Econometric Theory, **34**, 2018, 574–597.
doi:10.1017/S0266466617000093

NONPARAMETRIC INSTRUMENTAL VARIABLES AND REGULAR ESTIMATION

JINYONG HAHN AND ZHIPENG LIAO
University of California, Los Angeles

This paper investigates whether there can exist regular estimators in models characterized by nonparametric instrumental variable (NPIV). We show by a number of examples that regular estimation is impossible in general for nonlinear functionals.

1. INTRODUCTION

There has been a lot of interest in nonparametric objects characterized as solutions to linear integral equations. The nonparametric instrumental variable (NPIV) model takes the form

$$Y_1 = h_0(Y_2) + \varepsilon \text{ with } \mathbb{E}[\varepsilon|X] = 0, \tag{1}$$

where $Y = (Y_1, Y_2)$ are endogenous variables, X is the (nonparametric) instrumental variable (IV) for Y_2, and $h_0(\cdot)$ is the infinite dimensional parameter in some function space $\mathcal{H}$. The literature has focused on identification and consistent estimation of the infinite dimensional parameter $h_0(\cdot)$, which is subject to the well-known ill-posed inverse problem. See, e.g., Newey and Powell (2003), Carrasco, Florens, and Renault (2007), Florens, Johannes, and van Bellegem (2011), Chen and Reiss (2011), and d'Haultfoeuille (2011).

In this paper, we pose a semiparametric question, and study certain statistical properties of some *functionals* of $h_0(\cdot)$. Specifically, we study whether a regular estimator may exist for a finite dimensional parameter $\theta_0 = \mathbb{E}[\psi(h_0(Y_2))]$, where $\psi(\cdot)$ is a real-valued (and possibly nonlinear) function. In the exogenous case where $Y_2 = X$, it is well known that a regular estimator exists as long as the $h_0(\cdot)$ can be estimated at the speed of $o_p\left(n^{-1/4}\right)$, and the asymptotic variance of the regular estimator can be characterized in a straightforward way. See, e.g., Newey (1994). For the NPIV case where $Y_2 \neq X$, it is well known that the rate of convergence for estimation of the nonparametric object $h_0(\cdot)$ is slower than in the exogenous case. See, e.g., Horowitz (2011). Our objective is to understand whether such a slow rate in NPIV affects inference of the functionals of $h_0(\cdot)$.

We acknowledge helpful comments by anonymous referees. We are particularly grateful to the Editor and a Co-Editor for their considerable input. Address correspondence to Jinyong Hahn, Department of Economics, UCLA, Box 951477, Los Angeles, CA 90095-1477, USA; e-mail: hahn@econ.ucla.edu.

 © Cambridge University Press 2017

https://doi.org/10.1017/S0266466617000093 Published online by Cambridge University Press

We show by a few examples that regular estimation is impossible in general for *nonlinear* functionals of $h_0(\cdot)$ in models with NPIV, at least without restricting the parameter space. This is done by following van der Vaart's (1991, Sect. 2) argument, and deriving a mathematical contradiction. Regular estimability is often understood to be equivalent to the feasibility of root-n-consistent estimation,[1] and hence, our result implies that root-n-consistent estimation is impossible for many estimands in models characterized by NPIV.

2. OVERVIEW

Our model is given by

$$\mathbb{E}[Y_1 - h_0(Y_2)|X] = 0, \tag{2}$$

$$\mathbb{E}[\psi(h_0(Y_2))] = \theta_0, \tag{3}$$

where the expectations are taken under the true joint distribution F_0 of $Z = (Y_1, Y_2, X)$, and $\psi(\cdot)$ is a known function. Note that equations (2) and (3) are a special case of the model considered in Ai and Chen (2007, 2012), among others, who proposed a $\sqrt{n}$-consistent estimator of θ_0. We discuss the restrictive nature of a specific assumption in Ai and Chen (2007), and argue how the feasibility of regular estimation may help us understand the problem better. We then go on to discuss a certain implication of van der Vaart's (1991) analyses, which forms a basis of our counterexamples afterwards.

In Appendix A, we examine Ai and Chen's (2007) Assumption 4.4 in detail. For the case $\psi(h_0(Y_2)) = \frac{1}{2}(h_0(Y_2))^2$, it is shown that their Assumption 4.4 essentially requires that $\mathbb{E}\left[(h_n(Y_2))^2\right] = o(n^{-\frac{1}{2}})$ for any sequence $h_n(\cdot)$ such that $\mathbb{E}[(\mathbb{E}[h_n(Y_2)|X])^2] = o(n^{-\frac{1}{2}})$. This is problematic because the rate on the conditional mean $\mathbb{E}[h_n(Y_2)|X]$ does not restrict the conditional variance $\mathrm{Var}(h_n(Y_2)|X)$.

Ai and Chen's (2007) Assumption 4.4 is phrased as a sufficient condition for the $\sqrt{n}$-consistency and asymptotic normality results. Because it is not a necessary condition, it is logically possible that their results may be valid even when Assumption 4.4 is violated. We show here that it is not the case; specifically, regular estimation is impossible in general, at least when no restriction is imposed on the parameter space.

Our argument is based on deriving a mathematical contradiction using van der Vaart's (1991) analyses, and we will start by reviewing some of the relevant materials there. Let $\mathcal{F}$ denote a collection of probability distributions, and let $F \in \mathcal{F}$ denote a fixed point. Let $L_2(F)$ be the set of measurable, F-square integrable functions, with inner product $\langle s_1, s_2 \rangle_F = \int s_1 s_2 dF$ and the corresponding norm $\|\cdot\|_F$. Let $\mathcal{F}(F)$ be a collection of maps $t \to F_t$ from an interval $(0, \epsilon) \subset \mathbb{R}$ to $\mathcal{F}$, satisfying for some $s \in L_2(F)$,

$$\int \left(\frac{\sqrt{dF_t} - \sqrt{dF}}{t} - \frac{1}{2} s \sqrt{dF} \right)^2 \to 0 \text{ as } t \to 0, \tag{4}$$

https://doi.org/10.1017/S0266466617000093 Published online by Cambridge University Press

where dF_t and dF denote the densities of F_t and F respectively. To $\mathcal{F}(F)$ corresponds a tangent space $T(F)$, consisting of all elements s above. We assume that $t \to F_{at}$ is in $\mathcal{F}(F)$ for every $a \in \mathbb{R}^+$ whenever this is true of the map $t \to F_t$. Now, $\kappa : \mathcal{F} \to (\mathbf{B}, \|\cdot\|)$ is *differentiable* at F relative to $\mathcal{F}(F)$ if there exists a continuous, linear map $\kappa'_F : (T(F), \|\cdot\|_F) \to (\mathbf{B}, \|\cdot\|)$ such that

$$\frac{\kappa(F_t) - \kappa(F)}{t} \to \kappa'_F(s) \text{ as } t \to 0 \tag{5}$$

for every path $t \to F_t$ in $\mathcal{F}(F)$. We can take $\mathbf{B} = \mathbb{R}$ for our purpose, although we keep the $\mathbf{B}$ symbol to make comparison with van der Vaart (1991) easier.

Let T_n denote an estimator based on n IID observations $Z_1, \ldots, Z_n$. The estimator T_n is *regular* if for every path $t \to F_t$ in $\mathcal{F}(F)$ and $a_n \to a \in \mathbb{R}$, the distribution of $\sqrt{n}(T_n - \kappa(F_{a_n/\sqrt{n}}))$ under $F_{a_n/\sqrt{n}}$ converges weakly to some fixed tight Borel law L on $\mathbf{B}$, and the limiting distribution L is the same for every path $\{F_t\}$. As in Newey (1994), we define T_n to be asymptotically linear if $\sqrt{n}(T_n - \kappa(F_0)) = n^{-1/2}\sum_{i=1}^n \zeta(Z_i) + o_p(1)$ with $\mathbb{E}[\zeta(Z_i)] = 0$ and $\mathrm{Var}(\zeta(Z_i)) < \infty$ under F_0. Most familiar econometric estimators are asymptotically linear, and without too much loss of practical generality, we restrict our attention to the class of asymptotically linear estimators.

Our result is based on the following proposition:

PROPOSITION 2.1. *Suppose that $\kappa : \mathcal{F} \to (\mathbf{B}, \|\cdot\|)$ is not differentiable at F relative to $\mathcal{F}(F)$. Then there cannot exist a regular asymptotically linear estimator of $\kappa(F_0)$.*

Proof. Because of asymptotic linearity, the joint distribution of

$$\left(\sqrt{n}(T_n - \kappa(F_0)), n^{-1/2}\sum_{i=1}^n s(Z_i)\right) = \left(n^{-1/2}\sum_{i=1}^n \zeta(Z_i), n^{-1/2}\sum_{i=1}^n s(Z_i)\right) + o_p(1)$$

under F_0 converges weakly for every $s \in T(F)$. The conclusion therefore follows as a trivial corollary of van der Vaart (1991, Thm. 2.1). ■

In our application, the $F = F_0$ is the true distribution of (Y_1, Y_2, X) which satisfies (2). We will assume that the joint distribution of (Y_2, X) satisfies the completeness condition. The set $\mathcal{F}$ is the collection of all distributions of (Y_1, Y_2, X) such that there exists some h satisfying (2). The $\kappa(F) = \mathbb{E}[\psi(h_0(Y_2))] = \theta_0$.

Our result is based on a simple reasoning. We consider a path F_t such that (4) is satisfied with $s = 0$ almost surely (a.s.). If the path is differentiable relative to $\mathcal{F}(F)$, we should have $(\kappa(F_t) - \kappa(F))/t \to 0$ in view of (5). If $(\kappa(F_t) - \kappa(F))/t \not\to 0$, we should conclude that there cannot exist a regular asymptotically linear estimator.

3. COUNTER-EXAMPLES

In this section, we present a few counter-examples where (8) is satisfied. The contradiction that we will derive is based on a reduced form characterization of

https://doi.org/10.1017/S0266466617000093 Published online by Cambridge University Press

the model (1):

$$Y_1 = r_0(X) + u,$$
$$Y_2 = \pi_0(X) + v,$$

where $r_0(X) = \mathbb{E}[Y_1|X]$ and $\pi_0(X) = \mathbb{E}[Y_2|X]$.

3.1. Path Considered

We will fix π_0 as well as the distribution of (u, v, X), and consider a path obtained by perturbing r_0. Under the path indexed by the perturbation of r_0, the conditional distribution of Y_2 given X is fixed, because the joint distribution of (v, X) and the conditional mean function $\pi_0(\cdot)$ are fixed. By keeping the conditional distribution of Y_2 given X fixed, we maintain completeness and identification of h over the path.

Let $g_0(u, v, x)$ denote the true joint density of (u, v, X). The path F_t we consider is such that its derivative $dF_t = f_t$ is given by

$$f_t = g_0[y_1 - r_t(x), y_2 - \pi_0(x), x], \tag{6}$$

where r_t denotes the perturbation of r_0. It can be shown that under some regularity conditions, $s = 0$ is the score of the path F_t constructed in (6) if r_t satisfies $t^{-1}(r_t - r_0) \to 0$ in quadratic mean. (See Lemma B.1 in the Appendix for the theoretical justification.)

In the examples below, we will find r_t (thereby F_t) such that (i)

$$\mathbb{E}\left[\left(t^{-1}[r_t(X) - r_0(X)]\right)^2\right] \to 0; \tag{7}$$

yet $\theta_t = \kappa(F_t)$ is such that (ii)

$$\frac{\theta_t - \theta_0}{t} \not\to 0. \tag{8}$$

In view of Proposition 2.1, we can conclude that there cannot exist a regular estimator of θ_0. For simplicity, we will work with $t = n^{-1/2}$.

