NONPARAMETRIC INSTRUMENTAL VARIABLE ESTIMATION OF STRUCTURAL QUANTILE EFFECTS

PATRICK GAGLIARDINI OLIVIER SCAILLET

Abstract. We study the asymptotic distribution of Tikhonov Regularized estimation of quantile structural effects implied by a nonseparable model. The nonparametric instrumental variable estimator is based on a minimum distance principle. We show that the minimum distance problem without regularization is locally ill-posed, and consider penalization by the norms of the parameter and its derivatives. We derive pointwise asymptotic normality and develop a consistent estimator of the asymptotic variance. We study the small sample properties via simulation results, and provide an empirical illustration to estimation of nonlinear pricing curves for telecommunications services in the U.S.

Key words: Nonparametric Quantile Regression, Instrumental Variable, Ill-Posed Inverse Problems, Tikhonov Regularization, Nonlinear Pricing Curve.
JEL classification: C13, C14, D12, G12. AMS 2000 classification: 62G08, 62G20, 62P20.

Date: September 2011.

We are very grateful to Victor Chernozhukov who was a co-author of the version that went through the review process of Econometrica. We also thank him for providing the tightness argument for the consistency proof. We thank the editor and referees for helpful suggestions. We thank J. Hausman and G. Sidak for generously providing us with the data on telecommunications services in the U.S. We are especially grateful to X. Chen for comments that have led to major improvements in the paper. We also thank G. Dhaene, R. Koenker, O. Linton, E. Mammen, M. Wolf, and seminar participants at Athens University, Zurich University, Bern University, Boston University, Queen Mary, Greqam Marseille, Mannheim University, Leuven University, Toulouse University, ESEM 2008 and SITE 2009 workshop for helpful comments. The authors acknowledge the support provided by the Swiss National Science Foundation through the National Center of Competence in Research: Financial Valuation and Risk Management (NCCR FINRISK).

1. Introduction

We propose and analyze estimators of the nonseparable model $Y = g(X, U)$, where the error U is independent of the instrument Z, and has a uniform distribution $U \sim \mathcal{U}(0,1)$ (Chernozhukov and Hansen (2005), Chernozhukov, Imbens and Newey (2007)). The function $g(x,u)$ is strictly monotonic increasing w.r.t. $u \in [0,1]$. The variable X has compact support $\mathcal{X} = [0,1]$ and is potentially endogenous. The variables Y and Z have compact supports $\mathcal{Y} \subset [0,1]$ and $\mathcal{Z} = [0,1]^{d_Z}$. The parameter of interest is the quantile structural effect $\varphi_0(x) = g(x,\tau)$ on $\mathcal{X}$ for a given $\tau \in (0,1)$. The function φ_0 measures the structural impact of the regressor X on the τ-quantile of the dependent variable Y. Formally it satisfies the conditional quantile restriction

$$P\left[Y \leq \varphi_0(X) \mid Z\right] = P\left[g(X,U) \leq g(X,\tau) \mid Z\right] = P\left[U \leq \tau \mid Z\right] = \tau, \tag{1.1}$$

which yields the endogenous quantile regression representation (Horowitz and Lee (2007)):

$$Y = \varphi_0(X) + V, \qquad\qquad P\left[V \leq 0 \mid Z\right] = \tau. \tag{1.2}$$

The main contribution of this paper is the derivation of the large sample distribution of a Tikhonov regularized estimator of φ_0. This is the first distributional result in the literature on nonlinear problems, and it is noteworthy because of a fundamental difficulty of linearization of a nonlinear ill-posed problem such as (1.2), as pointed out in Horowitz and Lee (2007). Even though this paper focuses on a particular case (1.2) of a nonlinear ill-posed problem, the results of the paper are conceptually amenable to other problems. Indeed, the nonsmooth case (1.2) analyzed here is in some sense the hardest; so our analysis could be applied to other problems, such as nonlinear ill-posed pricing problems in finance (see e.g. Egger and Engl (2005), Chen and Ludvigson (2009)) along similar lines.

