimators attain the relevant efficiency bounds.

In this article we use a method due to [1] to derive the semiparametric efficiency bound for this model in a straightforward manner, independent of assumptions regarding identification or type of estimator. Model (1.1) is not addressed in [1], and we extend their method to this model without making any more restrictions on the conditional distribution of ε given W. We then use our bound to compare to results for conditiona mean- and quantile location identification conditions. This method makes it easy to derive the general bound (i.e., for the "least-favorable parametric submodel") in a clear manner without going through the usual two-step style calculation, as represented for example by [12] and [13], and we view it as a complement to that model, which appears to be more well-suited to calculations for special cases (like efficiency bounds for the class of M-estimators, for example).

2. General assumptions

We assume the random variables $y \in \mathbb{R}$, $W = [\, X' \; Z' \,]' \in \mathbb{R}^p$, and $\varepsilon \in \mathbb{R}$ have densities $q_0^2(y|W)$, $b_0^2(W)$ and $\gamma_0^2(\varepsilon|W)$ respectively. The likelihood associated with an observation (y, W) is

$$\mathscr{L}(\theta; y, W) = q_0^2(y|W) b_0^2(W)$$

which could alternatively be expressed using the conditional density of the additive error term ε, because the model (1.1) implies q_0 and γ_0 satisfy the equation $q_0(y|W) = \gamma_0(y - g_0(X'\theta_0) + Z'\beta_0|W)$. Assume $g_0 \in \mathrm{L}^2(\mathbb{R}, \lambda)$ and has a derivative $g_0' \in \mathrm{L}^2(\mathbb{R}, \lambda)$, where the notation $\mathrm{L}^2(\mathcal{A}, \mu)$ denotes the space of square-integrable functions on some domain $\mathcal{A}$ with respect to some measure μ, and λ is Lebesgue measure. Because g_0 is unknown, we make the definitions $X_0 = X'\theta_0$ (where X' denotes "X transpose" below) and $W_0 = [\, x_0 \; z' \,]' \in \mathbb{R}^{p_0}$, and impose the "index restriction" $q_0(y|W) = q_0(y|W_0)$ (equivalently, $\gamma_0(\varepsilon|W) = \gamma_0(\varepsilon|W_0)$). This differs slightly from previous partial-index models in the literature (e.g., [12]) that assumed the variance function was a fully nonparametric function of W. As is pointed out by [14], when considering estimation, such models may suffer from the curse of dimensionality. Our results are relevant for models with variance functions that generally depend only on W_0. This restriction has some precedence in the literature; for example, [4] restrict their attention to similar cases. Finally, we note that we implicitly assume the model is identified. In practice, this would mean for example, assuming the first element of θ_0 is normalized to 1 and the first element of X is continuously distributed, as well as using trimming in an estimator to ensure the positivity of the density of $X'\theta_0$; however, we abstract away from these details to focus on the technique used to derive the efficiency bound.

We make minimal assumptions regarding b_0^2, the marginal density of W: we assume that b_0 is a member of the space $\mathcal{B}$, where

$$\mathcal{B} = \left\{ b \in \mathrm{L}^2(\mathbb{R}^p, \lambda) : b^2(w) > 0, \int_{\mathbb{R}^p} b^2(w)\mathrm{d}w = 1 \right\},$$

and the additional identification assumption that $\mathrm{E}\left[(W - \mathrm{E}\left[W|X_0\right])(W - \mathrm{E}\left[W|X_0\right])'\right]$ exists and is nonsingular (one could assume only a generalized inverse, as in [10] for a model similar to (1.1)). Assume that $\gamma_0 \in \Gamma$, where

$$\Gamma = \left\{ \gamma : \mathbb{R} \times \mathbb{R}^p \to \mathbb{R} : \gamma(u|W) = \gamma(u|X'\theta_0, Z), \gamma^2(u|W) > 0, \int_{\mathbb{R}} \gamma^2(u|W)\mathrm{d}u = 1, \right.$$
$$\left. \gamma(u|W) \text{ is bounded and continuous, and } \int_{\mathbb{R}} \left(\gamma'(u|W)\right)^2 \mathrm{d}u < \infty, \text{ all } w.p.1 \right\}$$

where γ' refers to a partial derivative with respect to u — that is, $\gamma'(u|w) = \partial\gamma(u|w)/\partial u$. Further conditions that γ_0 must satisfy will be specified below, depending on the identification condition imposed on ε. Because of the aforementioned equation of q_0 with γ_0, the space of functions $\mathcal{Q} \ni q$ is essentially the same as Γ described here, and we simply rely on Γ as the relevant space.

To derive a semiparametric efficiency bound for this model, we follow the strategy of [1] — in order to consider the likelihood functions of one-dimensional submodels local to the true model, we organize local deviations from the model using real-valued $t \in [0, t_0]$ for some $t_0 > 0$. Let $\xi := (\theta', \beta')'$, and consider a curve $t \mapsto (\xi_t, g_t, \gamma_t, b_t)$ from $[0, t_0]$ into $\mathbb{R}^p \times \mathrm{L}^2(\mathbb{R}, \lambda) \times \Gamma \times \mathcal{B}$ that passes through $(\xi_0, g_0, \gamma_0, b_0)$ at $t = 0$. The score function for the model with respect to t (treating t as if it were the parameter to be estimated) is

$$S_0 = \frac{2\dot{q}(y|W)}{q_0(y|W)} + \frac{2\dot{b}(W)}{b_0(W)} \tag{2.1}$$

where $\dot{q} = \frac{\mathrm{d}}{\mathrm{d}t} q_t(y|W)|_{t=0}$ is tangent to q_t at $t = 0$ and all other "dotted" quantities are defined analogously. The Fisher information for the parameter ξ is

$$i_F = \mathrm{E}\left[S_0^2\right] = \mathrm{E}\left[\frac{4\dot{q}^2(y|W)}{q_0^2(y|W)} + \frac{4\dot{b}^2(W)}{b_0^2(W)}\right]$$

(the zero-mean property of score functions implies that the cross term in the quadratic is zero). Therefore

$$\begin{aligned} i_F &= 4\mathrm{E}\left[\frac{\dot{q}^2(y|W)}{q_0^2(y|W)}\right] + 4\int_{\mathbb{R}^p} \dot{b}^2(w)\mathrm{d}w \\ &= 4\mathrm{E}_W\left[\int_{\mathbb{R}} \dot{q}^2(y|W)\mathrm{d}y\right] + 4\int_{\mathbb{R}^p} \dot{b}^2(w)\mathrm{d}w. \end{aligned}$$