3.2. Example 1 - Nonlinear Functional under Mild Ill-Posedness

Suppose that (Y_2, X) has the joint density

$$b_0(y_2, x) = \sum_{k=1}^{\infty} \sqrt{\rho_k}\varphi_k(y_2)\varphi_k(x), \text{ for any } (y_2, x) \in [0,1] \times [0,1], \tag{9}$$

where $\varphi_1(x) = 1$, $\rho_1 = 1$ and for $k > 1$, $\varphi_k(x) = \sqrt{2}\cos[(k-1)\pi x]$ and $\rho_k = k^{-a}$ ($a > 2$) is a decaying sequence.[2] It is clear that the marginal distributions of Y_2 and X are uniform on $[0, 1]$. Moreover, $\{\varphi_k\}_{k=1}^{\infty}$ is a set of orthonormal basis functions in $L_2([0, 1])$ (see, e.g., Bierens (2013)). Consider an arbitrary function $r_0 \in L_2([0, 1])$ such that the integral equation $\mathbb{E}[h(Y_2)|X] = r_0(X)$ has a unique

https://doi.org/10.1017/S0266466617000093 Published online by Cambridge University Press

solution $h_0 \in L_2([0,1])$ almost surely. Our parameter of interest is

$$\theta_0 = \frac{1}{2}\mathbb{E}\left[h_0^2(Y_2)\right]. \tag{10}$$

We consider the following path/perturbation of $r_0(x)$:

$$r_{n^{-1/2}}(x) = r_0(x) + n^{-\frac{1}{2}} \sum_{k=1}^{\infty} \phi_{n^{-1/2},k} \varphi_k(x),$$

where $\phi_{n^{-1/2},k} = n^{-\frac{1}{8}} k^{-\frac{1}{2}-\frac{a}{4}} I\left\{k \leq n^{\frac{3}{2a}}\right\}$. For this path, it can be shown[3] that (7) and (8) are satisfied, i.e.,

$$\lim_{n\to\infty} \mathbb{E}\left[n \left|r_{n^{-1/2}}(X) - r_0(X)\right|^2\right] = 0, \quad \lim_{n\to\infty} n^{\frac{1}{2}}\left(\theta_{n^{-1/2}} - \theta_0\right) = \frac{1}{a},$$

and therefore, we can conclude that there cannot exist a regular estimator of θ_0.

Despite the simplistic appearance of this example, we would like to emphasize that the problem is inherent in any nonlinear functional of h. For this purpose, note that

$$\begin{aligned} n^{\frac{1}{2}}\left(\theta_{n^{-1/2}} - \theta_0\right) &= n^{\frac{1}{2}}\left\{\mathbb{E}\left[\frac{1}{2}h^2_{n^{-1/2}}(Y_2)\right] - \mathbb{E}\left[\frac{1}{2}h_0^2(Y_2)\right]\right\} \\ &= \mathbb{E}\left[n^{\frac{1}{2}}\left(h_{n^{-1/2}}(Y_2) - h_0(Y_2)\right)h_0(Y_2)\right] + \frac{1}{2}\mathbb{E}\left[n^{\frac{1}{2}}\left(h_{n^{-1/2}}(Y_2) - h_0(Y_2)\right)^2\right]. \end{aligned}$$

Although the details of the proof[4] are technical in nature, the gist of our contradiction is based on the fact that the second order term $\frac{1}{2}\mathbb{E}\left[n^{\frac{1}{2}}\left(h_{n^{-1/2}}(Y_2) - h_0(Y_2)\right)^2\right]$ does not converge to zero. Our parameter (10) of interest has a convenient expression that facilitates analysis of such a higher order term, yet the problem is fundamentally due to the nonlinearity. As such, we expect that such problems with higher order terms would be the norm rather than the exception for other nonlinear problems as well. Our next example shows that a qualitatively identical problem arises under severe ill-posedness as well.

3.3. Example 2 - Quadratic Functional under Severe Ill-Posedness

Suppose that (Y_2, X) is joint normal with mean zero, unit variance and covariance $\rho^{\frac{1}{2}} \in (0,1)$. Let $p_0(\cdot)$ denote the standard normal density. Also let

$$\varphi_k(\cdot) = \frac{(-1)^k}{\sqrt{k!}} \frac{p_0^{(k)}(\cdot)}{p_0(\cdot)} \tag{11}$$

denote the k-th ($k = 0, 1, 2, \ldots$) normalized Hermite polynomial, where $p_0^{(k)}(\cdot)$ is the k-th derivative of $p_0(\cdot)$ and we define $0! = 1$. Let $L_2(R, p_0)$ denote the set of functions such that for any $f(\cdot) \in L_2(R, p_0)$, $\int_R f(x)^2 p_0(x)\,dx < \infty$. It is well known that the Hermite polynomials $\{\varphi_k\}_{k=0}^{\infty}$ is a set of orthonormal basis functions in $L_2(R, p_0)$ (see, e.g., Johnston (2014)). Consider an arbitrary function

https://doi.org/10.1017/S0266466617000093 Published online by Cambridge University Press

$r_0 \in L_2(R, p_0)$ such that the integral equation $\mathbb{E}[h(Y_2)|X] = r_0(X)$ has a unique solution $h_0 \in L_2(R, p_0)$ almost surely. Our parameter of interest is

$$\theta_0 = \frac{1}{2}\mathbb{E}\left[h_0^2(Y_2)\right].$$

We consider the following perturbation of $r_0(x)$:

$$r_{n^{-1/2}}(x) = r_0(x) + n^{-\frac{1}{2}} \sum_{k=0}^{\infty} \phi_{n^{-1/2},k}\varphi_k(x),$$

where $\phi_{n^{-1/2},k} = n^{\frac{1}{4}}\rho^{\frac{k}{2}} I\{n^{\frac{1}{2}} \leq k < n^{\frac{1}{2}} + 1\}$ is a real sequence. It can be shown[5] that (7) and (8) are satisfied, i.e.,

$$\lim_{n\to\infty} \mathbb{E}\left[\left|n^{\frac{1}{2}}\left[r_{n^{-1/2}}(X) - r_0(X)\right]\right|^2\right] = 0, \quad \lim_{n\to\infty} n^{\frac{1}{2}}\left(\theta_{n^{-1/2}} - \theta_0\right) = \frac{1}{2},$$

and therefore, we can conclude that there cannot exist a regular estimator of θ_0.

3.4. Restriction of Parameter Space

In both examples, one may wonder whether we worked with too large a space in which a possible value of r can reside. In fact, an anonymous referee pointed out that $\sqrt{n}$-consistent estimation is feasible if the restricted parameter space consists of rs that can be written as $r(x) = \sum_{k=1}^{\infty} \phi_k \varphi_k(x)$ such that $\sup_k k^b |\phi_k| < C$ for some positive $C < \infty$ and $b > a + 1$. For $\widehat{\phi}_k = n^{-1}\sum_{i=1}^{n} Y_{2i}\varphi_k(X_i)$, and $M_n = \left\lfloor n^{\frac{1}{2b}} \right\rfloor$, consider a natural estimator $\widehat{\theta} = 2^{-1}\sum_{k=1}^{M_n} \widehat{\phi}_k^2 \rho_k^{-1}$. In order to show that the $\widehat{\theta}$ is $\sqrt{n}$-consistent, one can note that

$$\widehat{\theta} - \theta_0 = \sum_{k=1}^{M_n} \frac{\phi_k\left(\widehat{\phi}_k - \phi_k\right)}{\rho_k} + \frac{1}{2}\sum_{k=1}^{M_n} \frac{\left(\widehat{\phi}_k - \phi_k\right)^2}{\rho_k} + \frac{1}{2}\sum_{k=M_n+1}^{\infty} \frac{\phi_k^2}{\rho_k}.$$

It can be shown that the second term and third remainder term on the right are of order $o_p\left(n^{-1/2}\right)$ under such a restriction, and the asymptotic distribution is therefore determined by the first term, which is typically asymptotically normal under standard conditions.

While mathematically correct, the estimator above is predicated on the assumption that the ρ_k and b are *known* to the econometrician, and as such, it is an infeasible estimator in practice. The perturbation/path that we considered in Example 1 was such that the joint density of (Y_2, X) is fixed, so one might argue that the above estimator is a "feasible" estimator along the path. We do not think such an argument is persuasive. In any case, regular estimability needs to be verified for every possible path, including the one that perturbs the joint density of (Y_2, X), so it would not be sensible to use the above estimator as evidence of the feasibility of regular estimation.

https://doi.org/10.1017/S0266466617000093 Published online by Cambridge University Press

We therefore consider a different path where the joint density of (Y_2, X) is perturbed, while fixing $r_0(X)$ as well as the conditional density of u given (Y_2, X). This is an extreme case where the parameter space for r is so restricted that it consists of a singleton. Let $\varphi_1(x) = 1$ and $\varphi_k(x) = \cos[(k-1)\pi x]$ for any $k > 1$. We will assume that $r_0(x) = \sum_{k=1}^{\infty} \phi_k \varphi_k(x) = \sum_{k=1}^{\infty} k^{-b} \varphi_k(x)$ for some $b > 0$ sufficiently large yet finite and fixed. We let

$$b_0(y_2, x) = 1 + 2\sqrt{C} \sum_{k=2}^{\infty} k^{-\frac{a}{2}} \varphi_k(y_2) \varphi_k(x),$$

$$b_{n^{-1/2}}(y_2, x) = b_0(y_2, x) + 2\sqrt{C} \sum_{k=n^{\delta}}^{\infty} \left(\frac{1}{k^{a'/2} + k^{a/2}} - \frac{1}{k^{a/2}} \right) \varphi_k(y_2) \varphi_k(x) \quad \textbf{(12)}$$

for some $\delta > 1/(a-1)$ and a' such that $1 + \delta(a' + a - 4b + 2) < 0$ and $1 + \delta(2a' - 4b + 2) \geq 0$. Adopting the restriction $b \geq a + 1$ as suggested by the anonymous referee, we can see that such an a' always exists, and it can be shown[6] that $\int n\left(\sqrt{b_{n^{-1/2}}(y_2, x)} - \sqrt{b_0(y_2, x)}\right)^2 \to 0$, but $\mathbb{E}\left[n^{\frac{1}{2}}\left(h^2_{n^{-1/2}}(Y_2) - h_0^2(Y_2)\right)\right] \nrightarrow 0$. This is the same kind of contradiction discussed in Example 1, except that we now consider a path such that the joint density of (Y_2, X) is perturbed, which essentially reflects uncertainty about a, or the decay rate of the orthonormal representation.

The above example is such that for every $b_0(y_2, x)$, there exists a virtually indistinguishable joint density of (Y_2, X) whose spectrum decays with a fast rate such that regular estimation becomes infeasible. As such, we are uncomfortable imposing a restriction on the parameter space, and claiming feasibility of regular estimation. In case a restriction is imposed on a, and the further restriction $b > a + 1$ is employed, it is insightful to discuss the mathematical implication of the restriction $b > a + 1$. According to Hall and Horowitz (2005, Thm. 4.4), the optimal rate of convergence under the L_2-norm for estimation of h_0 is $O\left(n^{-(2\beta-1)/2(2\beta+\alpha)}\right)$ for some α and β. The two parameters α and β in Hall and Horowitz (2005) correspond to our a and $b - a/2$, respectively. It follows that the condition $b > a + 1$ implies that the optimal rate of convergence under the L_2-norm is of order $o\left(n^{-1/4}\right)$. In other words, the nonparametric estimator of h_0 can have a rate of convergence $o\left(n^{-1/4}\right)$ when $b > a + 1$, which is a classic condition under which $\sqrt{n}$-consistency can be shown to be feasible, as discussed by Newey (1994, pp. 1364–1365). We should also note that the optimal rate is $O_p\left((\log n)^{-s}\right)$ for Example 2 (Horowitz, 2011, p. 366), so such a mathematical possibility should be ruled out for the severely ill-posed case.