We build on a series of fundamental papers on ill-posed endogenous mean regressions (Ai and Chen (2003), Darolles, Fan, Florens, and Renault (2011), Newey and Powell (2003), Hall and Horowitz (2005), Horowitz (2007), Blundell, Chen, and Kristensen (2007)), and the review paper by Carrasco, Florens, and Renault (CFR, 2007). The main issue in nonparametric estimation with endogeneity is overcoming ill-posedness of the associated inverse problem. It occurs since the mapping of the reduced form parameter (that is, the distribution of the data) into the structural parameter (that is, the instrumental regression function) is not continuous. We need a regularization of the estimation to recover consistency. Here we follow Gagliardini and Scaillet (GS, 2011a) and study a Tikhonov Regularized (TiR) estimator (Tikhonov (1963a,b), Groetsch (1984), Kress (1999)). We achieve regularization by adding a compactness-inducing penalty term, the Sobolev norm, to a functional minimum distance criterion. For nonparametric instrumental variable estimation of endogenous quantile regression

(NIVQR), Chernozhukov, Imbens and Newey (2007) discuss identification and estimation via a constrained minimum distance criterion. Horowitz and Lee (2007) give optimal consistency rates for a L^2-norm penalized estimator.

In independent work for a general setting, Chen and Pouzo (2009, 2011) study semiparametric sieve estimation of conditional moment models based on possibly nonsmooth generalized residual functions. Specifically, Chen and Pouzo (2009) focus on the semiparametric efficiency, asymptotic normality, and a weighted bootstrap procedure for the finite-dimensional parameter, and use a finite-dimensional sieve to estimate the functional parameter. They cover partially linear IVQR as a particular example. Chen and Pouzo (2011) give an in-depth, unifying treatment of convergence rates of penalized sieve-based estimators, and characterize when the sieve or the penalization dominates the convergence rates. Our and results of Chen and Pouzo (2009, 2011) are complementary to each other (their results do not nest our results, and vice versa, since we work in an infinite-dimensional parameter space, while Chen and Pouzo (2011) work in finite-dimensional parameter spaces of increasing dimensions). The most important difference is the derivation of pointwise asymptotic normality which is not available in Chen and Pouzo (2009, 2011). Our other specific contributions for NIVQR include a proof of ill-posedness and a proof of consistency under weak conditions on the penalization parameter.

We organize the rest of the paper as follows. In Section 2, we prove local ill-posedness, and clarify the importance of including a derivative in the penalization. In Section 3, we prove consistency of our Q-TiR estimator. In Section 4, we show pointwise asymptotic normality, and introduce a consistent estimator of the asymptotic variance. In Section 5, we provide computational experiments and present an empirical illustration to estimation of nonlinear pricing curves for telecommunications services in the U.S. In the Appendix, we gather the technical assumptions and some proofs. We place all omitted proofs in the online supplementary materials (Gagliardini and Scaillet (2011b)).

2. Ill-posedness in nonseparable models

From (1.1), the quantile structural effect φ_0 is a solution of the nonlinear functional equation $\mathcal{A}(\varphi_0) = \tau$, where the operator $\mathcal{A}$ is defined by

$$\mathcal{A}(\varphi)(z) = \int F_{Y|X,Z}(\varphi(x)|x,z)\, f_{X|Z}(x|z)dx, \quad z \in \mathcal{Z},$$

and $F_{Y|X,Z}$ and $f_{X|Z}$ denote the c.d.f. of Y given X, Z, and the p.d.f. of X given Z, respectively. Alternatively, in terms of the conditional c.d.f. $F_{U|X,Z}$ of U given X, Z, we can rewrite $\mathcal{A}(\varphi)(z) = \int F_{U|X,Z}\left(g^{-1}(x,\varphi(x))|x,z\right) f_{X|Z}(x|z)dx$, where $g^{-1}(x,.)$ denotes the generalized inverse of function $g(x,.)$ w.r.t. its second argument. The functional parameter φ_0