As noted in [1], the above conditions are not enough to generally characterize informative efficiency bounds. We require that the functions considered in Fisher information calculations come from the subset of *feasible* score functions, which are those functions in $\mathbb{R}^p \times \mathrm{L}^2(\mathbb{R}, \lambda) \times \Gamma \times \mathcal{B}$ that are locally (i.e., for $t \in [0, t_0]$) linear in t (see [1, p. 27] for a longer discussion of this condition). Feasible score functions are $\dot{\tau} \in \dot{\mathcal{T}}$, where $\dot{\tau} = (\dot{\xi}, \dot{g}, \dot{\gamma}, \dot{b})$ denote parts that go towards defining a score function S_0. Exactly which functions are feasible (i.e., belong in $\dot{\mathcal{T}}$) depends on the identification conditions presented below. However, we can specify many characteristics of these functions given model (1.1). Regardless of the identification condition, we require that for any model, $\dot{g} \in \mathrm{L}^2(\mathbb{R}, b_{00}^2)$, where b_{00}^2 is defined as the density of X_0, and $\dot{\xi} \in \mathbb{R}^p$ (feasible $\dot{b}$ are described below). We take as our parameter of interest $\rho(\tau) := c'\xi$, where $c \in \mathbb{R}^p$ is arbitrary and included so that the parameter is real-valued. For the purposes of our efficiency calculation, we assume that this is differentiable relative to the tangent set (which will be defined below on a case-by-case basis); that is, we assume $\lim_{t \to 0} \frac{1}{t}(c'\xi_t - c'\xi) = c'\dot{\xi} = \dot{\rho}(\dot{\tau})$ for all $\dot{\tau}$ in the tangent set. We define the Fisher inner product for any feasible $\dot{\tau}_1$ and $\dot{\tau}_2$ by

$$\langle \dot{\tau}_1, \dot{\tau}_2 \rangle_F = 4\mathrm{E}_W\left[\int_{\mathbb{R}} \dot{q}_1(y|W)\dot{q}_2(y|W)\mathrm{d}y\right] + 4\int_{\mathbb{R}^p} \dot{b}_1(w)\dot{b}_2(w)\mathrm{d}w$$

The Riesz Representation Theorem implies that there exists a τ^* such that

$$c'\dot{\xi} = \dot{\rho}(\dot{\tau}) = \langle \tau^*, \dot{\tau} \rangle_F.$$

See [15, p. 363] or [1, § 2] for more on this topic. We calculate the efficiency bound by first solving for τ^* using the above equality. Next, the efficiency bound is calculated by evaluating the Fisher information i_F at (q^*, b^*); that is, the information bound is equal to $\langle \tau^*, \tau^* \rangle_F$.

Here we generally characterize $\dot{\mathcal{T}}$, the space of feasible score functions, although the cases considered below will have additional specific features that depend on identification condition. We use $\overline{\text{lin}}\,\mathcal{Z}$ to denote the closed linear span of elements of a set $\mathcal{Z}$ (defining tangent spaces using closed spaces to ensure existence and uniqueness of τ^* (cf. [15, p. 363])). Because the model considered here is so similar to the partially linear model of [1, §8], we may use their characterization of the tangent space of vectors near γ_0; that is, Lemma B.2 of [1] implies that

$$\overline{\text{lin}}\ T(\Gamma, \gamma_0) = \left\{ \dot{\gamma} : \mathrm{L}^2(\mathbb{R} \times \mathbb{R}^{p_0}, \lambda \times \tilde{b}_{00}^2) : \int_{\mathbb{R}} \dot{\gamma}(u|W_0)\gamma_0(u|W)\mathrm{d}u = 0 \ w.p.1 \right\}$$

where $\tilde{b}_{00}^2$ is the density of W_0. We note in passing that the proof of this assertion is constructive, using parametric submodels chosen to be in the tangent space; more recently, [16] have proposed a different construction for dynamic models of conditional quantiles. [1, Lemma B.1] implies that the closed linear span of vectors $\dot{b}$ tangent to $\mathcal{B}$ at b_0 is

$$\overline{\text{lin}}\ T(\mathcal{B}, b_0) = \left\{ \dot{b} \in \mathrm{L}^2(\mathbb{R}^p, \lambda) : \int_{\mathbb{R}^p} \dot{b}(w)b_0(w)\mathrm{d}w = 0 \right\}$$

Such a tangent set characterization is referred to in [15, p. 363-4] as the "maximal tangent set" for $(\mathcal{B}, b_0)$. For feasibility, we require $\dot{g} \in \mathrm{L}^2(\mathbb{R}, b_{00}^2)$. Collecting these conditions results in a complete characterization of the tangent space $\dot{\mathcal{T}}$:

$$\dot{\mathcal{T}} = \mathbb{R}^p \times \mathrm{L}^2(\mathbb{R}, b_{00}^2) \times \overline{\text{lin}}\, T(\Gamma, \gamma_0) \times \overline{\text{lin}}\, T(\mathcal{B}, b_0). \tag{2.2}$$

$\dot{\tau} = (\dot{\xi}, \dot{g}, \dot{\gamma}, \dot{b}) \in \dot{\mathcal{T}}$ is a necessary condition for feasible score functions (as functions of $\dot{\tau}$) in efficiency calculations below. Feasible functions $\dot{q}$ are constructed from these parts via the identity (arrived at by differentiating $q_t(y|W) = \gamma_t(y - g_0(X'\theta_t) - Z'\beta_t|W)$ with respect to t)

$$\dot{q}(y|W) = \dot{\gamma}(\varepsilon|W_0) - \gamma_0'(\varepsilon|W)\left(\dot{g}(X_0) + g_0'(X_0)X'\dot{\theta} + Z'\dot{\beta}\right), \tag{2.3}$$

and these functions are used in Fisher information calculations that provide the semiparametric efficiency bound.

3. The model with independent errors

Suppose that the model of interest is model (1.1), where ε is independent of W, and the semiparametric efficiency bound for estimators of ξ_0 is desired. One example discussed in [1, §8] is the semiparametric efficiency bound for estimators of the finite-dimensional parameter in a partially linear model with ε distributed independently of W. We use the solution method of [1] for this model, which represents a minor extension of their results. In the next Section we consider a solution method when the shape of γ_0 and q_0 may depend on covariates.

Independence of ε implies that $q_0(y|W) = \gamma_0(\varepsilon)$. Here $\gamma \in \Gamma'$, which is simpler than Γ defined above:

$$\Gamma' = \Big\{ \gamma : \mathbb{R} \to \mathbb{R} : \gamma^2(\varepsilon) > 0, \int_{\mathbb{R}} \gamma^2(u) \mathrm{d}u = 1, \\ \gamma(\varepsilon) \text{ is bounded and continuous, and } \int_{\mathbb{R}} \left(\gamma'(u)\right)^2 \mathrm{d}u < \infty \Big\}$$

and

$$\overline{\mathrm{lin}}\ T(\Gamma', \gamma_0) = \left\{ \dot\gamma : \mathrm{L}^2(\mathbb{R}, \lambda) : \int_{\mathbb{R}} \dot\gamma(u)\gamma_0(u)\mathrm{d}u = 0 \right\}.$$

Feasible $\dot\gamma$ are elements of this space. Therefore vectors $\dot\tau$ can be written $\dot\tau = (\dot\xi, \dot g, \dot\gamma, \dot b) \in \dot{\mathcal{T}}'$, where $\dot{\mathcal{T}}' = \mathbb{R}^p \times \mathrm{L}^2(\mathbb{R}, b_{00}^2) \times \overline{\mathrm{lin}}\, T(\Gamma', \gamma_0) \times \overline{\mathrm{lin}}\, T(\mathcal{B}, b_0)$. Theorem 3.1 collects the vector τ^* and the semiparametric efficiency bound for this model. Below we use $\mathrm{E}\left[X\right]^{-1}$ to denote $(\mathrm{E}\left[X\right])^{-1}$ in simple expressions, in order to reduce the great number of parentheses that would otherwise result.