Finally, we note that the parameter space restriction discussed above implies that even if one were willing to impose the parameter space restriction and claim feasibility of regular estimation, one would have to accept that the nature of "regularity" in NPIV is different from the exogenous case, i.e., when $Y_2 = X$ for the model (2) and (3). It is often the case that some smoothness condition is imposed on h_0 to ensure that its natural estimator $\widehat{h}$ converges at a fast enough

https://doi.org/10.1017/S0266466617000093 Published online by Cambridge University Press

rate, which in turn ensures that the intuitive sample analog estimator $\widehat{\theta}$ of θ_0 is $\sqrt{n}$-consistent. This does not mean that the same kind of parameter space restriction needs to be imposed on the perturbation of h_0 in the exogenous case, even though it may sound contradictory at first. In order to appreciate the subtlety, it is helpful to note that there are two kinds of parameter spaces implicitly discussed there; one needs to consider the parameter space for the h_0, as well as the parameter space for the perturbation, say $h_{n^{-1/2}}$, around h_0. It can be shown by using Le Cam's Third Lemma that $\sqrt{n}\left(\widehat{\theta}-\theta_{n^{-1/2}}\right)$ has the same asymptotic distribution as long as $\theta_{n^{-1/2}} = \frac{1}{2}\mathbb{E}\left[h^2_{n^{-1/2}}(Y_2)\right]$ is such that $F_{n^{-1/2}}$ satisfies (4), and no other restriction needs to be imposed on $h_{n^{-1/2}}$ for this purpose. On the other hand, differentiability in the NPIV case requires that we rule out more paths than is usual under exogeneity. As a consequence, the "regularity" in the endogenous case should be interpreted as a limited version even with parameter space restriction.[7]

4. DISCUSSION

The argument in previous sections is superficially similar to Chamberlain (1986) and Hahn (1994) in that impossibility of regular estimation was established. The difference is that those two papers used characterization of semiparametric efficiency bounds through the "standard" arguments[8] based on projection algebra in some Hilbert space. Our analysis does not require the standard projection argument; the projection argument is predicated on the assumption that the basic differentiability property is satisfied, and our discussion established that such a differentiability property does not hold (at least when the parameter space is not restricted).[9]

In all our counter-examples, the differentiability condition was shown to fail through the "higher order" term, which suggests that for parameters that are "linear" in h, we may apply standard arguments.[10] Indeed, Severini and Tripathi (2012) worked with $\mathbb{E}[\ell_0(Y_2)h(Y_2)]$ in the NPIV case, and required that the solution $v^*_{\ell_0}$ of the integral equation $\mathbb{E}\left[v_{\ell_0}(X)|Y_2\right] = \ell_0(Y_2)$ is well defined in order for κ to be differentiable at F *relative to* $\mathcal{F}(F)$, i.e., they noted that the existence of $v^*_{\ell_0}$ is necessary for $\sqrt{n}$-consistency even for linear functionals of h.[11] Severini and Tripathi's (2012) discussion is related to van der Vaart (1991, Sect. 7), who studies the linear functional $\kappa(h) = \int \ell_0(z)h(z)dz$ in a semiparametric mixture model, where the observable X has the density

$$p_h(x) = \int p(x,z)h(z)dz \tag{13}$$

and $p(x,z)$ is the conditional density of X given Z and is assumed to be a known element in the exponential family. Van der Vaart (1991) shows that for $\kappa(h)$ to be $\sqrt{n}$-consistently estimable, there should exist some square integrable function $g(x)$ satisfying

$$\ell_0(z) = \int g(x)p(x,z)dx. \tag{14}$$

https://doi.org/10.1017/S0266466617000093 Published online by Cambridge University Press

We note in passing that van der Vaart's (1991, Sect. 7) result has an implication for measurement error problems in econometrics. Suppose that $X = (X_1, X_2)$ constitutes two independent measurements of Z, and that we know the measurement errors $U_1 = X_1 - Z$ and $U_2 = X_2 - Z$ are such that their distributions are independent $N(0,1)$. Suppose that our objective of interest is $\kappa(h) = \int \ell_0(z)h(z)dz$, where again $\ell_0(z)$ is a known function and $h(z)$ is the (unknown) density of Z. His analysis goes through unchanged, because in this case, we can write the density of observable X in terms of (13) with $p(x,z) = \phi_0(x_1 - z, x_2 - z)$, where $\phi_0(\cdot,\cdot)$ denotes the standard bivariate normal density with correlation zero. Hence, we can conclude that $\int \ell_0(z)h(z)dz$ is a differentiable functional (i.e., regular estimation is potentially possible) only if there exists a function $g(x_1, x_2)$ such that $\mathbb{E}[g(X_1, X_2)|Z] = \ell_0(Z)$. In other words, regular estimation of $\mathbb{E}[\ell_0(Z)]$ is impossible if there is no such g. Utilizing the sufficiency of $\overline{X} = (X_1 + X_2)/2$, we can conclude that every function of Z that can be written as $\mathbb{E}[g(X_1, X_2)|Z]$ for some g can also be written as $\mathbb{E}\left[\tilde{g}\left(\overline{X}\right)|Z\right]$ for some $\tilde{g}$, and vice versa. Given that $\overline{X}$, conditional on Z, has a normal distribution with mean Z and variance 1/2, we can find some $\ell_0(Z)$ for which there exists no $\tilde{g}$ such that $\ell_0(Z) = \mathbb{E}\left[\tilde{g}\left(\overline{X}\right)|Z\right]$.[12] For such $\ell_0(Z)$, we should conclude that regular estimation of $\mathbb{E}[\ell_0(Z)]$ is impossible.[13] This contradicts Schennach (2004, Thm. 2), who proposes a $\sqrt{n}$-consistent estimator for arbitrary $\mathbb{E}[\ell_0(Z)]$; Schennach (2004) works with a more difficult model where the distributions of the measurement errors are nonparametrically specified, but van der Vaart (1991, Sect. 7) implies that it cannot be true even for the simpler model where the error distribution is parametrically specified.

5. SUMMARY

In this paper, we followed van der Vaart (1991), and developed a few examples where regular estimation is impossible. The examples worked with quadratic terms that do not converge to zero. One can easily extend these examples, and conclude that regular estimation is impossible for expectations of strictly concave (or convex) functions, as long as the second derivatives are bounded away from zero.[14]

Our analysis leads to the following conclusions. (i) Regular estimation is impossible with severe ill-posedness; (ii) it may be possible with mild ill-posedness by imposing restrictions on a parameter space, but the restriction most likely requires that the nonparametric estimator of h_0 can have a rate of convergence $o\left(n^{-1/4}\right)$.

NOTES

1. Chamberlain (1986), e.g., adopts a similar intuition/convention; he derives impossibility of regular estimation for a binary response model with median restriction, and summarizes the finding in the abstract by writing that it is impossible to attain root-n-consistency.

2. The condition $a > 2$ ensures that the series representation of the joint density $b_0(y_2, x)$ is convergent uniformly in $(y_2, x) \in [0,1] \times [0,1]$.

3. See Lemmas C.1 and C.2 in Appendix C.

https://doi.org/10.1017/S0266466617000093 Published online by Cambridge University Press

4. See Lemmas C.1 and C.2 in Appendix C.
5. See Lemmas C.3 and C.4 in Appendix C.
6. See Lemma D.2 in Appendix D.
7. We thank another anonymous referee, who gave us very helpful comments on this issue.
8. See, e.g., Newey (1990) and van der Vaart (1991, Sects. 3–6).
9. Put another way, we did not use the classic argument contained in van der Vaart's (1991) Theorem 3.1, and worked directly with his Theorem 2.1.
10. See, e.g., van der Vaart (1991, Sects. 3–6).
11. In Appendix E, we provide a counter example to show that in general, regular estimation is impossible for the linear functionals $E\left[\ell_0(Y_2)h(Y_2)\right]$ when ℓ_0 is unkown and has to be estimated.
12. Van der Vaart (1991, p. 190) discusses one such example, e.g., $\ell_0(z) = I\{z \leq z_0\}$ where z_0 is a fixed point and $I\{\cdot\}$ denotes the indicator function.
13. We thank an anonymous referee, who provided us with an alternative argument that does not rely on sufficiency or normality: Let $T(g) = E\left[g(X_1, X_2)|Z\right]$. If T is compact then it cannot be surjective by Banach's open mapping theorem. It follows that there is always an ℓ_0 not in the range of T.
14. We also developed a cubic example where the quadratic term converges to zero, yet the nonregularity still arises. This argument is available upon request.

REFERENCES

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

Ai, C. & X. Chen (2007) Estimation of possibly misspecified semiparametric conditional moment restriction models with different conditioning variables. *Journal of Econometrics* 141, 5–43.

Ai, C. & X. Chen (2012) The semiparametric efficiency bound for models of sequential moment restrictions containing unknown functions. *Journal of Econometrics* 170, 442–457.

Bierens, H. (2013) The Hilbert space theoretical foundation of semi-nonparametric modeling. In A. Ullah, J. Racine, & L. Su (eds.), *Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, pp. 3–37. Oxford University Press.

Carrasco, M., J.-P. Florens, & E. Renault (2007) Linear inverse problems in structural econometrics. In J.J. Heckman & E.E. Leamer (eds.), *Handbook of Econometrics* 6B, pp. 5633–5751. Elsevier.

Chamberlain, G. (1986) Asymptotic efficiency in semi-parametric models with censoring. *Journal of Econometrics* 32, 189–218.

Chen, X. & M. Reiss (2011) On rate optimality for ill-posed inverse problems in econometrics. *Econometric Theory* 27, 497–521.

Florens, J.-P., J. Johannes, & S. van Bellegem (2011) Identification and estimation by penalization in nonparametric instrumental regression. *Econometric Theory* 27, 472–496.

d'Haultfoeuille, X. (2011) On the completeness condition in nonparametric instrumental problems. *Econometric Theory* 27, 460–471.

Hahn, J. (1994) The efficiency bound of the mixed proportional hazard model. *Review of Economic Studies* 61, 607–629.

Hall, P. & J.L. Horowitz (2005) Nonparametric methods for inference in the presence of instrumental variables. *Annals of Statistics* 33, 2904–2929.

Horowitz, J. (2011) Applied nonparametric instrumental variables estimation. *Econometrica* 79, 347–394.

Newey, W.K. (1990) Semiparametric efficiency bounds. *Journal of Applied Econometrics* 5, 99–135.

Newey, W.K. (1994) The asymptotic variance of semiparametric estimators. *Econometrica* 62, 1349–1382.

Newey, W.P. & J.L. Powell (2003) Instrumental variables for nonparametric models. *Econometrica* 71, 1565–1578.

Johnston, W. (2014) The weighted Hermite polynomials form a basis for $L_2(R)$. *American Mathematical Monthly* 121, 249–253.

https://doi.org/10.1017/S0266466617000093 Published online by Cambridge University Press

Schennach, S.M. (2004) Estimation of nonlinear models with measurement error. *Econometrica* 72, 33–75.
Severini, T.A. & G. Tripathi (2012) Efficiency bounds for estimating linear functionals of nonparametric regression models with endogenous regressors. *Journal of Econometrics* 170, 491–498.
Van der Vaart, A. (1991) On differentiable functionals. *Annals of Statistics* 19, 178–204.
Van der Vaart, A. (2002) Semiparametric statistics. In P. Bernard (ed.), *Lectures on Probability Theory and Statistics* (*Ecole d'Eté de Probabilités de Saint-Flour XXIX - 1999*), pp. 331–457. Springer.

APPENDIX A: Proofs for Section 2

We examine Ai and Chen (2007) for the model (2) and (3). We assume that the moment conditions are correctly specified. This implies that the (weak) metric $\|\cdot\|$ on the top of page 16 of Ai and Chen (2007) is the same as the one proposed in Ai and Chen (2003). We follow Ai and Chen's (2007) notational convention as much as possible, except that they denote the true value of the parameter by the $*$ subscript, but we use the 0 subscript instead.