belongs to a subset Θ of the weighted Sobolev space $H^l[0,1]$, $l \in \mathbb{N} \cup \{\infty\}$, that is the completion of $\left\{\varphi \in C^l[0,1] \mid \|\varphi\|_H < \infty\right\}$ w.r.t. the weighted Sobolev norm $\|\varphi\|_H := \langle \varphi, \varphi \rangle_H^{1/2}$, where $\langle \varphi, \psi \rangle_H := \sum_{s=0}^{l} a_s \langle \nabla^s \varphi, \nabla^s \psi \rangle$, $a_s > 0$, is the weighted Sobolev scalar product, and $\langle \varphi, \psi \rangle = \int \varphi(x)\psi(x)dx$. Below we focus on (i) $a_s = 1$, when $l < \infty$, (ii) $a_s = 1/s!$, when $l = \infty$. The former gives the classical Sobolev space of order l (Adams and Fournier (2003)) while the latter gives the Sobolev space $H^\infty[0,1] := \left\{\varphi \in C^\infty[0,1] \mid \sum_{s=0}^{\infty} \frac{1}{s!} \langle \nabla^s \varphi, \nabla^s \varphi \rangle < \infty\right\}$ of infinite order (Dubinskij (1986)). These Sobolev spaces are Hilbert spaces w.r.t. $\langle .,. \rangle_H$, and the embeddings $H^\infty[0,1] \subset H^l[0,1] \subset L^2[0,1]$, $l \in \mathbb{N}$, are compact (Adams and Fournier (2003), Theorem 6.3). We use the L^2-norm $\|\varphi\| = \langle \varphi, \varphi \rangle^{1/2}$ as consistency norm, and we assume that Θ is bounded and closed w.r.t. $\|.\|$.

We assume that φ_0 is globally identified on Θ (see Appendix C in Chernozhukov and Hansen (2005) for a discussion) and interior.

Assumption 1: (i) $\mathcal{A}(\varphi) - \tau = 0$, $\varphi \in \Theta$, *if and only if* $\varphi = \varphi_0$; (ii) *function* φ_0 *is an interior point of set* Θ *w.r.t. norm* $\|.\|$.

We use the conditional moment restriction $m(\varphi_0, z) := \mathcal{A}(\varphi_0)(z) - \tau = 0$, $z \in \mathcal{Z}$, and consider the criterion

$$Q_\infty(\varphi) := \frac{1}{\tau(1-\tau)} E\left[m(\varphi, Z)^2\right] =: \|\mathcal{A}(\varphi) - \tau\|^2_{L^2(F_Z,\tau)},$$

where F_Z denotes the marginal distribution of Z and $L^2(F_Z, \tau)$ denotes the L^2 space w.r.t. measure $F_Z/(\tau(1-\tau))$. The true structural function φ_0 minimizes this criterion function Q_∞.

The following proposition shows that the minimum distance problem above is locally ill-posed (see e.g. Definition 1.1 in Hofmann and Scherzer (1998)). There are sequences of increasingly oscillatory functions arbitrarily close to φ_0 that approximately minimize Q_∞ while not converging to φ_0. In other words, function φ_0 is not identified in Θ as an isolated minimum of Q_∞. Therefore, ill-posedness can lead to inconsistency of the naive analog estimators based on the empirical analog of Q_∞. In order to rule out these explosive solutions we use penalization.

Proposition 1: *Under Assumptions 1 (i) and A.3:* (a) *the problem is locally ill-posed, namely for any* $r > 0$ *small enough, there exist* $\varepsilon \in (0, r)$ *and a sequence* $(\varphi_n) \subset B_r(\varphi_0) := \left\{\varphi \in L^2[0,1] : \|\varphi - \varphi_0\| < r\right\}$ *such that* $\|\varphi_n - \varphi_0\| \geq \varepsilon$ *and* $Q_\infty(\varphi_n) \to Q_\infty(\varphi_0) = 0$; (b) *any sequence* $(\varphi_n) \subset B_r(\varphi_0)$ *such that* $\|\varphi_n - \varphi_0\| \geq \varepsilon$ *for* $r > \varepsilon > 0$ *and* $Q_\infty(\varphi_n) \to 0$ *satisfies*

$$\limsup_{n \to \infty} \|\nabla \varphi_n\| = +\infty.$$

The proof of result (a) gives explicit sequences (φ_n) generating ill-posedness. Since there is no general characterization of the ill-posedness of a nonlinear problem through conditions on its linearization, i.e., on the Frechet derivative of the operator (Engl, Kunisch and Neubauer