Theorem 3.1. *Assume the model is (1.1) and that ε is independent of W. Then the solution vector τ^* such that $\langle \tau^*, \dot\tau \rangle_F = c'\dot\xi$ for arbitrary $c \in \mathbb{R}^p$, and all $\dot\tau \in \dot{\mathcal{T}}'$ is*

$$\tau^* = (\xi^*, g^*, \gamma^*, b^*) = \left(\mathrm{E}\left[\tilde S \tilde S'\right]^{-1} c, -\mathrm{E}\left[\tilde W | X_0\right] \mathrm{E}\left[\tilde S \tilde S'\right]^{-1} c, 0, 0 \right).$$

where $\psi_0(\varepsilon|W_0) = -2\gamma_0'(\varepsilon)/\gamma_0(\varepsilon)$, $\tilde W = [\, g_0'(X_0)X' \;\; Z' \,]'$ and

$$\tilde S = \psi_0(\varepsilon) \left(\tilde W - \mathrm{E}\left[\tilde W | X_0\right] \right).$$

Furthermore, the semiparametric efficiency bound for estimators of ξ_0 *is*

$$\mathrm{E}\left[\tilde{S}\tilde{S}'\right]^{-1}.$$

We leave the proof of Theorem 3.1 here because it is illustrative of the technique used in the proof of the main Theorem presented in the next Section.

Proof. As discussed above, we find the semiparametric efficiency bound for estimators of $c'\xi_0$, because it is relatively simple to calculate for a scalar condition, and this generalizes immediately to the important bound, namely that for estimators of ξ_0. We find τ^* by solving the equation that it must satisfy to be feasible: $\langle \tau^*, \dot{\tau}\rangle_F = c'\dot{\xi}$ for all $\dot{\tau} \in \dot{\mathcal{T}}'$. Then the components of τ^* are assembled to find q^* using (2.3). The equation $\langle \tau^*, \dot{\tau}\rangle_F = c'\dot{\xi}$ is equivalent to

$$4\mathrm{E}\left[\frac{q^*(y|W)\dot{q}(y|W)}{q_0^2(y|W)}\right] + 4\int_{\mathbb{R}^p} b^*(w)\dot{b}(w)\mathrm{d}w = c'\dot{\xi},$$

or using (2.3),

$$4\mathbf{E}\left[\frac{\gamma^*(\varepsilon) - \gamma_0'(\varepsilon)\Big(g^*(X_0) + g_0'(X_0)X'\theta^* + Z'\beta^*\Big)}{\gamma_0(\varepsilon)}\times \right.$$
$$\left.\frac{\dot{\gamma}(\varepsilon) - \gamma_0'(\varepsilon)\left(\dot{g}(X_0) + g_0'(X_0)X'\dot{\theta} + Z'\dot{\beta}\right)}{\gamma_0(\varepsilon)}\right] + 4\int_{\mathbb{R}^p} b^*(w)\dot{b}(w)\mathrm{d}w = c'\dot{\xi}.$$

This implies a system of several equations:

$$\int_{\mathbb{R}^p} b^*(w)\dot{b}(w)\mathrm{d}w = 0$$
$$\mathrm{E}\left[\frac{\gamma^*(\varepsilon) - \gamma_0'(\varepsilon)\left(g^*(X_0) + g_0'(X_0)X'\theta^* + Z'\beta^*\right)}{\gamma_0^2(\varepsilon)}\dot{\gamma}(\varepsilon)\right] = 0$$
$$\mathrm{E}\left[\frac{\gamma^*(\varepsilon) - \gamma_0'(\varepsilon)\left(g^*(X_0) + g_0'(X_0)X'\theta^* + Z'\beta^*\right)}{\gamma_0^2(\varepsilon)}\gamma_0'(\varepsilon)\dot{g}(X_0)\right] = 0$$
$$-4\mathrm{E}\left[\frac{\gamma^*(\varepsilon) - \gamma_0'(\varepsilon)\left(g^*(X_0) + g_0'(X_0)X'\theta^* + Z'\beta^*\right)}{\gamma_0^2(\varepsilon)}\gamma_0'(\varepsilon)g_0'(X_0)X'\right]\dot{\theta} = c_1'\dot{\theta}$$
$$-4\mathrm{E}\left[\frac{\gamma^*(\varepsilon) - \gamma_0'(\varepsilon)\left(g^*(X_0) + g_0'(X_0)X'\theta^* + Z'\beta^*\right)}{\gamma_0^2(\varepsilon)}\gamma_0'(\varepsilon)Z'\right]\dot{\beta} = c_2'\dot{\beta}$$

where the arbitrary vector $c \in \mathbb{R}^p$ is broken into conformable subvectors c_1 and c_2 in the final two equations. The first equation implies that $b^* \equiv 0$; this is also a feasible function. From the next equation one sees that

$$\gamma^*(\varepsilon) = \gamma_0'(\varepsilon)\mathrm{E}\left[g^*(X_0) + g_0'(X_0)X'\theta^* + Z'\beta^*\right]$$

satisfies the equation for all $\dot{\gamma}$. The next equation implies similarly that

$$g^*(X_0) = -\mathrm{E}\left[g_0'(X_0)X'\theta^* + Z'\beta^*|X_0\right]$$

which is square integrable under our assumptions (and therefore feasible). Inserting this solution for g^* in the expression for γ^* one finds $\gamma^* \equiv 0$, which is also feasible. Now we use these solutions and consolidate the final two equations into one vector equation, recalling the definition $W_0 = [\,X_0 \; Z'\,]'$ and introducing notation $\tilde{W} = [\,g_0'(X_0)X' \; Z'\,]'$ as in the statement of the Theorem:

$$4\mathrm{E}\left[\frac{\gamma_0'(\varepsilon)\left(\tilde{W} - \mathrm{E}\left[\tilde{W}|X_0\right]\right)'\xi^*}{\gamma_0^2(\varepsilon)}\gamma_0'(\varepsilon)\tilde{W}'\right]\dot{\xi} = c'\dot{\xi}. \tag{3.1}$$

Here we make some more definitions: let $\psi_0(\varepsilon) = -2\gamma_0'(\varepsilon)/\gamma_0(\varepsilon)$ (which is equal to $\frac{\mathrm{d}}{\mathrm{d}u}\log\gamma_0^2(u)$), and let

$$\tilde{S} = \psi_0(\varepsilon)\left(\tilde{W} - \mathrm{E}\left[\tilde{W}|X_0\right]\right).$$

Then we rewrite the previous display as

$$\dot{\xi}'\mathrm{E}\left[\psi_0^2(\varepsilon)\left(\tilde{W} - \mathrm{E}\left[\tilde{W}|X_0\right]\right)\left(\tilde{W} - \mathrm{E}\left[\tilde{W}|X_0\right]\right)'\right]\xi^* = \dot{\xi}'\mathrm{E}\left[\tilde{S}\tilde{S}'\right]\xi^* = \dot{\xi}'c.$$