We assume that Assumptions 3.1–3.7 on page 14 and/or Assumptions 3.2–3.8 on page 16 of Ai and Chen (2007) are all satisfied. For simplicity of analysis, we will also make the following assumptions.

Assumption A.1. (i) $\psi(t)$ is three times continuously differentiable; (ii) for $\varkappa(t) = d\psi(t)/dt$,

$$\left.\frac{d\mathbb{E}\left[\psi\left(h_0(Y_2)+\tau w(Y_2)\right)\right]}{d\tau}\right|_{\tau=0} = \mathbb{E}\left[\varkappa\left(h_0(Y_2)\right)w(Y_2)\right];$$

(iii) there exists a unique $\xi(Y_2)$ with $\mathbb{E}[(\xi(Y_2))^2] < \infty$ such that $\mathbb{E}\left[\mathbb{E}\left[\xi(Y_2)|X\right]\middle| Y_2\right] = \varkappa\left(h_0(Y_2)\right)$.

As we will show below, Assumption A.1 ensures that the Riesz representor (Ai and Chen, 2007, Assumption 4.1) exists.

A.1. Ai and Chen's (2007, Assumption 4.1)

Let $\alpha = (h, \theta)$ and $\alpha_0 = (h_0, \theta_0)$. As in Ai and Chen (2007), we write

$$m_1(X,\alpha) \equiv \mathbb{E}\left[Y_1 - h(Y_2)|X\right],$$
$$m_2(\alpha) \equiv \mathbb{E}\left[\psi\left(h(Y_2)\right) - \theta\right],$$
$$m(X,\alpha) \equiv \begin{bmatrix} m_1(X,\alpha) \\ m_2(\alpha) \end{bmatrix}.$$

Under our assumption of correct specification, we have

$$m_1(X,\alpha_0) = 0,$$
$$m_2(\alpha_0) = 0,$$

and therefore, we can ignore the component that depends on the second derivative in the minimization problem in the middle of page 18 in Ai and Chen (2007); in the displayed

https://doi.org/10.1017/S0266466617000093 Published online by Cambridge University Press

equation that defines w^* in the middle of page 18, the second derivative term is multiplied by $m(X, \alpha_*)$ (in their notation), which is equal to zero.

By Assumption A.1(ii), we have

$$\begin{bmatrix} \frac{\partial m_1(X,\alpha_0)}{\partial \theta} \\ \frac{\partial m_2(\alpha_0)}{\partial \theta} \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}, \quad \begin{bmatrix} \frac{dm_1(X,\alpha_0)}{dh}[w] \\ \frac{dm_2(\alpha_0)}{dh}[w] \end{bmatrix} = \begin{bmatrix} -\mathbb{E}\left[w(Y_2)|X\right] \\ \mathbb{E}\left[\varkappa(h_0(Y_2))\, w(Y_2)\right] \end{bmatrix}$$

and hence

$$\frac{\partial m(X,\alpha_0)}{\partial \theta} - \frac{dm(X,\alpha_0)}{dh}[w] = \begin{bmatrix} \mathbb{E}\left[w(Y_2)|X\right] \\ -1 - \mathbb{E}\left[\varkappa(h_0(Y_2))\, w(Y_2)\right] \end{bmatrix}.$$

It follows that the w^* is computed as a solution to the following minimization problem:

$$\begin{aligned} &\min_w \mathbb{E}\left[\left(\frac{\partial m(X,\alpha_0)}{\partial \theta} - \frac{dm(X,\alpha_0)}{dh}[w]\right)'\left(\frac{\partial m(X,\alpha_0)}{\partial \theta} - \frac{dm(X,\alpha_0)}{dh}[w]\right)\right] \\ &\quad = \min_w \mathbb{E}\left[\left(-1 - \mathbb{E}\left[\varkappa(h_0(Y_2))\, w(Y_2)\right]\right)^2 + \left(\mathbb{E}\left[w(Y_2)|X\right]\right)^2\right]. \end{aligned}$$

Similarly, the metric defined in Ai and Chen (2007, p. 16) depends only on the first derivative term, and is equal to

$$\begin{aligned} \|\alpha\|^2 &= \mathbb{E}\left[\left\|\frac{dm(X,\alpha_0)}{d\alpha}[\alpha]\right\|_E^2\right] \\ &= \mathbb{E}\left[\left(\frac{\partial m(X,\alpha_0)}{\partial \theta}\theta + \frac{dm(X,\alpha_0)}{dh}[h]\right)'\left(\frac{\partial m(X,\alpha_0)}{\partial \theta}\theta + \frac{dm(X,\alpha_0)}{dh}[h]\right)\right] \\ &= \mathbb{E}\left[\left(\theta - \mathbb{E}\left[\varkappa(h_0(Y_2))\, h(Y_2)\right]\right)^2 + \left(\mathbb{E}\left[h(Y_2)|X\right]\right)^2\right] \end{aligned}$$

where $\|\cdot\|_E$ denotes the standard Euclidean norm (Ai and Chen, 2007, p. 14). Therefore, the inner product induced by the implicit norm $\|\alpha\|$ here is given by

$$\langle \alpha_1, \alpha_2 \rangle = \mathbb{E}\left[\begin{array}{c} \mathbb{E}\left[h_1(Y_2)|X\right]\mathbb{E}\left[h_2(Y_2)|X\right] \\ + \left(\theta_1 - \mathbb{E}\left[\varkappa(h_0(Y_2))\, h_1(Y_2)\right]\right)\left(\theta_2 - \mathbb{E}\left[\varkappa(h_0(Y_2))\, h_2(Y_2)\right]\right) \end{array}\right]. \tag{A.1}$$

A.2. Calculation of Riesz Representation

We derive the Riesz representor v^* by the definition given above Assumption 4.1 in Ai and Chen (2007). We start by characterizing the w^*.

LEMMA A.1. *Under Assumption A.1,*

$$w^*(Y_2) = -\frac{\xi(Y_2)}{1 + \mathbb{E}[\omega(X)^2]} \tag{A.2}$$

where $\omega(X) = \mathbb{E}\left[\xi(Y_2)|X\right]$.

Proof. Recall that w^* is the solution to

$$\begin{aligned} &\min_w \mathbb{E}\left[\left(\mathbb{E}\left[w(Y_2)|X\right]\right)^2 + \left(-1 - \mathbb{E}\left[\varkappa(h_0(Y_2))\, w(Y_2)\right]\right)^2\right] \\ &\quad = \min_w \left[\mathbb{E}\left(\mathbb{E}\left[w(Y_2)|X\right]\right)^2 + \left(1 + \mathbb{E}\left[\omega(X)\, w(Y_2)\right]\right)^2\right]. \end{aligned} \tag{A.3}$$

https://doi.org/10.1017/S0266466617000093 Published online by Cambridge University Press

Let $g(X) = \mathbb{E}\left[w(Y_2)|X\right]$. We can write

$$\begin{aligned}
&\mathbb{E}\left[\left(\mathbb{E}\left[w(Y_2)|X\right]\right)^2 + \left(1+\mathbb{E}\left[\omega(X)w(Y_2)\right]\right)^2\right] \\
&\quad = \mathbb{E}\left[g(X)^2\right] + 2\mathbb{E}[\omega(X)g(X)] + (\mathbb{E}[\omega(X)g(X)])^2 + 1 \\
&\quad = \mathbb{E}\left[\left|g(X) + \frac{\omega(X)}{1+\mathbb{E}[\omega(X)^2]}\right|^2\right] \\
&\qquad + \left|\mathbb{E}\left[\left(g(X) + \frac{\omega(X)}{1+\mathbb{E}[\omega(X)^2]}\right)\omega(X)\right]\right|^2 \\
&\qquad + 1 - \frac{\mathbb{E}[\omega(X)^2]}{(1+\mathbb{E}[\omega(X)^2])^2} - \frac{(\mathbb{E}[\omega(X)^2])^2}{(1+\mathbb{E}[\omega(X)^2])^2} \\
&\quad \geq \frac{1}{1+\mathbb{E}[\omega(X)^2]}.
\end{aligned}$$

The above inequality becomes equality if $g(X) = -\omega(X)/(1+\mathbb{E}[\omega(X)^2])$. This means that the minimizer of the problem in (A.3) is $w^*(Y_2)$ defined in (A.2). ■

LEMMA A.2. *Let* $D_{w^*} = \frac{\partial m(X,\alpha_0)}{\partial \theta} - \frac{dm(X,\alpha_0)}{dh}\left[w^*\right]$. *Then under Assumption A.1,*

$$D_{w^*} = \begin{bmatrix} -\frac{\omega(X)}{1+\mathbb{E}\left[\omega(X)^2\right]} \\ -\frac{1}{1+\mathbb{E}\left[\omega(X)^2\right]} \end{bmatrix}.$$

Proof.

$$\begin{aligned}
D_{w^*} &= \begin{bmatrix} \frac{dm_1(X,h_0,\theta_0)}{d\theta} - \frac{dm_1(X,h_0,\theta_0)}{dh}\left[w^*\right] \\ \frac{dm_2(h_0,\theta_0)}{d\theta} - \frac{dm_2(h_0,\theta_0)}{dh}\left[w^*\right] \end{bmatrix} = \begin{bmatrix} \mathbb{E}\left[w^*(Y_2)|X\right] \\ -1 - \mathbb{E}\left[\varkappa(h_0(Y_2))w^*(Y_2)\right] \end{bmatrix} \\
&= \begin{bmatrix} \mathbb{E}\left[w^*(Y_2)|X\right] \\ -1 - \mathbb{E}\left[w^*(Y_2)\omega(X)\right] \end{bmatrix} = \begin{bmatrix} -\frac{\omega(X)}{1+\mathbb{E}\left[\omega(X)^2\right]} \\ -1 + \frac{\mathbb{E}\left[\omega(X)^2\right]}{1+\mathbb{E}\left[\omega(X)^2\right]} \end{bmatrix} = \begin{bmatrix} -\frac{\omega(X)}{1+\mathbb{E}\left[\omega(X)^2\right]} \\ -\frac{1}{1+\mathbb{E}\left[\omega(X)^2\right]} \end{bmatrix}
\end{aligned}$$

where the fourth equality is based on the last lemma. ■

Under our correct specification assumption, the V_{w^*} term defined on page 19 of Ai and Chen (2007) is irrelevant for the calculation of v_θ^*, and we immediately have the following lemma.