(1989), Schock (2002)), this result does not follow from the ill-posedness of the linearized version of our problem. Under a stronger condition than Assumption 1 (i), namely local injectivity of $\mathcal{A}$, the definition of local ill-posedness is equivalent to $\mathcal{A}^{-1}$ being discontinuous in a neighborhood of $\mathcal{A}(\varphi_0)$ (see Engl, Hanke and Neubauer (2000), Chapter 10). Part (b) provides a theoretical underpinning for including the norm $\|\nabla\varphi\|$ of the derivative in the penalty term.

3. Consistency of the Q-TiR estimator

We consider a penalized criterion $L_T(\varphi) := Q_T(\varphi) + \lambda_T \|\varphi\|_H^2$, where $\lambda_T > 0$, P-a.s., and

$$Q_T(\varphi) := \frac{1}{T\tau(1-\tau)} \sum_{t=1}^{T} \hat{m}(\varphi, Z_t)^2 .$$

The conditional moment $m(\varphi, z)$ is estimated nonparametrically by

$$\hat{m}(\varphi, z) := \int \hat{F}_{Y|X,Z}(\varphi(x)|x,z)\,\hat{f}_{X|Z}(x|z)dx - \tau =: \hat{\mathcal{A}}(\varphi)(z) - \tau, \quad z \in \mathcal{Z}, \tag{3.1}$$

where $\hat{f}_{X|Z}$ and $\hat{F}_{Y|X,Z}$ denote kernel estimators of $f_{X|Z}$ and $F_{Y|X,Z}$ with bandwidth $h_T > 0$ and kernel K satisfying Assumption A.2.

Proposition 2: *Suppose λ_T is a stochastic sequence such that $\lambda_T > 0$, $\lambda_T \to 0$, P-a.s., and $\frac{1}{\lambda_T}\left(\frac{\log T}{T h_T^{d_Z+1}} + h_T^{2m}\right) = O_p(1)$, where $m \geq 2$ is the order of differentiability of the joint density $f_{X,Y,Z}$ of (X, Y, Z). Then, under Assumptions 1 (i) and A.1-A.3, the Q-TiR estimator $\hat{\varphi}$ defined by*

$$\hat{\varphi} := \arg\inf_{\varphi\in\Theta} Q_T(\varphi) + \lambda_T \|\varphi\|_H^2 , \tag{3.2}$$

is consistent, namely $\|\hat{\varphi} - \varphi_0\| \xrightarrow{p} 0$, as sample size $T \to \infty$.

Term $\lambda_T \|\varphi\|_H^2$ in definition (3.2) penalizes highly oscillating components of the estimated function induced by ill-posedness, and restores its consistency. In (3.2) we work with a function space-based estimator as in Horowitz and Lee (2007) (see also the suggestion in Newey and Powell (2003), p. 1573). In Section 5 we compute the estimator based on a finite large number of polynomials. The discrepancy between the function space-based estimator and the implemented estimator is of a numerical nature since our type of asymptotics does not rely on a sieve approach.