This implies

$$\xi^* = \mathrm{E}\left[\tilde{S}\tilde{S}'\right]^{-1}c,$$

and after solving for τ^* one arrives at

$$\tau^* = (\xi^*, g^*, \gamma^*, b^*) = \left(\mathrm{E}\left[\tilde{S}\tilde{S}'\right]^{-1}c, -\mathrm{E}\left[\tilde{W}|X_0\right]\mathrm{E}\left[\tilde{S}\tilde{S}'\right]^{-1}c, 0, 0\right). \tag{3.2}$$

Now consider q^*. Using the components of the above solution,

$$\begin{aligned}
q^*(y|W) &= -\gamma_0'(\varepsilon)\left(\tilde{W} - \mathrm{E}\left[\tilde{W}|X_0\right]\right)'\mathrm{E}\left[\tilde{S}\tilde{S}'\right]^{-1}c \\
&= \frac{1}{2}\gamma_0(\varepsilon)\psi_0(\varepsilon)\left(\tilde{W} - \mathrm{E}\left[\tilde{W}|X_0\right]\right)'\mathrm{E}\left[\tilde{S}\tilde{S}'\right]^{-1}c \\
&= \frac{1}{2}\gamma_0(\varepsilon)\tilde{S}'\mathrm{E}\left[\tilde{S}\tilde{S}'\right]^{-1}c
\end{aligned} \tag{3.3}$$

This relies on the fact that $q_0(y|W) = \gamma_0(\varepsilon)$. Expression (3.3) will be used to extend the model to more general cases where the density of ε may depend on covariates.

Calculating the Fisher information of τ^* results in the lower bound for estimators of $c'\xi_0$: in this case,

$$\begin{aligned}
i_F &= 4\mathrm{E}\left[\frac{(q^*(y|W))^2}{q_0^2(y|W)}\right] + 4\mathrm{E}\left[(b^*(W))^2\right] \\
&= c'\mathrm{E}\left[\mathrm{E}\left[\tilde{S}\tilde{S}'\right]^{-1}\tilde{S}\tilde{S}'\mathrm{E}\left[\tilde{S}\tilde{S}'\right]^{-1}\right]c \\
&= c'\mathrm{E}\left[\tilde{S}\tilde{S}'\right]^{-1}c,
\end{aligned}$$

This implies that the lower bound for estimators of ξ_0 is $\mathrm{E}\left[\tilde{S}\tilde{S}'\right]^{-1}$ because c is arbitrary. □

Note that the $\tilde{S}$ given in the Theorem above is not exactly analogous to the efficient score given in other work (where the efficient score is given by projecting the score function on a tangent space). Because we work with square-root densities, the calculations are slightly different — for example, $\tilde{S}$ above is not necessarily mean zero, and calculating the efficiency bound results from evaluating the variance of S_0 given in (2.1) instead of $\tilde{S}$. However, the efficiency bound is most easily expressed in this context using $\tilde{S}$; of course, the resulting efficiency bound is the same, and only the calculations leading to this bound are different.

4. Likelihood when ε is heteroskedastic

Now suppose that the model is (1.1), but the density of ε is allowed to change its shape with W. How might one extend Theorem 3.1 to account for γ_0 that depends on ε and W? In this Section we keep the notation of the previous section, with one important exception. Here we use ψ_0 to refer to functions of ε conditional on W; that is, let $\psi_0(\varepsilon|W) = -2\gamma_0'(\varepsilon|W)/\gamma_0(\varepsilon|W)$. Thus, in this section we assume $\dot{\mathcal{T}}$ is the one specified in (2.2). Theorem 4.1 presents our result, which is that the efficiency bound takes the same basic form as in the independent case, but with a more complicated score function.

Theorem 4.1. *Assume the model is (1.1) and that the shape of the density of ε may depend on W. Then the solution vector τ^* such that $\langle \tau^*, \dot{\tau} \rangle_F = c'\dot{\xi}$ for arbitrary $c \in \mathbb{R}^p$, and all $\dot{\tau} \in \dot{\mathcal{T}}$ is*

$$\tau^* = (\xi^*, g^*, \gamma^*, b^*) = \left(\mathrm{E}\left[\tilde{S}\tilde{S}'\right]^{-1} c, -\mathrm{E}\left[\tilde{W}|X_0\right] \mathrm{E}\left[\tilde{S}\tilde{S}'\right]^{-1} c, 0, 0 \right).$$

where $\psi_0(\varepsilon|W_0) = -2\gamma_0'(\varepsilon|W)/\gamma_0(\varepsilon|W)$, $\tilde{W} = [\, g_0'(X_0)X' \;\; Z' \,]'$ and

$$\tilde{S} = \psi_0(\varepsilon|W) \left(\tilde{W} - \frac{\mathrm{E}\left[\psi_0^2(\varepsilon|W)\tilde{W}|X_0\right]}{\mathrm{E}\left[\psi_0^2(\varepsilon|W)|X_0\right]} \right)$$

Furthermore, the semiparametric efficiency bound for estimators of ξ_0 is

$$\mathrm{E}\left[\tilde{S}\tilde{S}'\right]^{-1}.$$

Proof. As before, the function τ^* must satisfy $\langle \tau^*, \dot{\tau}\rangle_F = c'\dot{\xi}$ for all $\dot{\tau}$; in other words,

$$4\mathbf{E}\left[\frac{\gamma^*(\varepsilon|W_0) - \gamma_0'(\varepsilon|W)\Big(g^*(X_0) + g_0'(X_0)X'\theta^* + Z'\beta^*\Big)}{\gamma_0(\varepsilon|W)}\times\right.$$
$$\left.\frac{\dot{\gamma}(\varepsilon|W_0) - \gamma_0'(\varepsilon|W)\left(\dot{g}(X_0) + g_0'(X_0)X'\dot{\theta} + Z'\dot{\beta}\right)}{\gamma_0(\varepsilon|W)}\right] + 4\int_{\mathbb{R}^p} b^*(w)\dot{b}(w)\mathrm{d}w = c'\dot{\xi}$$

for all $\dot{\tau} \in \dot{\mathcal{T}}$. The only difference between this and the analogous expression in the previous Section is that now γ is a function of ε and W.