LEMMA A.3. *Under Assumption A.1,*

$$v_\theta^* = \left(\mathbb{E}\left[D_{w^*}' D_{w^*}\right]\right)^{-1} = 1 + \mathbb{E}\left[\omega(X)^2\right].$$

According to the definition of v^* right above Assumption 4.1 in Ai and Chen (2007), we have

$$v^* = \begin{bmatrix} v_\theta^* \\ v_h^* \end{bmatrix} = \begin{bmatrix} v_\theta^* \\ -w^* \times v_\theta^* \end{bmatrix} = \begin{bmatrix} 1+\mathbb{E}\left[\omega(X)^2\right] \\ \xi(Y_2) \end{bmatrix}. \tag{A.4}$$

https://doi.org/10.1017/S0266466617000093 Published online by Cambridge University Press

In order to verify that the v^* above is indeed correctly calculated, we can use the inner product in (A.1) and verify that

$$\langle v^*, \alpha - \alpha_0 \rangle = \mathbb{E}\left[\begin{array}{c} \mathbb{E}[\xi(Y_2)|X]\mathbb{E}[(h-h_0)(Y_2)|X] \\ +\left(-\left(1+\mathbb{E}\left[\omega(X)^2\right]\right)+\mathbb{E}[\omega(X)\xi(Y_2)]\right)(-(\theta-\theta_0)+\mathbb{E}[\omega(X)(h-h_0)(Y_2)]) \end{array}\right]$$
$$= \mathbb{E}[\omega(X)\mathbb{E}[(h-h_0)(Y_2)|X]] - (-(\theta-\theta_0)+\mathbb{E}[\omega(X)(h-h_0)(Y_2)]) = \theta-\theta_0. \quad \textbf{(A.5)}$$

A.3. Implication of Assumption 4.4

Assumption 4.4 is only used once in Ai and Chen (2007), and it is to provide a justification of the second last displayed equation on page 39 there. We will show that their Assumption 4.4 is equivalent to

$$\begin{aligned} &\mathbb{E}\left[\frac{d^2\varkappa(h_0+th)}{dt^2}v^*_{n,h}(Y_2)\right]\mathbb{E}\left[\psi(h_0+th)-(\theta_0+t\theta)\right] \\ &\quad +2\mathbb{E}\left[\frac{d\varkappa(h_0+th)}{dt}v^*_{n,h}(Y_2)\right]\mathbb{E}\left[\frac{d\psi(h_0+th)}{dt}-\theta\right] \\ &\quad +\left(-v^*_{n,\theta}+\mathbb{E}\left[\varkappa(h_0+th)v^*_{n,h}(Y_2)\right]\right)\mathbb{E}\left[\frac{d^2\psi(h_0+th)}{dt^2}\right] \\ &= o\left(n^{-\frac{1}{2}}\right) \end{aligned} \quad \textbf{(A.6)}$$

uniformly over $t \in [0,1]$ and (θ, h) such that $\theta = o\left(n^{-\frac{1}{4}}\right)$, $\|h\|_s = o(1)$ and $\mathbb{E}[(\mathbb{E}[h(Y_2)|X])^2] = o\left(n^{-\frac{1}{2}}\right)$, where $\|\cdot\|_s$ is the strong metric used in Ai and Chen (2007). Here, $v^*_n = (v^*_{n,\theta}, v^*_{n,h}(Y_2))'$ are such that $v^*_{n,\theta} = v^*_\theta + O\left(n^{-\frac{1}{4}}\right)$, and $\mathbb{E}[(\mathbb{E}[v^*_{n,h}(Y_2) - v^*_h(Y_2)|X])^2] = O\left(n^{-\frac{1}{2}}\right)$, where v^*_θ and v^*_h are defined in (A.4).

Equation (A.6) may appear messy and complicated, but it is straightforward to show that it is violated in general. (The requirement (A.6) is trivially satisfied if $\psi(h(Y_2)) = h(Y_2)$, but it is not an interesting problem. After all, we can see that $\theta_0 = \mathbb{E}\left\{\mathbb{E}\left[h_0(Y_2)|X\right]\right\} = \mathbb{E}\left\{\mathbb{E}\left[Y_1|X\right]\right\} = \mathbb{E}\left[Y_1\right]$, so θ_0 can be estimated without even addressing the ill-posed inverse problem.) The derivatives $\mathbb{E}[d^j\varkappa(h_0+th)/dt^j]$ and $\mathbb{E}[d^j\psi(h_0+th)/dt^j]$ (for $j=1,2$) often fail to converge to zero at a reasonably fast rate to ensure that (A.6) holds even when $\mathbb{E}[(\mathbb{E}[h(Y_2)|X])^2] = o\left(n^{-\frac{1}{2}}\right)$ and $\|h\|_s = o(1)$, which can be seen by considering for example, $\psi(h(Y_2)) = \mathbb{E}\left[(h(Y_2))^m\right]$ for any $m \neq 0$ or 1. The problem is that even though $\mathbb{E}[(\mathbb{E}[h(Y_2)|X])^2] = o(n^{-\frac{1}{2}})$ and $\|h\|_s = o(1)$, it does not necessarily follow that $\mathbb{E}[h(Y_2)^2] = o(n^{-\frac{1}{2}})$; after all, $\mathbb{E}[h(Y_2)^2] = \mathbb{E}[(\mathbb{E}[h(Y_2)|X])^2] + \mathbb{E}[\mathrm{Var}(h(Y_2)|X)]$ and $\mathbb{E}[\mathrm{Var}(h(Y_2)|X)]$ may converge to zero very slowly.

A.4. Proof of (A.6)

We now prove (A.6). Let $v^*_n = (v^*_{n,\theta}, v^*_{n,h})'$ such that $\left\|v^*_n - v^*\right\| = O(n^{-\frac{1}{4}})$ as discussed in Assumption 4.2' of Ai and Chen (2007). We note that

https://doi.org/10.1017/S0266466617000093 Published online by Cambridge University Press

$$\left\|v_n^*-v^*\right\|^2=\mathbb{E}\left[\left(-\left(v_{n,\theta}^*-\left(1+\mathbb{E}\left[\omega(X)^2\right]\right)\right)+\mathbb{E}\left[\omega(X)\mathbb{E}\left[v_{n,h}^*(Y_2)-\xi(Y_2)\middle|X\right]\right]\right)^2\right]$$
$$+\mathbb{E}\left[\left(\mathbb{E}\left[v_{n,h}^*(Y_2)-\xi(Y_2)\middle|X\right]\right)^2\right]$$

so it is equivalent to the assumption that

$$v_{n,\theta}^*=1+\mathbb{E}[\omega(X)^2]+O\left(n^{-\frac{1}{4}}\right) \text{ and } \mathbb{E}\left[\left(\mathbb{E}\left[v_{n,h}^*(Y_2)-\xi(Y_2)\middle|X\right]\right)^2\right]=O\left(n^{-\frac{1}{2}}\right).$$

Because

$$\begin{bmatrix}\frac{\partial m_1(X,\alpha)}{\partial\theta}\\ \frac{\partial m_2(\alpha)}{\partial\theta}\end{bmatrix}=\begin{bmatrix}0\\-1\end{bmatrix},\quad \begin{bmatrix}\frac{dm_1(X,\alpha)}{dh}[w]\\ \frac{dm_2(\alpha)}{dh}[w]\end{bmatrix}=\begin{bmatrix}-\mathbb{E}\left[w(Y_2)\middle|X\right]\\ \mathbb{E}\left[\varkappa(h(Y_2))w(Y_2)\right]\end{bmatrix},$$

we have

$$\left(\frac{dm(X,\alpha)}{d\alpha}\left[v_n^*\right]\right)'m(X,\alpha)=-\mathbb{E}\left[v_{n,h}^*(Y_2)\middle|X\right]\mathbb{E}\left[Y_1-h(Y_2)\middle|X\right]$$
$$+\left(-v_{n,\theta}^*+\mathbb{E}\left[\varkappa(h)v_{n,h}^*(Y_2)\right]\right)\mathbb{E}\left[\psi(h)-\theta\right]$$

and hence

$$\mathbb{E}\left[\left(\frac{dm(X,\alpha_0+t\alpha)}{d\alpha}\left[v_n^*\right]\right)'m(X,\alpha_0+t\alpha)\right]$$
$$=-\mathbb{E}\left\{\mathbb{E}\left[v_{n,h}^*(Y_2)\middle|X\right]\mathbb{E}\left[Y_1-h_0(Y_2)-th(Y_2)\middle|X\right]\right\}$$
$$+\left(-v_{n,\theta}^*+\mathbb{E}\left[\varkappa(h_0+th)v_{n,h}^*(Y_2)\right]\right)\mathbb{E}\left[\psi(h_0+th)-(\theta_0+t\theta)\right].$$

Assumption 4.4 of Ai and Chen (2007) requires that

$$\sup_{0\le t\le 1}\frac{d^2}{dt^2}\mathbb{E}\left[\left(\frac{dm(X,\alpha_0+t\alpha)}{d\alpha}\left[v_n^*\right]\right)'m(X,\alpha_0+t\alpha)\right]=o\left(n^{-\frac{1}{2}}\right)$$

uniformly over (θ,h) such that $\alpha\in\mathcal{N}_{on}$, i.e.,

$$\theta=o\left(n^{-\frac{1}{4}}\right),\ \mathbb{E}[(\mathbb{E}[h(Y_2)|X])^2]=o\left(n^{-\frac{1}{2}}\right) \text{ and } \|h\|_s=o(1).$$

In other words, Assumption 4.4 requires that

$$\frac{d^2}{dt^2}\mathbb{E}\left[\left(\frac{dm(X,\alpha_0+t\alpha)}{d\alpha}\left[v_n^*\right]\right)'m(X,\alpha_0+t\alpha)\right]$$
$$=\mathbb{E}\left[\frac{d^2\varkappa(h_0+th)}{dt^2}v_{n,h}^*(Y_2)\right]\mathbb{E}\left[\psi(h_0+th)-(\theta_0+t\theta)\right]$$
$$+2\mathbb{E}\left[\frac{d\varkappa(h_0+th)}{dt}v_{n,h}^*(Y_2)\right]\mathbb{E}\left[\frac{d\psi(h_0+th)}{dt}-\theta\right]$$
$$+\left(-v_{n,\theta}^*+\mathbb{E}\left[\varkappa(h_0(Y_2)+th(Y_2))v_{n,h}^*(Y_2)\right]\right)\mathbb{E}\left[\frac{d^2\psi(h_0+th)}{dt^2}\right]$$
$$=o\left(n^{-\frac{1}{2}}\right), \tag{A.7}$$

uniformly over $t\in[0,1]$ and over $\alpha\in\mathcal{N}_{on}$.

https://doi.org/10.1017/S0266466617000093 Published online by Cambridge University Press

A.5. Special Case

For the special case $\psi(h_0(Y_2)) = \frac{1}{2}(h_0(Y_2))^2$, we have

$$\begin{aligned}\varkappa(h(Y_2)) &= h(Y_2),\\ \mathbb{E}\left[\mathbb{E}\left[\xi(Y_2)\,|X\right]|Y_2\right] &= h_0(Y_2),\\ \omega(X) &= \mathbb{E}\left[\xi(Y_2)\,|X\right],\\ v_h^*(Y_2) &= \xi(Y_2),\\ v_\theta^* &= 1+\mathbb{E}[(\omega(X)^2)],\end{aligned}$$

and the requirement (A.7) becomes

$$\begin{aligned}&2\mathbb{E}\left[h(Y_2)v_{n,h}^*(Y_2)\right]\mathbb{E}\left[(h_0(Y_2)+th(Y_2))h(Y_2)-\theta\right]\\ &\quad+\left(-v_{n,\theta}^*+\mathbb{E}\left[(h_0(Y_2)+th(Y_2))v_{n,h}^*(Y_2)\right]\right)\mathbb{E}\left[(h(Y_2))^2\right]=o\left(n^{-\frac{1}{2}}\right)\end{aligned}$$

uniformly over $t\in[0,1]$ and over (θ,h) such that $\theta = o(n^{-\frac{1}{4}})$, $\|h\|_s = o(1)$, $\mathbb{E}[(\mathbb{E}[h(Y_2)\,|X])^2] = o\left(n^{-\frac{1}{2}}\right)$, $v_{n,\theta}^* = v_\theta^* + O\left(n^{-\frac{1}{4}}\right)$ and $\mathbb{E}[(\mathbb{E}[v_{n,h}^*(Y_2) - v_h^*(Y_2)\,|X])^2] = O\left(n^{-\frac{1}{2}}\right)$, where $\|\cdot\|_s$ is the strong norm (e.g., the sup-norm or the L_2-norm) used in Ai and Chen (2007), $v_h^*(Y_2)$ is such that $\mathbb{E}\left[\mathbb{E}\left[v_h^*(Y_2)\,|X\right]|Y_2\right] = h_0(Y_2)$ and $v_\theta^* = 1+\mathbb{E}[(\mathbb{E}\left[v_h^*(Y_2)\,|X\right])^2]$. An obvious problem of this requirement is that even though $\mathbb{E}[(\mathbb{E}[h(Y_2)\,|X])^2] = o\left(n^{-\frac{1}{2}}\right)$ and $\|h\|_s = o(1)$, it does not necessarily follow that $\mathbb{E}[h(Y_2)v_{n,h}^*(Y_2)]$ and $\mathbb{E}[(h(Y_2))^2]$ will converge to zero sufficiently fast such that it holds.