To show Proposition 2 we use two results. First, the Sobolev penalty implies that the sequence of estimates $\hat{\varphi}$ for $T \in \mathbb{N}$ is tight in $\left(L^2[0,1], \|.\|\right)$. This induces an effective compactification of the parameter space: there exists a compact set that contains $\hat{\varphi}$ for any large T with probability $1-\delta$, for any arbitrarily small $\delta > 0$. Second, we obtain a suitable uniform convergence result for Q_T on an infinite-dimensional and possibly non-totally bounded parameter set Θ by exploiting the specific expression of $\hat{m}(\varphi, z)$ given in (3.1). We are able to reduce the sup over Θ to a sup over a bounded subset of a finite-dimensional space. Proposition 6.2 in Chen and Pouzo (2011) states a consistency result for nonparametric additive IVQR using a series estimator for $m(\varphi, z)$ under similar conditions. In the rest of the paper we assume a deterministic regularization parameter λ_T. The assumption in Proposition 2 becomes $h_T \asymp T^{-\eta}$ and $\lambda_T \asymp T^{-\gamma}$ for $0 < \eta < \frac{1}{d_Z+1}$, $0 < \gamma < \min\{1 - \eta(d_Z + 1), 2m\eta\}$; the relation $a_T \asymp b_T$, for positive sequences a_T and b_T, means that a_T/b_T is bounded away from 0 and ∞ as $T \to \infty$.

4. Asymptotic distribution of the Q-TiR estimator

In this section we derive a feasible asymptotic normality theorem. After deriving the first-order condition (Section 4.1) we show how to control the error induced by linearization of the problem under suitable smoothness assumptions (Sections 4.2 and 4.3). The validity of a Bahadur-type representation for the functional estimator makes it possible to show asymptotic normality (Section 4.4). We then provide a consistent estimator for the asymptotic variance (Section 4.5).

4.1. **First-order condition.** The asymptotic expansion of the Q-TiR estimator is derived by following the same steps as in the usual finite-dimensional setting. To cope with the functional nature of φ_0, we exploit an appropriate notion of differentiation to get the first-order condition. More precisely, we introduce the operator from $L^2[0,1]$ to $L^2(F_Z, \tau)$ corresponding to the Frechet derivative $A := D\mathcal{A}(\varphi_0)$ of operator $\mathcal{A}$ at φ_0:

$$A\varphi(z) = \int f_{X,Y|Z}(x, \varphi_0(x)|z)\varphi(x)\,dx,$$

and the operator from $L^2[0,1]$ to $L^2(\hat{F}_Z, \tau)$ corresponding to the Frechet derivative $\hat{A} := D\hat{\mathcal{A}}(\hat{\varphi})$ of operator $\hat{\mathcal{A}}$ at $\hat{\varphi}$ (see Appendix A.4):

$$\hat{A}\varphi(z) = \int \hat{f}_{X,Y|Z}(x, \hat{\varphi}(x)|z)\varphi(x)\,dx,$$

where $z \in \mathcal{Z}$ and $\varphi \in L^2[0,1]$. The space $L^2(\hat{F}_Z, \tau)$ is endowed with the scalar product $\langle \psi_1, \psi_2 \rangle_{L^2(\hat{F}_Z,\tau)} := \frac{1}{T\tau(1-\tau)} \sum_{t=1}^T \psi_1(Z_t)\,\psi_2(Z_t)$. Under Assumptions A.4 (ii)-(iii), operator A is compact, which implies that the linearized version of our problem is ill-posed. Under Assumption 1 (ii), the Q-TiR estimator satisfies w.p.a. 1 the first-order condition

$$\begin{aligned} 0 &= \frac{d}{d\varepsilon} L_T(\hat{\varphi} + \varepsilon\varphi)\Big|_{\varepsilon=0} = \frac{2}{T\tau(1-\tau)} \sum_{t=1}^T \left(\hat{\mathcal{A}}(\hat{\varphi})(Z_t) - \tau\right) \hat{A}\varphi(Z_t) + 2\lambda_T \langle \hat{\varphi}, \varphi \rangle_H \\ &= 2\left\langle \hat{A}^*\left(\hat{\mathcal{A}}(\hat{\varphi}) - \tau\right) + \lambda_T \hat{\varphi}, \varphi \right\rangle_H, \end{aligned} \tag{4.1}$$