$$\int_{\mathbb{R}^p} b^*(w)\dot{b}(w)\mathrm{d}w = 0$$
$$\mathrm{E}\left[\frac{\gamma^*(\varepsilon|W_0) - \gamma_0'(\varepsilon|W)\left(g^*(X_0) + g_0'(X_0)X'\theta^* + Z'\beta^*\right)}{\gamma_0^2(\varepsilon|W)}\dot{\gamma}(\varepsilon|W_0)\right] = 0$$
$$\mathrm{E}\left[\frac{\gamma^*(\varepsilon|W_0) - \gamma_0'(\varepsilon|W)\left(g^*(X_0) + g_0'(X_0)X'\theta^* + Z'\beta^*\right)}{\gamma_0^2(\varepsilon|W)}\gamma_0'(\varepsilon|W)\dot{g}(X_0)\right] = 0$$
$$-4\mathrm{E}\left[\frac{\gamma^*(\varepsilon|W_0) - \gamma_0'(\varepsilon|W)\left(g^*(X_0) + g_0'(X_0)X'\theta^* + Z'\beta^*\right)}{\gamma_0^2(\varepsilon|W)}\gamma_0'(\varepsilon|W)g_0'(X_0)X'\right]\dot{\theta} = c_1'\dot{\theta}$$
$$-4\mathrm{E}\left[\frac{\gamma^*(\varepsilon|W_0) - \gamma_0'(\varepsilon|W)\left(g^*(X_0) + g_0'(X_0)X'\theta^* + Z'\beta^*\right)}{\gamma_0^2(\varepsilon|W)}\gamma_0'(\varepsilon|W)Z'\right]\dot{\beta} = c_2'\dot{\beta}$$

The first equation implies $b^* \equiv 0$ once again. It is difficult to determine γ^* simply by looking at the next equation. After our experience with the independent case, we claim that $\gamma^* \equiv 0$ is a solution. Indeed the zero function is feasible; we need only to show that with this proposed solution we can find feasible solutions for the other elements of τ^*, and because the tangent space is closed, the Riesz Representation Theorem implies τ^* is unique.

Setting $\gamma^* \equiv 0$ and rewriting the remaining equations results in this system of equations:

$$\mathrm{E}\left[\psi_0^2(\varepsilon|W)\left(g^*(X_0) + g_0'(X_0)X'\theta^* + Z'\beta^*\right)\dot{g}(X_0)\right] = 0 \tag{4.1}$$
$$\mathrm{E}\left[\psi_0^2(\varepsilon|W)\left(g^*(X_0) + g_0'(X_0)X'\theta^* + Z'\beta^*\right)g_0'(X_0)X'\right]\dot{\theta} = c_1'\dot{\theta} \tag{4.2}$$
$$\mathrm{E}\left[\psi_0^2(\varepsilon|W)\left(g^*(X_0) + g_0'(X_0)X'\theta^* + Z'\beta^*\right)Z'\right]\dot{\beta} = c_2'\dot{\beta} \tag{4.3}$$

First rewrite these equations in terms of $\tilde{W}$ as was defined in the previous Section:

$$\mathrm{E}\left[\psi_0^2(\varepsilon|W)\left(g^*(X_0) + \tilde{W}'\xi^*\right)\dot{g}(X_0)\right] = 0 \tag{4.4}$$
$$\mathrm{E}\left[\psi_0^2(\varepsilon|W)\left(g^*(X_0) + \tilde{W}'\xi^*\right)\tilde{W}'\right]\dot{\xi} = c'\dot{\xi}. \tag{4.5}$$

The first equation implies

$$g^*(X_0) = -\frac{\mathrm{E}\left[\psi_0^2(\varepsilon|W)\tilde{W}'|X_0\right]}{\mathrm{E}\left[\psi_0^2(\varepsilon|W)|X_0\right]}\xi^*. \tag{4.6}$$

Plugging this into the second equation, one finds

$$\dot{\xi}'\mathrm{E}\left[\psi_0^2(\varepsilon|W)\tilde{W}\left(\tilde{W}-\frac{\mathrm{E}\left[\psi_0^2(\varepsilon|W)\tilde{W}|X_0\right]}{\mathrm{E}\left[\psi_0^2(\varepsilon|W)|X_0\right]}\right)'\right]\xi^*=\dot{\xi}'c. \tag{4.7}$$

The above equation can be rewritten

$$\dot{\xi}'\mathrm{E}\left[\tilde{S}\tilde{S}'\right]\xi^*=\dot{\xi}'c, \tag{4.8}$$

where

$$\tilde{S}=\psi_0(\varepsilon|W)\left(\tilde{W}-\frac{\mathrm{E}\left[\psi_0^2(\varepsilon|W)\tilde{W}|X_0\right]}{\mathrm{E}\left[\psi_0^2(\varepsilon|W)|X_0\right]}\right) \tag{4.9}$$

which provides a solution for ξ^*. Because all the other elements of τ^* satisfy all the necessary conditions, $\gamma^*\equiv 0$ is correct as well.

Gathering all the components of τ^* results in the following summary:

$$\xi^*=\mathrm{E}\left[\tilde{S}\tilde{S}'\right]^{-1}c \tag{4.10}$$

$$g^*(X_0)=-\frac{\mathrm{E}\left[\psi_0^2(\varepsilon|W)\tilde{W}'|X_0\right]}{\mathrm{E}\left[\psi_0^2(\varepsilon|W)|X_0\right]}\mathrm{E}\left[\tilde{S}\tilde{S}'\right]^{-1}c \tag{4.11}$$

$$\gamma^*\equiv 0,\quad b^*\equiv 0 \tag{4.12}$$

Putting these parts together results in

$$q^*(y|W)=\frac{1}{2}\gamma_0(\varepsilon|W)\tilde{S}'\mathrm{E}\left[\tilde{S}\tilde{S}'\right]^{-1}c$$

with $\tilde{S}$ given in (4.9).

Then the same calculations as in the independent case imply that the semiparametric efficiency bound for estimators of $c'\xi_0$ is $c'\mathrm{E}\left[\tilde{S}\tilde{S}'\right]^{-1}c$ and the bound for estimators of ξ_0 is $\mathrm{E}\left[\tilde{S}\tilde{S}'\right]^{-1}$, because c is arbitrary. □

This bound is similar to "Regression Model I" of [17, p.105-6] and Theorem 4 of [3]. It is uncommon to see the bound discussed in this section achieved by estimators in the literature. We are only aware of [13], which considers efficient estimation of a single index model related to (1.1) by setting $Z=0$. Interestingly, [13] offers something more general than it appears is claimed at first glance. There it is stated that the expected value of y conditional on covariates is of interest, so one might expect it is only as efficient as the specialized bound discussed below in Section 5.1. However, this expected value condition is only an identification condition — note that without more conditions, the model considered in Theorem 4.1 is not necessarily identified. Their estimator is a pseudo-maximum likelihood estimator that does not limit the investigator to consider only the expected value of y. Accordingly, the efficiency bound that they derive matches that given in Theorem 4.1. This is generally smaller than the bound one reaches if one estimates only the conditional expectation of y. As mentioned above, this more restrictive bound will be considered in Subsection 5.1. Estimators that reach this bound require an estimate of the likelihood function as in [13]; this is the strategy also followed by, for example, the L-estimators of [18] and [19] or section 4.4.1 of [20] for the linear model.