For simplicity, we consider the case $v_n^* = v^*$, i.e., we consider the special case of Assumption 4.2' of Ai and Chen (2007). For $t=0$, we obtain

$$2\mathbb{E}[h(Y_2)\xi(Y_2)]\mathbb{E}[h_0(Y_2)h(Y_2)-\theta]+\left(-v_\theta^*+\mathbb{E}[h_0(Y_2)\xi(Y_2)]\right)\mathbb{E}\left[(h(Y_2))^2\right]=o\left(n^{-\frac{1}{2}}\right). \tag{A.8}$$

Because

$$\begin{aligned}&\mathbb{E}[h_0(Y_2)\xi(Y_2)] = \mathbb{E}\left[\mathbb{E}\left[\omega(X)\,|Y_2\right]\xi(Y_2)\right] = \mathbb{E}[\omega(X)\xi(Y_2)] = \mathbb{E}\left[\omega(X)\mathbb{E}\left[\xi(Y_2)\,|X\right]\right] = \mathbb{E}[(\omega(X))^2],\\ &\mathbb{E}[h_0(Y_2)h(Y_2)] = \mathbb{E}\left[\mathbb{E}\left[\omega(X)\,|Y_2\right]h(Y_2)\right] = \mathbb{E}[\omega(X)h(Y_2)] = \mathbb{E}\left[\omega(X)\mathbb{E}\left[h(Y_2)\,|X\right]\right],\end{aligned}$$

we may rewrite (A.8) as

$$\begin{aligned}&2\mathbb{E}\left[h(Y_2)\xi(Y_2)\right]\left(\mathbb{E}\left[\omega(X)\mathbb{E}\left[h(Y_2)\,|X\right]\right]-\theta\right)\\ &\quad+\left(-(1+\mathbb{E}[(\omega(X))^2])+\mathbb{E}[(\omega(X))^2]\right)\mathbb{E}\left[(h(Y_2))^2\right]\\ &=2\mathbb{E}\left[h(Y_2)\xi(Y_2)\right]\left(\mathbb{E}\left[\omega(X)\mathbb{E}\left[h(Y_2)\,|X\right]\right]-\theta\right)-\mathbb{E}\left[(h(Y_2))^2\right]=o\left(n^{-\frac{1}{2}}\right).\end{aligned} \tag{A.9}$$

Because $\mathbb{E}[(\omega(X))^2] < \infty$ and $\mathbb{E}[(\mathbb{E}[h(Y_2)\,|X])^2] = o(n^{-\frac{1}{2}})$, we have

$$\left|\mathbb{E}\left[\omega(X)\mathbb{E}\left[h(Y_2)\,|X\right]\right]\right| \le \sqrt{\mathbb{E}\left[(\omega(X))^2\right]}\sqrt{\mathbb{E}\left[\left(\mathbb{E}\left[h(Y_2)\,|X\right]\right)^2\right]} = o\left(n^{-\frac{1}{4}}\right). \tag{A.10}$$

https://doi.org/10.1017/S0266466617000093 Published online by Cambridge University Press

Moreover,

$$\begin{aligned}\mathbb{E}\left[h(Y_2)\xi(Y_2)\right] &= \mathbb{E}\left[\mathbb{E}\left[h(Y_2)\middle|X\right]\xi(Y_2)\right]+\mathbb{E}\left[\left(h(Y_2)-\mathbb{E}\left[h(Y_2)\middle|X\right]\right)\xi(Y_2)\right]\\ &= \mathbb{E}\left[\mathbb{E}\left[h(Y_2)\middle|X\right]\mathbb{E}\left[\xi(Y_2)\middle|X\right]\right]+\mathbb{E}\left[\left(h(Y_2)-\mathbb{E}\left[h(Y_2)\middle|X\right]\right)\xi(Y_2)\right]\\ &= o\left(n^{-\frac{1}{4}}\right)+\mathbb{E}\left[\left(h(Y_2)-\mathbb{E}\left[h(Y_2)\middle|X\right]\right)\xi(Y_2)\right], \qquad \textbf{(A.11)}\end{aligned}$$

where the last equality is by (A.10). By $\theta = o\left(n^{-\frac{1}{4}}\right)$ and $\mathbb{E}[(\mathbb{E}[h(Y_2)|X])^2] = o\left(n^{-\frac{1}{2}}\right)$, and the results in (A.9) and (A.11), we can further rewrite (A.8) as

$$\mathbb{E}\left[\left(h(Y_2)-\mathbb{E}\left[h(Y_2)\middle|X\right]\right)\xi(Y_2)\right]o(n^{-1/4})+\mathbb{E}\left[\left(h(Y_2)-\mathbb{E}\left[h(Y_2)\middle|X\right]\right)^2\right] = o\left(n^{-\frac{1}{2}}\right)$$

which suggests that the requirement (A.7) implies $\mathbb{E}\left[\left(h(Y_2)-\mathbb{E}\left[h(Y_2)\middle|X\right]\right)^2\right] = o\left(n^{-\frac{1}{2}}\right)$. (This can be verified by considering a special case where there exists some ϖ with $\mathbb{E}\left[(\varpi(X))^2\right] < \infty$ such that $\mathbb{E}\left[\varpi(X)\middle|Y_2\right] = \xi(Y_2)$. It then follows that $\mathbb{E}\left[\left(h(Y_2)-\mathbb{E}\left[h(Y_2)\middle|X\right]\right)\xi(Y_2)\right] = \mathbb{E}\left[\left(h(Y_2)-\mathbb{E}\left[h(Y_2)\middle|X\right]\right)\mathbb{E}\left[\varpi(X)\middle|Y_2\right]\right] = 0$, which simplifies the requirement (A.8).)

APPENDIX B: Mean Square Differentiability

We will show that under some regularity conditions, $s = 0$ is the score of the path F_t constructed in (6) if r_t satisfies $t^{-1}(r_t - r_0) \to 0$ in quadratic mean. Let $g_0(u, v|x)$ denote the conditional density of (u, v) given X. We use the following regularity condition:

Assumption B.1. (i) The joint density $\sqrt{g_0(u, v, x)}$ is continuously differentiable in u for all u, v and x; (ii) for any x and any $\mathbf{r} \in \mathbb{R}$, we have

$$I(\mathbf{r}|x) = \int \left(\frac{\frac{\partial}{\partial u} g_0(u-\mathbf{r}, v|x)}{g_0(u-\mathbf{r}, v|x)}\right)^2 g_0(u-\mathbf{r}, v|x) < M,$$

where M is a fixed finite constant and integration is taken with respect to (u, v) under some dominating measure.

Assumption B.1 is similar to the conditions in Lemma 1.8 of van der Vaart (2002). It is clear that $I(\mathbf{r}|x)$ denotes the information of the (conditional) location model $g_0(u-\mathbf{r}, v|x)$. Hence Assumption B.1(ii) requires that the information of the location model is bounded from above. As a simple illustration, we consider the conditional distribution of (u, v) given X is standard normal with correlation. Then

$$g_0(u-\mathbf{r}, v|x) \propto \exp\left(-\frac{1}{2}\begin{bmatrix} u-\mathbf{r} & v\end{bmatrix}\begin{bmatrix}1 & \delta\\ \delta & 1\end{bmatrix}^{-1}\begin{bmatrix} u-\mathbf{r}\\ v\end{bmatrix}\right)$$

which implies that the score for calculation of the information is equal to $-\frac{1}{\delta^2-1}(\mathbf{r}-u+v\delta)$. Hence the information of this (conditional) Gaussian location model is $\frac{1}{1-\delta^2}$ which is bounded uniformly over $\mathbf{r}$ and x.

LEMMA B.1. *Suppose that Assumption B.1 holds. Then $s = 0$ is the score of the path F_t constructed in (6) where r_t satisfies $t^{-1}(r_t - r_0) \to 0$ in quadratic mean.*

https://doi.org/10.1017/S0266466617000093 Published online by Cambridge University Press

Proof. By definition, it is sufficient to show that

$$\int \left(\frac{\sqrt{dF_t} - \sqrt{dF}}{t} \right)^2 \to 0 \text{ as } t \to 0. \tag{B.1}$$

Let $\widetilde{r}_t = r_t - r_0$ and $g_{x,0}(x)$ denote the marginal density of X. By Assumption B.1(i),

$$\begin{aligned} \sqrt{f_t} - \sqrt{f_0} &= \left(\sqrt{g_0(u - \widetilde{r}_t(x), v|x)} - \sqrt{g_0(u, v|x)} \right) \sqrt{g_{x,0}(x)} \\ &= \frac{1}{2} \frac{\frac{\partial}{\partial u} g_0\left(u - \widetilde{r}_t^*(x), v|x\right)}{\sqrt{g_0\left(u - \widetilde{r}_t^*(x), v|x\right)}} \widetilde{r}_t(x) \sqrt{g_{x,0}(x)}, \end{aligned} \tag{B.2}$$

where $\widetilde{r}_t^*$ is between $\widetilde{r}_t$ and 0. Using (B.2), we have

$$\begin{aligned} \int \left(\frac{\sqrt{dF_t} - \sqrt{dF}}{t} \right)^2 &= \frac{1}{4} \int \int \int \left(\frac{\frac{\partial}{\partial u} g_0\left(u - \widetilde{r}_t^*(x), v|x\right)}{\sqrt{g_0\left(u - \widetilde{r}_t^*(x), v|x\right)}} \right)^2 \left(\frac{\widetilde{r}_t(x)}{t} \right)^2 g_{x,0}(x) du dv dx \\ &= \frac{1}{4} \int \left[\int \int \left(\frac{\frac{\partial}{\partial u} g_0\left(u - \widetilde{r}_t^*(x), v|x\right)}{\sqrt{g_0\left(u - \widetilde{r}_t^*(x), v|x\right)}} \right)^2 du dv \right] \left(\frac{\widetilde{r}_t(x)}{t} \right)^2 g_{x,0}(x) dx \\ &= \frac{1}{4} \int I\left(\widetilde{r}_t^*|x\right) \left(\frac{\widetilde{r}_t(x)}{t} \right)^2 g_{x,0}(x) dx \\ &\le \frac{M}{4} \int \left(\frac{r_t(x) - r_0(x)}{t} \right)^2 g_{x,0}(x) \to 0 \text{ as } t \to 0, \end{aligned}$$

where the first inequality is by Assumption B.1(ii). ■

APPENDIX C: Proofs for Section 3

LEMMA C.1. *In Example 1, we have* $\mathbb{E}\left[n \left| r_{n^{-1/2}}(X) - r_0(X) \right|^2 \right] \to 0$ *as* $n \to \infty$.