for all $\varphi \in H^l[0,1]$, where the second line in (4.1) comes from the definition of the operator $\hat{A}^*$ through $\left\langle \hat{A}^*\psi, \varphi \right\rangle_H := \frac{1}{T\tau(1-\tau)} \sum_{t=1}^{T} \psi(Z_t)\, \hat{A}\varphi(Z_t)$, for any $\varphi \in H^l[0,1]$ and $\psi \in L^2(\hat{F}_Z, \tau)$. Operator $\hat{A}^*$ is the adjoint of $\hat{A}$ w.r.t. the scalar products $\langle ., . \rangle_H$ on $H^l[0,1]$ and $\langle ., . \rangle_{L^2(\hat{F}_Z,\tau)}$ on $L^2(\hat{F}_Z, \tau)$. It is the empirical counterpart of the adjoint operator A^* of A w.r.t. the Sobolev scalar product on $H^l[0,1]$ and scalar product $\langle ., . \rangle_{L^2(F_Z,\tau)}$ on $L^2(F_Z, \tau)$. The operator A^* can be characterized in terms of the adjoint $\tilde{A}$ of A w.r.t. the $L^2[0,1]$ scalar product, defined by $\tilde{A}\psi(x) = \frac{1}{\tau(1-\tau)} \int f_{X,Y,Z}(x, \varphi_0(x), z)\psi(z)dz$. For $l = 1$ we have $A^* = \mathcal{D}^{-1}\tilde{A}$, where $\mathcal{D}^{-1}$ denotes the inverse of operator $\mathcal{D} : H_0^2[0,1] \to L^2[0,1]$ with $H_0^2[0,1] := \left\{\varphi \in H^2[0,1] : \nabla\varphi(0) = \nabla\varphi(1) = 0\right\}$ and $\mathcal{D}\varphi = \left(1 - \nabla^2\right)\varphi$ (see the supplementary materials for the derivation and the characterization with $l > 1$). From (4.1) holding for all $\varphi \in H^l[0,1]$, estimate $\hat{\varphi}$ satisfies the nonlinear integro-differential equation of Type II:

$$\hat{A}^*\left(\hat{\mathcal{A}}(\hat{\varphi}) - \tau\right) + \lambda_T\hat{\varphi} = 0. \tag{4.2}$$

4.2. **Highlighting the nonlinearity issue.** We can rewrite Equation (4.2) by using the second-order expansion

$$\widehat{\mathcal{A}}(\hat{\varphi}) = \widehat{\mathcal{A}}(\varphi_0) + \hat{A}_0\Delta\hat{\varphi} + \hat{R},$$

where

$$\hat{A}_0\varphi(z) := \int \hat{f}_{X,Y|Z}(x, \varphi_0(x)|z)\varphi(x)dx, \;\; \Delta\hat{\varphi} := \hat{\varphi} - \varphi_0,$$

and $\hat{R} := \hat{R}(\hat{\varphi}, \varphi_0)$ is the second-order residual term (see Lemma A.4). Then, after rearranging terms and performing an asymptotic expansion, we get the decomposition (Appendix A.4):

$$\Delta\hat{\varphi} = (\lambda_T + A^*A)^{-1} A^*\hat{\zeta} + \mathcal{B}_T + \mathcal{R}_T + \hat{\mathcal{K}}_T(\Delta\hat{\varphi}), \tag{4.3}$$

where $\hat{\zeta}(z) := \int\int (\tau - 1\{y \le \varphi_0(x)\}) \frac{\Delta\hat{f}_{X,Y,Z}(x,y,z)}{f_Z(z)}dydx$, $\Delta\hat{f} = \hat{f} - f$, and