5. Conditional mean- and quantile identification

The result presented in Theorem 4.1 holds generally for both identification conditions presented below because it is written in terms of the likelihood. However, most practical estimators must deal with the fact that the likelihood is unknown. M-estimators constitute a class that enjoy some optimality properties and provide clear rules with regards to efficiency. We focus on these efficiency results here for two special M-estimators, those designed to estimate models when it may be assumed that the error term ε has mean zero, and when the α-quantile of its distribution is zero. We first consider the model with mean-zero error term, and then one in which the α^{th} quantile of the conditional distribution of ε is 0, and we discuss several estimators proposed in the literature (many of which are M-estimators) in this light. M-estimators remain quite attractive, even if they may appear less efficient than the optimum computed in Theorem 4.1 above. Namely, they can be tailored to balance considerations of robustness or efficiency (either of which may be an issue with small data sets), as the solutions to optimization problems they are computationally attractive, and they have a well-developed and tractable asymptotic theory. Furthermore, the conditional mean model and the conditional α-quantile model are easily estimated using M-estimators and have a "structural" meaning that is easy to interpret. The efficiency bound calculated in Theorem 4.1 is the bound for the finite-dimensional parameter of model (1.1), but the resulting regression model may then be a conditional *location* model instead of a model of e.g. the conditional mean of the response given covariates.

Nevertheless, Lemmas 5.1 and 5.2 show the necessary tradeoff that one makes when using (for example) an M-estimator. Focusing on, for example, one moment of the error distribution in the case of the conditional mean model, has the advantages listed above, but one loses efficiency as compared to the general bound presented in Theorem 4.1. The bounds obtained for M-estimators in [12] (or quasi-MLE methods represented by [3]) are for a smaller tangent space than the one considered here. That makes the methods used here complementary to those represented by, say, [12] or [17] — here we give efficiency bounds for model (1.1), rather than for a class of estimators of that model, in the spirit of Stein's "least-favorable submodel", and does not depend on the form of the estimators used. There does not appear to be a good way to tailor this method to classes of estimators. However, in the Lemmas below we show the difference between the efficiency bound for *all* estimators of a model and estimators of a specific form, like M-estimators.

5.1. Conditional mean identification

Now we specialize model (1.1) to conditional mean-zero error models; that is, to models that satisfy $\mathrm{E}[\varepsilon|W] = 0$ or $\mathrm{E}[y|W] = g_0(X_0)+Z'\beta_0$. We also require $\int_{\mathbb{R}} y^2 q_0^2(y|W)\mathrm{d}y < \infty$ $w.p.1$, where this second condition is an additional assumption that must be satisfied so that the bound is informative. Arguments given in [12] show that an M-estimator with weights depending on the inverse of $\sigma^2(\varepsilon|W)$ has a variance function equal to the efficiency bound (whether or not such a weighting function is feasible in practice). The semiparametric efficiency bound for M-estimators of ξ_0 in this model is

$$\Sigma_\mu = \mathrm{E}\left[\sigma(W)^{-2}\left(\tilde{W} - \frac{\mathrm{E}\left[\sigma^{-2}(W)\tilde{W}|X_0\right]}{\mathrm{E}\left[\sigma^{-2}(W)|X_0\right]}\right)\left(\tilde{W} - \frac{\mathrm{E}\left[\sigma^{-2}(W)\tilde{W}|X_0\right]}{\mathrm{E}\left[\sigma^{-2}(W)|X_0\right]}\right)'\right]^{-1} \tag{5.1}$$

where $\sigma^2(W) = \mathrm{E}\left[\varepsilon^2|W\right]$. The following Lemma shows that this bound is larger (in a positive semidefinite sense) than the bound given in Theorem 4.1. The proof of this Lemma is contained in the Appendix. This result comes from the fact that the model in this subsection only specifies one moment of the conditional distribution of the error conditional on covariates.

Lemma 5.1. *For model (1.1) with error term specified to have mean equal to zero,*

$$\Sigma_\mu - \mathrm{E}\left[\tilde{S}\tilde{S}'\right]^{-1}$$

is positive semidefinite, where Σ_μ is defined in (5.1) and $\tilde{S}$ is defined in Theorem 4.1.

An efficient estimator for the generalized partial linear single-index model has been proposed in [3]. More recently estimation of model 1.1 with homoskedastic errors was reconsidered in [10]. There they showed that accounting for the imposition of identification conditions on θ_0 can results in a more efficient estimator of θ_0 — although not β_0 — than was previously thought possible. This is also addressed for a generalized single-index model by [21].

Efficiency via weighting in a local-linear estimator of the single-index model (that is, model (1.1) with $Z = 0$) is considered in [22], who notes that as long as heteroskedasticity only depends on X_0, weighting by the inverse of the variance function results in an efficient estimator. In a single-index model, $\tilde{W} = g_0'(X_0)X$, and $\gamma_0(\varepsilon|X) = \gamma_0(\varepsilon|X_0)$. In the special case that the variance also depends only on X_0, the efficiency bound is

$$\mathrm{E}\left[\sigma(X_0)^{-2}\left(g_0'(X_0)\right)^2\left(X - \mathrm{E}\left[X|X_0\right]\right)\left(X - \mathrm{E}\left[X|X_0\right]\right)'\right]^{-1},$$

which agrees with the variance given in Theorem 5.2 of [22] when using the optimal weights $\omega(X_0) = \sigma^{-2}(X_0)$. The efficiency bound above also matches equation (4.4) of [12] when specialized to the mean-zero identification condition. When ε is homoskedastic, the variance function is not a function of X and the expression for the bound reduces further to

$$\sigma^2\mathrm{E}\left[\left(g_0'(X_0)\right)^2\left(X - \mathrm{E}\left[X|X_0\right]\right)\left(X - \mathrm{E}\left[X|X_0\right]\right)'\right]^{-1}. \tag{5.2}$$

All the above estimators are for the model (1.1) or the single-index model — that is, the model that results from setting $Z = 0$. We note that, although we focus on these two models, these results could be further specialized to the partially-linear and linear models. See [1] for more discussion of the bound in those models when error terms are independent. Our results augment the results of [1] to the case where the density of ε may depend on W.

5.2. Conditional quantile identification

Next we assume the model (1.1) is identified using the restriction $Q_{y|W}(\alpha|W) = g_0(X_0) + Z'\beta_0$, or $Q_{\varepsilon|W}(\alpha|W) = 0$ *w.p.*1. Expressed in terms of conditional density functions of ε in terms of models in the tangent space $\dot{\mathcal{T}}$, this means $\int_{-\infty}^{0}\gamma_t^2(u|W)\mathrm{d}u = \alpha$ *w.p.*1.

The semiparametric efficiency bound for M-estimators of ξ_0 under this restriction is

$$\Sigma_\alpha = \mathrm{E}\left[\frac{\gamma_0^4(0|W)}{\alpha(1-\alpha)}\left(\tilde{W} - \frac{\mathrm{E}\left[\gamma_0^4(0|W)\tilde{W}|X_0\right]}{\mathrm{E}\left[\gamma_0^4(0|W)|X_0\right]}\right)\left(\tilde{W} - \frac{\mathrm{E}\left[\gamma_0^4(0|W)\tilde{W}|X_0\right]}{\mathrm{E}\left[\gamma_0^4(0|W)|X_0\right]}\right)'\right]^{-1}. \tag{5.3}$$

The above bound for the α-quantile case matches that given in [12] when specializing the identification condition to a conditional quantile condition.

Lemma 5.2 shows that this bound is larger (in a positive definite sense) than the bound given in Theorem 4.1. Its proof is also contained in the Appendix. Once again, this bound is larger than the bound in Theorem 4.1 because only one quantile of the conditional distribution is specified in this model.