Proof. By definition,

$$\mathbb{E}\left[n \left| r_{n^{-1/2}}(X) - r_0(X) \right|^2 \right] = \mathbb{E}\left[\left| \sum_{k=1}^{\infty} \phi_{n^{-1/2},k} \varphi_k(X) \right|^2 \right] = \sum_{k=1}^{\infty} \phi_{n^{-1/2},k}^2 \tag{C.1}$$

where the last equality holds because $\{\varphi_k\}_{k=1}^{\infty}$ is an orthonormal basis. For any real number a, we use $\lfloor a \rfloor$ to denote its integer part. Then

$$\sum_{k=1}^{\infty} \phi_{n^{-1/2},k}^2 = n^{-\frac{1}{4}} \sum_{k=1}^{\lfloor n^{\frac{3}{2a}} \rfloor} k^{-1-\frac{a}{2}} \le n^{-\frac{1}{4}} \sum_{k=1}^{\infty} k^{-1-\frac{a}{2}} \to 0, \text{ as } n \to \infty. \tag{C.2}$$

■

LEMMA C.2. *In Example 1, we have* $n^{\frac{1}{2}} \left(\theta_{n^{-1/2}} - \theta_0 \right) \to \frac{1}{a} \in (0, \infty)$ *as* $n \to \infty$.

https://doi.org/10.1017/S0266466617000093 Published online by Cambridge University Press

Proof. We see that

$$\sqrt{n}\left[h_{n^{-1/2}}(y_2)-h_0(y_2)\right]=\sum_{k=1}^{\infty}\phi_{n^{-1/2},k}\rho_k^{-\frac{1}{2}}\varphi_k(y_2),$$

where $\phi_{n^{-1/2},k}$ and ρ_k satisfy

$$n^{-\frac{1}{2}}\sum_{k=1}^{\infty}\phi^2_{n^{-1/2},k}\rho_k^{-1}=n^{-\frac{3}{4}}\sum_{k=1}^{\lfloor n^{\frac{3}{2a}}\rfloor}k^{-1+\frac{a}{2}}\to\int_0^1 x^{-1+\frac{a}{2}}dx=\frac{2}{a}\in(0,\infty). \tag{C.3}$$

(In (C.3), we used

$$\begin{aligned}n^{-\frac{3}{4}}\sum_{k=1}^{\lfloor n^{\frac{3}{2a}}\rfloor}k^{-1+\frac{a}{2}}&=n^{-\frac{3}{4}}\left(\lfloor n^{\frac{3}{2a}}\rfloor\right)^{\frac{a}{2}}\left(\frac{1}{\lfloor n^{\frac{3}{2a}}\rfloor}\sum_{k=1}^{\lfloor n^{\frac{3}{2a}}\rfloor}\left(\frac{k}{\lfloor n^{\frac{3}{2a}}\rfloor}\right)^{-1+\frac{a}{2}}\right)\\&=n^{-\frac{3}{4}}n^{\frac{3}{4}}\left(\int_0^1 x^{-1+\frac{a}{2}}dx+o(1)\right)\\&=\int_0^1 x^{-1+\frac{a}{2}}dx+o(1).)\end{aligned}$$

Moreover, by the Cauchy-Schwarz inequality,

$$\sum_{k=1}^{\infty}\phi_{n^{-1/2},k}\phi_k\rho_k^{-1}\le\sqrt{\sum_{k=1}^{\infty}\phi^2_{n^{-1/2},k}}\sqrt{\sum_{k=1}^{\infty}\phi_k^2\rho_k^{-2}}.$$

As long as $\sum_{k=1}^{\infty}\phi_k^2\rho_k^{-2}<\infty$, which is the case because $r_0=0$, (C.2) implies that

$$\sum_{k=1}^{\infty}\phi_{n^{-1/2},k}\phi_k\rho_k^{-1}\to 0,\text{ as }n\to\infty. \tag{C.4}$$

Given these conditions, we see that

$$\begin{aligned}n^{\frac{1}{2}}\left(\theta_{n^{-1/2}}-\theta_0\right)&=n^{\frac{1}{2}}\left\{\mathbb{E}\left[\frac{1}{2}h^2_{n^{-1/2}}(Y_2)\right]-\mathbb{E}\left[\frac{1}{2}h_0^2(Y_2)\right]\right\}\\&=\mathbb{E}\left[n^{\frac{1}{2}}\left(h_{n^{-1/2}}(Y_2)-h_0(Y_2)\right)h_0(Y_2)\right]+\frac{1}{2}\mathbb{E}\left[n^{\frac{1}{2}}(h_{n^{-1/2}}(Y_2)-h_0(Y_2))^2\right]\\&=\sum_{k=1}^{\infty}\phi_{n^{-1/2},k}\phi_k\rho_k^{-1}+\frac{1}{2\sqrt{n}}\sum_{k=1}^{\infty}\phi^2_{n^{-1/2},k}\rho_k^{-1}\to\frac{1}{a}>0,\text{ as }n\to\infty.\end{aligned}$$ ■

LEMMA C.3. *In Example 2,* $\mathbb{E}\left[\left|n^{\frac{1}{2}}\left[r_{n^{-1/2}}(X)-r_0(X)\right]\right|^2\right]\to 0$ *as* $n\to\infty$.

Proof.

$$\mathbb{E}\left[\left|n^{\frac{1}{2}}\left[r_{n^{-1/2}}(X)-r_0(X)\right]\right|^2\right]=\sum_{k=1}^{\infty}\phi^2_{n^{-1/2},k}\le n^{\frac{1}{2}}\rho^{n^{\frac{1}{2}}}\to 0,\text{ as }n\to\infty.$$ ■

LEMMA C.4. *In Example 2,* $n^{\frac{1}{2}}\left(\theta_{n^{-1/2}}-\theta_0\right)\to\frac{1}{2}>0$ *as* $n\to\infty$.

https://doi.org/10.1017/S0266466617000093 Published online by Cambridge University Press

Proof. We see that

$$\sqrt{n}\left[h_{n^{-1/2}}(y_2)-h_0(y_2)\right]=\sum_{k=0}^{\infty}\frac{\phi_{n^{-1/2},k}}{\rho^{k/2}}\varphi_k(y_2),$$

where $\phi_{n^{-1/2},k}$ and ρ satisfy

$$n^{-\frac{1}{2}}\sum_{k=0}^{\infty}\phi_{n^{-1/2},k}^2\rho^{-k}=1 \text{ for all } n. \tag{C.5}$$

Given these conditions, we see that

$$\begin{aligned}n^{\frac{1}{2}}\left(\theta_{n^{-1/2}}-\theta_0\right)&=n^{\frac{1}{2}}\left\{\mathbb{E}\left[\frac{1}{2}h_{n^{-1/2}}^2(Y_2)\right]-\mathbb{E}\left[\frac{1}{2}h_0^2(Y_2)\right]\right\}\\&=\mathbb{E}\left[n^{\frac{1}{2}}\left(h_{n^{-1/2}}(Y_2)-h_0(Y_2)\right)h_0(Y_2)\right]\\&\quad+\frac{1}{2}\mathbb{E}\left[n^{\frac{1}{2}}\left(h_{n^{-1/2}}(Y_2)-h_0(Y_2)\right)^2\right]\\&=\sum_{k=0}^{\infty}\frac{\phi_{n^{-1/2},k}\phi_k}{\rho^k}+\frac{n^{-\frac{1}{2}}}{2}\sum_{k=0}^{\infty}\frac{\phi_{n^{-1/2},k}^2}{\rho^k}\to\frac{1}{2}>0, \text{ as } n\to\infty,\end{aligned}$$

where $\sum_{k=0}^{\infty}\phi_{n^{-1/2},k}\phi_k\rho^{-k}\to 0$ as $(n\to\infty)$ is implied by the Cauchy-Schwarz inequality. ■

APPENDIX D: Proofs for Section 3.4

LEMMA D.1. *If $a>2$, then*

$$b_0(y_2,x)=1+2\sum_{k=2}^{\infty}\sqrt{Ck^{-a}}\cos\left[(k-1)\pi y_2\right]\cos\left[(k-1)\pi x\right]$$

is bounded from above and bounded away from zero as long as C is sufficiently small.

Proof. It suffices to prove that

$$\left|\sum_{k=2}^{\infty}k^{-\frac{a}{2}}\cos\left[(k-1)\pi y_2\right]\cos\left[(k-1)\pi x\right]\right|<\infty.$$

For this purpose, we note that

$$\left|\sum_{k=2}^{\infty}k^{-\frac{a}{2}}\cos\left[(k-1)\pi y_2\right]\cos\left[(k-1)\pi x\right]\right|\le\sum_{k=2}^{\infty}k^{-\frac{a}{2}}$$

and the term in the right-hand side of the above inequality is absolutely convergent as long as $-\frac{a}{2}<-1$. ■

https://doi.org/10.1017/S0266466617000093 Published online by Cambridge University Press

LEMMA D.2. *Suppose that $a > 2$ and C is sufficiently small. Also suppose that*

$$\frac{1}{a-1} < \delta. \tag{D.1}$$

We then have

$$\int \left(\frac{\sqrt{b_{n^{-1/2}}(y_2,x)} - \sqrt{b_0(y_2,x)}}{n^{-1/2}} \right)^2 \to 0.$$

If

$$\frac{1}{2} + \delta\left(\frac{a'+a}{2} - 2b + 1\right) < 0,$$
$$\frac{1}{2} + \delta\left(a' - 2b + 1\right) \geq 0,$$

then we also have

$$\mathbb{E}\left[n^{\frac{1}{2}} \left(h^2_{n^{-1/2}}(Y_2) - h^2_0(Y_2) \right) \right] \nrightarrow 0.$$

Proof. By the previous lemma, we note that $b_0(y_2,x)$ is bounded away from zero, say by B. We have

$$b_{n^{-1/2}}(y_2,x) - b_0(y_2,x) = 2\sqrt{C} \sum_{k=n^\delta}^{\infty} \frac{-k^{-a/2}}{1+k^{(a-a')/2}} \varphi_k(y_2)\varphi_k(x)$$

and therefore

$$\begin{aligned}\int \left(b_{n^{-1/2}}(y_2,x) - b_0(y_2,x)\right)^2 &= 4C \sum_{k=n^\delta}^{\infty} \left(\frac{-k^{-a/2}}{1+k^{(a-a')/2}} \right)^2 \\ &\leq 4C \sum_{k=n^\delta}^{\infty} \left(k^{-a/2}\right)^2 = 4C \sum_{k=n^\delta}^{\infty} k^{-a} = O\left(n^{\delta(-a+1)}\right).\end{aligned}$$

Because

$$\begin{aligned}\left(\frac{\sqrt{b_{n^{-1/2}}(y_2,x)} - \sqrt{b_0(y_2,x)}}{n^{-1/2}} \right)^2 &= \frac{n\left(b_{n^{-1/2}}(y_2,x) - b_0(y_2,x)\right)^2}{\left(\sqrt{b_{n^{-1/2}}(y_2,x)} + \sqrt{b_0(y_2,x)}\right)^2} \\ &\leq \frac{n\left(b_{n^{-1/2}}(y_2,x) - b_0(y_2,x)\right)^2}{B}\end{aligned}$$

we have

$$\int \left(\frac{\sqrt{b_{n^{-1/2}}(y_2,x)} - \sqrt{b_0(y_2,x)}}{n^{-1/2}} \right)^2 \leq \frac{n}{B} \int \left(b_{n^{-1/2}}(y_2,x) - b_0(y_2,x)\right)^2 = O\left(n^{1+\delta(-a+1)}\right)$$

which together with $\delta > (a-1)^{-1}$ implies the first conclusion.

https://doi.org/10.1017/S0266466617000093 Published online by Cambridge University Press

With (12), we have that

$$h_{n^{-1/2}}(y_2) = \frac{1}{\sqrt{C}}\sum_{k=1}^{n^\delta-1}\frac{k^{-b}}{\sqrt{k^{-a}}}\varphi_k(y_2) + \frac{1}{\sqrt{C}}\sum_{k=n^\delta}^{\infty}\left(\frac{k^{-b}}{\sqrt{k^{-a'}}} + \frac{k^{-b}}{\sqrt{k^{-a}}}\right)\varphi_k(y_2)$$
$$= \frac{1}{\sqrt{C}}\sum_{k=1}^{\infty}\frac{k^{-b}}{\sqrt{k^{-a}}}\varphi_k(y_2) + \frac{1}{\sqrt{C}}\sum_{k=n^\delta}^{\infty}\frac{k^{-b}}{\sqrt{k^{-a'}}}\varphi_k(y_2)$$
$$= h_0(y_2) + \frac{1}{\sqrt{C}}\sum_{k=n^\delta}^{\infty}\frac{k^{-b}}{\sqrt{k^{-a'}}}\varphi_k(y_2).$$