$$\mathcal{B}_T := \left[(\lambda_T + A^*A)^{-1} A^*A - 1\right]\varphi_0.$$

The Bahadur-type representation (4.3) of the Q-TiR estimator (see e.g. Koenker (2005)) is key to our asymptotic normality result. The stochastic term $(\lambda_T + A^*A)^{-1} A^*\hat{\zeta}$ corresponds to the leading term yielding asymptotic normality of the standard exogenous quantile regression. The deterministic term $\mathcal{B}_T$ is the bias function induced by regularization. The remainder term $\mathcal{R}_T$ accounts for kernel estimation of operator $\mathcal{A}$ and its expression is given in (A.6) below. The nonlinearity term $\hat{\mathcal{K}}_T(\Delta\hat{\varphi})$ is defined by $\hat{\mathcal{K}}_T(\Delta\hat{\varphi}) := -\left(\lambda_T + \hat{A}_0^*\hat{A}_0\right)^{-1}\hat{A}_0^*\hat{R}$, where $\hat{A}_0^*$ is defined as $\hat{A}^*$, but with φ_0 substituted for $\hat{\varphi}$.

The major difference between our ill-posed setting and standard finite-dimensional parametric estimation problems, or well-posed functional estimation problems, concerns the behaviour of the nonlinearity term $\hat{\mathcal{K}}_T(\Delta\hat{\varphi})$. Controlling $\hat{\mathcal{K}}_T(\Delta\hat{\varphi})$ is a fundamental difficulty of linearization of a nonlinear ill-posed inverse problem. We prove in Lemma A.5 that $\hat{\mathcal{K}}_T(\Delta\hat{\varphi})$ satisfies a quadratic bound

$$\left\|\hat{\mathcal{K}}_T(\Delta\hat{\varphi})\right\| \leq \frac{C}{\sqrt{\lambda_T}}\|\Delta\hat{\varphi}\|^2, \tag{4.4}$$

w.p.a. 1, with a suitable constant C. In the RHS of (4.4) the coefficient of the quadratic bound diverges as the sample size increases. Hence, the usual argument that the quadratic nonlinearity term is negligible w.r.t. the first-order term no matter the convergence rate of the latter, does not apply. Still, under Assumptions 2-4 discussed below and ensuring $\|\Delta\hat{\varphi}\| = o_p\left(\sqrt{\lambda_T}\right)$, we can control the nonlinearity term $\hat{\mathcal{K}}_T(\Delta\hat{\varphi})$ and the remainder term $\mathcal{R}_T$.

4.3. **Assumptions for asymptotic normality.** Let $\varphi_\lambda := \arg\inf_{\varphi\in\Theta} Q_\infty(\varphi) + \lambda\|\varphi\|_H^2$, with $\lambda > 0$, denote the nonlinear Tikhonov solution in the population. We will make the following assumptions as well as the more technical Assumptions A.4-5 stated in the Appendix.

Assumption 2: *The solution φ_λ is unique and the equation $D\mathcal{A}(\varphi)^*(\mathcal{A}(\varphi)-\tau)+\lambda\varphi = 0$, $\varphi\in\Theta$, admits the unique solution $\varphi = \varphi_\lambda$, for all small $\lambda > 0$.*

This assumption involves the first-order condition for minimization of the penalized minimum distance criterion in the population. It rules out local extrema over Θ different from the global minimum φ_λ.

Assumption 3: *$A\varphi = 0$ for $\varphi \in L^2[0,1]$ if and only if $\varphi = 0$.*

This is an injectivity condition for the Frechet derivative A, that is, a local identification assumption. Since operator A is such that $A\varphi(z) = E\left[\left\{\frac{\partial g}{\partial u}(X,\tau)\right\}^{-1}\varphi(X)|Z=z, U=\tau\right]$ and $\partial g/\partial u > 0$, Assumption 3 is equivalent to completeness of X by Z conditional on $U=\tau$; see Chernozhukov, Imbens and Newey (2007) for examples and further discussion on the relationship with completeness.