Lemma 5.2. *For model (1.1) with error term specified to have α-quantile equal to zero,*

$$\Sigma_\alpha - \mathrm{E}\left[\tilde{S}\tilde{S}'\right]^{-1}$$

is positive semidefinite, where Σ_α is defined in (5.3) and $\tilde{S}$ is defined in Theorem 4.1.

Recently [23] have discussed the case of general M-estimators for single-index models (i.e., with $Z = 0$), although their focus was on properties of a large class of M-estimators and not efficiency for the two special cases considered here. See also [8] for a thorough characterization of regularity conditions required for M-estimation of the single-index model (and model (1.1) by extension). In [8] an M-estimator for a single-index quantile regression is proposed; their estimator is not semiparametrically efficient, but their focus was on deriving a weak set of regularity conditions under which asymptotic distributions for a large set of estimators could be verified.

Efficient estimation of the partially-linear model $y = g_0(X) + Z'\beta_0 + \varepsilon$ was considered in [24], where $Q_{\varepsilon|W}(\tau|W) = 0$; this is a special case of the result of the bound for conditional α-quantile models given above. There may be reasons for considering the model (1.1), namely because this model may suffer from the curse of dimensionality (because g_0 is a multivariate function of X). Our results show that this estimator is not generally efficient. However, the discussion in [24] implies that weighting the estimator would result in efficiency only when the scale of the distribution of the error depends on X, and proposes a one-step estimator when the variance function is a general function of W (in that case weighting would not be sufficient to produce an efficient estimator of the type considered in the paper). This estimator remains quite attractive, however, due to the simplicity with which it may be implemented and its feasibility.

A local linear single-index quantile regression estimator is proposed in [7] — that is, an estimator of model (1.1) with $Z = 0$ and $Q_{\varepsilon|X}(\alpha|X) = 0$. They show that their estimator has the following asymptotic covariance matrix:

$$\Sigma = \alpha(1-\alpha)C_1^{-1}C_0C_1^{-1}, \tag{5.4}$$

where

$$C_j = \mathrm{E}\left[\left(\gamma_0^2(0|X)\right)^j \left(g_0'(X_0)\right)^2 (X - \mathrm{E}\left[X|X_0\right])(X - \mathrm{E}\left[X|X_0\right])'\right]. \tag{5.5}$$

It can be seen that this estimator is not efficient because the variance matrix does not match Σ_α given in (5.3). When the scale of the distribution of the error term only depends on X_0, the argument given in [25, p.161] implies that $C_1^{-1}C_0C_1^{-1} - C_2^{-1}$ is positive semidefinite (provided all the C_j are p.s.d.); C_2 is what results from using optimal weights $\omega(X_0) = \gamma_0^2(0|X_0)$, and it is what results from simplifying (5.3) to the case where the scale only depends on X_0. See [26] for a discussion of weighted linear median regression. As is pointed out in [27] and [26] for linear quantile regression models, weighted estimators should have smaller variance than unweighted estimators. However, under more general heteroskedasticity of ε it may not be possible to achieve efficiency simply by weighting; see the discussion in [24, p. 11]. In this case, sample splitting (used in [27]) or a one-step estimator (used in [24]) could be used. Another estimator was proposed in [28]; this estimator is not efficient, although it has the great advantage that it converges almost surely and is more efficient than average derivative estimators as represented by, say [29]. Finally, when ε is independent of X, the bound for the single index model becomes

$$\frac{\alpha(1-\alpha)}{\gamma_0^4(0)}\mathrm{E}\left[\left(g_0'(X_0)\right)^2 (X - \mathrm{E}\left[X|X_0\right])(X - \mathrm{E}\left[X|X_0\right])'\right]^{-1}$$

similar to expression (5.2).

6. Conclusion

We calculate the semiparametric efficiency bounds for the partially linear single-index model using the simple solution model demonstrated in [1]. Our results allow us to compare the asymptotic variance of popular M-estimators with this bound and to show that they are different in general.

Appendix A. Proof of Lemmas 5.1 and 5.2

Proof of Lemma 5.1. In order to find the information bound, we must find the minimum value that the information can take under the mean-zero condition. When considering the minimum value that ψ_0 may take, note that (cf. [20, p. 51])

$$
\begin{aligned}
1 = \left(2\int_{\mathbb{R}} u\gamma_0'(u|W)\gamma_0(u|W)\mathrm{d}u\right)^2 &\le \int_{\mathbb{R}} \psi_0^2(u|W)\gamma_0^2(u|W)\mathrm{d}u \int_{\mathbb{R}} u^2\gamma_0^2(u|W)\mathrm{d}u \\
&= \mathrm{E}\left[\psi_0^2(\varepsilon|W)|W\right]\sigma^2(W)
\end{aligned}
$$

($\sigma^2(W)$ is finite because the second moment of ε conditional on W is assumed to exist). Therefore we have the bound $\mathrm{E}\left[\psi_0^2(\varepsilon|W)|W\right] \ge \sigma^{-2}(W)$ for almost all W, where $\sigma^{-2}(W)$ denotes $(\sigma^2(W))^{-1}$. It can be verified that this inequality holds with equality at the normal model; that is, $\gamma_0^2(\varepsilon|W) = (2\pi\sigma^2(W))^{-1/2}\exp\{\frac{-\varepsilon^2}{2\sigma^2(W)}\}$.

Now consider the matrices

$$
\Xi = \begin{bmatrix} \psi_0^2(\varepsilon|W)\tilde{W}\tilde{W}' & \mathrm{E}\left[\psi_0^2(\varepsilon|W)\tilde{W}|X_0\right] \\ \mathrm{E}\left[\psi_0^2(\varepsilon|W)\tilde{W}'|X_0\right] & \mathrm{E}\left[\psi_0^2(\varepsilon|W)|X_0\right] \end{bmatrix} \tag{A.1}
$$

and

$$
\Omega = \begin{bmatrix} \sigma^{-2}(W)\tilde{W}\tilde{W}' & \mathrm{E}\left[\sigma^{-2}(W)\tilde{W}|X_0\right] \\ \mathrm{E}\left[\sigma^{-2}(W)\tilde{W}'|X_0\right] & \mathrm{E}\left[\sigma^{-2}(W)|X_0\right] \end{bmatrix}. \tag{A.2}
$$

Then $\mathrm{E}[\Xi - \Omega]$ is positive semidefinite since (using the relationship $\mathrm{E}\left[\psi_0^2(\varepsilon|W)|W\right] \ge \sigma^{-2}(W)$)

$$
\mathrm{E}[\Xi - \Omega] = \mathrm{E}\left[\left(\psi_0^2(\varepsilon|W) - \sigma^{-2}(W)\right)[\tilde{W}', 1]\begin{bmatrix}\tilde{W} \\ 1\end{bmatrix}\right]. \tag{A.3}
$$

This means that $\mathrm{E}\left[\Omega^{-1} - \Xi^{-1}\right]$ is positive semidefinite, and using a partitioned inverse formula for the upper-left $p \times p$ submatrix of each matrix, we have the result. □