Now, we have

$$\mathbb{E}\left[n^{1/2}\left(h_{n^{-1/2}}(Y_2) - h_0(Y_2)\right)h_0(Y_2)\right] = n^{\frac{1}{2}}\left(\frac{1}{C}\sum_{k=n^\delta}^{\infty}k^{\frac{a'+a}{2}-2b}\right) = O\left(n^{\frac{1}{2}+\delta\left(\frac{a'+a}{2}-2b+1\right)}\right)$$

$$\mathbb{E}\left[n^{1/2}\left(h_{n^{-1/2}}(Y_2) - h_0(Y_2)\right)^2\right] = n^{\frac{1}{2}}\left(\frac{1}{C}\sum_{k=n^\delta}^{\infty}k^{a'-2b}\right) = O\left(n^{\frac{1}{2}+\delta(a'-2b+1)}\right).$$

Therefore, we conclude

$$\mathbb{E}\left[n^{\frac{1}{2}}\left(h^2_{n^{-1/2}}(Y_2) - h_0^2(Y_2)\right)\right] \nrightarrow 0$$

if

$$\frac{1}{2} + \delta\left(\frac{a'+a}{2} - 2b + 1\right) < 0 \text{ and } \frac{1}{2} + \delta\left(a' - 2b + 1\right) \geq 0. \qquad \blacksquare$$

APPENDIX E: Example 3 - Linear Functional under Mild Ill-Posedness

Suppose that (Y_2, X) has the joint density specified in (9). Consider an arbitrary function $r_0 \in L_2([0,1])$ such that the integral equation $\mathbb{E}[h(Y_2)|X] = r_0(X)$ has a unique solution $h_0 \in L_2([0,1])$ almost surely. Now we are interested in the functional value:

$$\theta_0 = \mathbb{E}\left[\ell_0(Y_2)h_0(Y_2)\right]$$

where $\ell_0(Y_2)$ is some function with $\mathbb{E}\left[(\ell_0(Y_2))^2\right] < \infty$. We assume that the solution $v^*_{\ell_0}$ of the integral equation $\mathbb{E}\left[v_{\ell_0}(X)|Y_2\right] = \ell_0(Y_2)$ exists with $\mathbb{E}\left[(v^*_{\ell_0}(X))^2\right] < \infty$.

If $\ell_0(\cdot)$ is *known*, then the problem as in Example 1 does not exist. This is because we can write

$$\theta_0 = \mathbb{E}\left[v^*_{\ell_0}(X)\mathbb{E}\left[h_0(Y_2)|X\right]\right] = \mathbb{E}\left[v^*_{\ell_0}(X)r_0(X)\right].$$

Under the sub-models induced by the perturbation $r_{n^{-1/2}}$, if there is

$$\mathbb{E}\left[\left|n^{\frac{1}{2}}\left[r_{n^{-1/2}}(X) - r_0(X)\right]\right|^2\right] \to 0 \text{ as } n \to \infty,$$

https://doi.org/10.1017/S0266466617000093 Published online by Cambridge University Press

then by the Hölder inequality and $\mathbb{E}\left[(v^*_{\ell_0}(X))^2\right] < \infty$,

$$\left|\frac{\theta_{n^{-1/2}} - \theta_0}{n^{-1/2}}\right| \leq \|v^*_{\ell_0}\|_F \left\|n^{\frac{1}{2}}\left[r_{n^{-1/2}}(X) - r_0(X)\right]\right\|_F \to 0 \text{ as } n \to \infty.$$

See, e.g., Severini and Tripathi (2012) for related discussion.

On the other hand, one may be interested in the case where $\ell_0(Y_2)$ needs to be estimated. We will suppose that $\ell_0(\cdot)$ is a solution to the following non-ill-posed problem

$$\mathbb{E}\left[Y_3 - \ell_0(Y_2)\,|Y_2\right] = 0 \tag{E.1}$$

and see if θ_0 can be estimated regularly. We will consider the perturbations

$$r_{n^{-1/2}}(x) = r_0(x) + n^{-\frac{1}{2}} \sum_{k=1}^{\infty} \phi_{n^{-1/2},k} \varphi_k(x), \tag{E.2}$$

$$h_{n^{-1/2}}(y_2) = h_0(y_2) + n^{-\frac{1}{2}} \sum_{k=1}^{\infty} \phi_{n^{-1/2},k} \rho_k^{-\frac{1}{2}} \varphi_k(y_2), \tag{E.3}$$

$$\ell_{n^{-1/2}}(y_2) = \ell_0(y_2) + n^{-\frac{1}{2}} \sum_{k=1}^{\infty} \chi_{n^{-1/2},k} \rho_k^{-\frac{1}{2}} \varphi_k(y_2) \tag{E.4}$$

where $\phi_{n^{-1/2},k} = n^{-\frac{1}{8}} k^{-\frac{1}{2}-\frac{a}{8}} I\{k \leq n^{\frac{3}{a}}\}$ and $\chi_{n^{-1/2},k} = \phi_{n^{-1/2},k} k^{-\frac{a}{2}} = \phi_{n^{-1/2},k} \rho_k^{1/2}$.

It is clear that as $n \to \infty$, we have

$$\mathbb{E}\left[\left|n^{\frac{1}{2}}\left[r_{n^{-1/2}}(X) - r_0(X)\right]\right|^2\right] = \sum_{k=1}^{\infty} \phi^2_{n^{-1/2},k} = n^{-\frac{1}{4}} \sum_{k=1}^{\lfloor n^{\frac{3}{a}} \rfloor} k^{-1-\frac{a}{4}} \to 0, \tag{E.5}$$

and similarly

$$\mathbb{E}\left[\left|n^{\frac{1}{2}}\left[\ell_{n^{-1/2}}(Y_2) - \ell_0(Y_2)\right]\right|^2\right] = \sum_{k=1}^{\infty} \chi^2_{n^{-1/2},k} \rho_k^{-1} = \sum_{k=1}^{\infty} \phi^2_{n^{-1/2},k} \to 0. \tag{E.6}$$

(In (E.5), we used the fact that the partial sum $\sum_{k=1}^{\lfloor n^{\frac{3}{a}} \rfloor} k^{-1-\frac{a}{4}}$ is convergent because $1 + \frac{a}{4} > 1$.)

Let $g_0(u, v, e, x)$ denote the joint density of (u, v, e, X), where $e = Y_3 - \ell_0(Y_2)$. Then the path $F_{n^{-1/2}}$ we consider is such that its derivative $dF_{n^{-1/2}} = f_{n^{-1/2}}$ is given by

$$f_{n^{-1/2}} = g_0\left[y_1 - r_{n^{-1/2}}(x), y_2 - \pi_0(x), y_3 - \ell_{n^{-1/2}}(y_2), x\right],$$

where the perturbations $r_{n^{-1/2}}$ and $\ell_{n^{-1/2}}$ are defined in (E.2) and (E.4) respectively. Under conditions similar to Assumption B.1, we can use similar arguments in the proof of Lemma B.1 to show that the score of the above path is $s = 0$.

Lemma D.3 below implies that (8) is satisfied in this example.

LEMMA D.3. *$n^{\frac{1}{2}}\left(\theta_{n^{-1/2}} - \theta_0\right) \to \frac{4}{a} \in (0, \infty)$ as $n \to \infty$.*

https://doi.org/10.1017/S0266466617000093 Published online by Cambridge University Press

Proof. We see that

$$\begin{aligned}
&\sqrt{n}\left\{\mathbb{E}\left[\ell_{n^{-1/2}}(Y_2)h_{n^{-1/2}}(Y_2)\right]-\mathbb{E}\left[\ell_0(Y_2)h_0(Y_2)\right]\right\}\\
&\quad=\mathbb{E}\left[\ell_0(Y_2)\sum_{k=1}^{\infty}\phi_{n^{-1/2},k}\rho_k^{-\frac{1}{2}}\varphi_k(Y_2)\right]+\mathbb{E}\left[h_0(Y_2)\sum_{k=1}^{\infty}\chi_{n^{-1/2},k}\rho_k^{-\frac{1}{2}}\varphi_k(Y_2)\right]\\
&\qquad+n^{-\frac{1}{2}}\mathbb{E}\left[\left(\sum_{k=1}^{\infty}\chi_{n^{-1/2},k}\rho_k^{-\frac{1}{2}}\varphi_k(Y_2)\right)\left(\sum_{k=1}^{\infty}\phi_{n^{-1/2},k}\rho_k^{-\frac{1}{2}}\varphi_k(Y_2)\right)\right]\\
&\quad=\sum_{k=1}^{\infty}\phi_{n^{-1/2},k}\chi_k\rho_k^{-1}+\sum_{k=1}^{\infty}\chi_{n^{-1/2},k}\phi_k\rho_k^{-1}+n^{-\frac{1}{2}}\sum_{k=1}^{\infty}\chi_{n^{-1/2},k}\phi_{n^{-1/2},k}\rho_k^{-1}. \qquad \textbf{(E.7)}
\end{aligned}$$

By (E.5) and (E.6), we get that $\sum_{k=1}^{\infty}\phi_{n^{-1/2},k}^2\to 0$ and $\sum_{k=1}^{\infty}\chi_{n^{-1/2},k}^2\le\sum_{k=1}^{\infty}\phi_{n^{-1/2},k}^2\to 0$ as $n\to\infty$, which together with the Cauchy-Schwarz inequality imply that

$$\left|\sum_{k=1}^{\infty}\phi_{n^{-1/2},k}\chi_k\rho_k^{-1}\right|^2\le\sum_{k=1}^{\infty}\phi_{n^{-1/2},k}^2\sum_{k=1}^{\infty}\chi_k^2\rho_k^{-2}\le\sup_{k\ge1}\left|d_k^2\right|\sum_{k=1}^{\infty}\phi_{n^{-1/2},k}^2\sum_{k=1}^{\infty}k^{-2(c-a)}\to0, \qquad \textbf{(E.8)}$$

$$\left|\sum_{k=1}^{\infty}\chi_{n^{-1/2},k}\phi_k\rho_k^{-1}\right|^2\le\sum_{k=1}^{\infty}\chi_{n^{-1/2},k}^2\sum_{k=1}^{\infty}\phi_k^2\rho_k^{-2}=\sum_{k=1}^{\infty}\chi_{n^{-1/2},k}^2\sum_{k=1}^{\infty}k^{-2(b-a)}\to0. \qquad \textbf{(E.9)}$$

Moreover, we see that

$$\begin{aligned}
n^{-\frac{1}{2}}\sum_{k=1}^{\infty}\chi_{n^{-1/2},k}\phi_{n^{-1/2},k}\rho_k^{-1}&=n^{-\frac{1}{2}}\sum_{k=1}^{\lfloor n^{\frac{3}{a}}\rfloor}\left(n^{-\frac{1}{8}}k^{-\frac{1}{2}-\frac{a}{8}}\right)^2k^{-\frac{a}{2}}k^a\\
&=\sum_{k=1}^{\lfloor n^{\frac{3}{a}}\rfloor}\left(kn^{-\frac{3}{a}}\right)^{-1+\frac{a}{4}}n^{-\frac{3}{a}}\to\int_0^1x^{-1+\frac{a}{4}}dx=\frac{4}{a}\in(0,\infty)
\end{aligned}$$

as $n\to\infty$, which together with (E.7), (E.8) and (E.9) implies that

$$n^{\frac{1}{2}}\left(\theta_{n^{-1/2}}-\theta_0\right)=n^{\frac{1}{2}}\left\{\mathbb{E}\left[\ell_{n^{-1/2}}(Y_2)h_{n^{-1/2}}(Y_2)\right]-\mathbb{E}\left[\ell_0(Y_2)h_0(Y_2)\right]\right\}\to\frac{4}{a}\in(0,\infty). \qquad \blacksquare$$

https://doi.org/10.1017/S0266466617000093 Published online by Cambridge University Press