Proof of Lemma 5.2. For a model of the α-quantile of the distribution of ε conditional on W, in a manner similar to [20, p. 54], consider that

$$
\begin{aligned}
(1-\alpha)\gamma_0^4(0|W) &= (1-\alpha)\left(\int_{-\infty}^0 2\gamma_0'(u|W)\gamma_0(u|W)\mathrm{d}u\right)^2 \\
&= (1-\alpha)\left(\int_{-\infty}^0 \frac{2\gamma_0'(u|W)}{\gamma_0(u|W)}\gamma_0(u|W)\gamma_0(u|W)\mathrm{d}u\right)^2 \\
&\le (1-\alpha)\int_{-\infty}^0 \psi_0^2(u|W)\gamma_0^2(u|W)\mathrm{d}u\int_{-\infty}^0 \gamma_0^2(u|W)\mathrm{d}u \\
&= \alpha(1-\alpha)\int_{-\infty}^0 \psi_0^2(u|W)\gamma_0^2(u|W)\mathrm{d}u
\end{aligned}
$$

where as before, $\psi_0(\varepsilon|W) = -2\gamma_0'(\varepsilon|W)/\gamma_0(\varepsilon|W)$. Similarly,

$$\begin{aligned}
\alpha\gamma_0^4(0|W) &= \alpha\left(-\int_0^\infty 2\gamma_0'(u|W)\gamma_0(u|W)\mathrm{d}u\right)^2 \\
&= \alpha\left(\int_0^\infty \frac{2\gamma_0'(u|W)}{\gamma_0(u|W)}\gamma_0(u|W)\gamma_0(u|W)\mathrm{d}u\right)^2 \\
&\leq \alpha\int_0^\infty \psi_0^2(u|W)\gamma_0^2(u|W)\mathrm{d}u\int_0^\infty \gamma_0^2(u|W)\mathrm{d}u \\
&= \alpha(1-\alpha)\int_0^\infty \psi_0^2(u|W)\gamma_0^2(u|W)\mathrm{d}u.
\end{aligned}$$

Adding these two inequalities, one finds that

$$\frac{\gamma_0^4(0|W)}{\alpha(1-\alpha)} \leq \int_{\mathbb{R}} \psi_0^2(u|W)\gamma_0^2(u|W)\mathrm{d}u = \mathrm{E}\left[\psi_0^2(\varepsilon|W)|W\right]$$

for almost all W. It can be verified that this inequality holds with equality at the "asymmetric Laplace" model, i.e., the model with density $\gamma_0^2(\varepsilon|W) = \frac{\alpha(1-\alpha)}{\sigma(W)}\exp\{-\frac{1}{\sigma(W)}\varepsilon(\alpha - I(\varepsilon < 0))\}$. This implies the bound in the statement of the Corollary using a similar method to that used in Lemma 5.1. □

References

[1] T. Severini, G. Tripathi, A simplified approach to computing efficiency bounds in semiparametric models, Journal of Econometrics 102 (2001) 23–66.

[2] T. Severini, G. Tripathi, Semiparametric efficiency bounds for microeconometric models: A survey, Foundations and Trends in Econometrics 6 (2-3) (2013) 163–397.

[3] R. Carroll, J. Fan, I. Gijbels, M. Wand, Generalized partly linear single-index models, Journal of the American Statistical Association 92 (438) (1997) 477–489.

[4] W. Härdle, P. Hall, H. Ichimura, Optimal smoothing in single-index models, The Annals of Statistics 21 (1993) 157–178.

[5] X. Chen, D. Pouzo, Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals, Journal of Econometrics 152 (2009) 46–60.

[6] X. Chen, D. Pouzo, Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals, Econometrica 80 (2012) 277–321.

[7] T. Wu, K. Yu, Y. Yu, Single-index quantile regression, Journal of Multivariate Analysis 101 (2010) 1607–1621.

[8] H. Ichimura, S. Lee, Characterization of the asymptotic distribution of semiparametric M-estimators, Journal of Econometrics 159 (2010) 252–266.

[9] Y. Xia, W. Härdle, Semi-parametric estimation of partially linear single-index models, Journal of Multivariate Analysis 97 (2010) 1162–1184.

[10] J. Wang, L. Xue, L. Zhu, Y. Chong, Estimation for a partial-linear single-index model, Annals of Statistics 38 (1) (2010) 246–274.

[11] Y. Fan, L. Zhu, Estimation of general semi-parametric quantile regression, Journal of Statistical Planning and Inference 143 (2013) 896–910.

[12] W. Newey, T. Stoker, Efficiency of weighted average derivative estimators and index models, Econometrica 61 (1993) 1199–1223.

[13] M. Delecroix, W. Härdle, M. Hristache, Efficient estimation in conditional single-index regression, Journal of Multivariate Analysis 86 (2003) 213–226.

[14] I. van Keilegom, L. Wang, Semiparametric modeling and estimation of heteroscedasticity in regression analysis of cross-sectional data, Electronic Journal of Statistics 4 (2010) 133:160.

[15] A. van der Vaart, Asymptotic Statistics, Cambridge, 1998.

[16] I. Komunjer, Q. Vuong, Semiparametric efficiency bound in time-series models for conditional quantiles, Econometric Theory 26 (2009) 383–405.
[17] P. Bickel, C. Klaasen, Y. Ritov, J. Wellner, Efficient and Adaptive Estimation for Semiparametric Models, Springer, 1998.
[18] R. Koenker, S. Portnoy, L-estimation for linear models, Journal of the American Statistical Association 82 (399) (1987) 851–857.
[19] S. Portnoy, R. Koenker, Adaptive L-estimation for linear models, The Annals of Statistics 17 (1) (1989) 362–381.
[20] M. Kosorok, Introduction to Empirical Processes and Semiparametric Inference, Springer, 2008.
[21] X. Cui, W. Härdle, L. Zhu, The EFM approach for single-index models, Annals of Statistics 39 (3) (2011) 1658–1688.
[22] H. Ichimura, Semiparametric least squares (SLS) and weighted SLS estimation of single-index models, Journal of Econometrics 58 (1993) 71–120.
[23] M. Delecroix, M. Hristache, V. Patilea, On semiparametric M-estimation in single-index regression, Journal of Statistical Planning and Inference 136 (2006) 730–769.
[24] S. Lee, Efficient semiparametric estimation of a partially linear quantile regression model, Econometric Theory 19 (2003) 1–31.
[25] R. Koenker, Quantile Regression, Cambridge, 2005.
[26] Q. Zhao, Asymptotically efficient median regression in the presence of heteroskedasticity of unknown form, Econometric Theory 17 (2001) 765–784.
[27] W. Newey, J. Powell, Efficient estimation of linear and type I censored regression models under conditional quantile restrictions, Econometric Theory 6 (1990) 295–317.
[28] E. Kong, Y. Xia, A single-index quantile regression model and its estimation, Econometric Theory 28 (2012) 730–768.
[29] P. Chaudhuri, K. Doksum, A. Samarov, On average derivative quantile regression, Annals of Statistics 25 (2) (1997) 715–744.