Assumption 4: (i) *The function φ_0 satisfies the source condition $\sum_{j=1}^{\infty}\frac{\langle\phi_j,\varphi_0\rangle_H^2}{\nu_j^{2\delta}} < \infty$, $\delta\in(1/2,1]$, where $\nu_j \searrow 0$ and ϕ_j are the eigenvalues and eigenfunctions of the compact, self-adjoint operator A^*A on $H^l[0,1]$, with $\|\phi_j\|_H = 1$.* (ii) $\sum_{j=1}^{\infty}\frac{\langle\phi_j,\varphi_0\rangle_H^2}{\nu_j} < 1/c^2$, *where $c := \sup_{x,y,z}\left|\nabla_y f_{X,Y|Z}(x,y|z)\right|$.* (iii) $\Gamma(\lambda) := \inf_{\substack{\varphi\in H^l[0,1]:\\ \|\varphi\|=1}}\|A\varphi\|_{L^2(F_Z,\tau)}^2 + \lambda\|\varphi\|_H^2 \geq C\lambda^a$, *for $C>0$, $a\in(0,1/2)$.*

The source condition in Assumption 4 (i) requires that function φ_0 can be well approximated by the elements of the eigenfunction basis $\{\phi_j : j\in\mathbb{N}\}$ associated with the larger eigenvalues of operator A^*A. The coefficients $\langle\phi_j,\varphi_0\rangle_H$ have to decrease to zero as $j\to\infty$ sufficiently fast compared to the eigenvalues ν_j. This condition controls the bias contribution as in the proof of Proposition 3.11 in CFR, and implies (see Appendix A.4)

$$\|\mathcal{B}_T\|_H = O\left(\lambda_T^{\delta}\right). \tag{4.5}$$

Hence, the larger the parameter δ, the smaller the regularization bias. Assumption 4 (ii) is the analog of Assumption 6 in Horowitz and Lee (2007) for Sobolev penalization. It controls the distance between the nonlinear Tikhonov solution in the population and φ_0 along the lines of Proposition 10.7 in Engl, Hanke and Neubauer (2000). Together with Assumption 4 (i) it implies $\|\varphi_{\lambda_T} - \varphi_0\|_H = O\left(\lambda_T^{\delta}\right)$ (see the proof of Lemma B.6 in the supplementary materials). Finally, Assumption 4 (iii) implies that

$$\inf_{\substack{\varphi\in\Theta: \\ \|\varphi-\varphi_0\|\geq d\sqrt{\lambda}}} Q_\infty(\varphi) + \lambda \|\varphi\|_H^2 - Q_\infty(\varphi_0) - \lambda \|\varphi_0\|_H^2 \geq C\lambda\Gamma(\lambda), \tag{4.6}$$

as $\lambda \to 0$, for any $C < d^2$ (see Lemma B.6 in the supplementary materials). Inequality (4.6) replaces the usual "identifiable uniqueness" condition (White and Wooldridge (1991)) $\inf_{\varphi\in\Theta:\|\varphi-\varphi_0\|\geq\varepsilon} Q_\infty(\varphi) - Q_\infty(\varphi_0) > 0$, which does not hold with ill-posedness (see Proposition 1 (a)). Function $\Gamma(\lambda)$ characterizes how well the penalized criterion distinguishes φ_0 from functions φ outside a neighbourhood of φ_0, when the radius $\sqrt{\lambda}$ of the neighbourhood shrinks to zero. The smaller the parameter a, the more effective the penalization restores "identifiable uniqueness". Inequality (4.6) is sharp in the sense that for a linear problem the LHS is equal to $d^2\lambda\Gamma(\lambda) + O\left(\lambda^{3/2}\right)$. Assumptions 2-4 are used to control the large deviation probability $P[\|\Delta\hat{\varphi}\| \geq d\sqrt{\lambda_T}]$, $d > 0$ (see Lemma B.7 in the supplementary materials). From (4.4) this yields an upper bound for the nonlinearity term $\hat{\mathcal{K}}_T(\Delta\hat{\varphi})$ (see Lemmas A.7 and A.9).

In order to discuss plausibility of Assumption 4 (iii), let $\{\tilde{\phi}_j : j \in \mathbb{N}\}$ be an orthonormal basis in $L^2[0,1]$ of eigenfunctions of $\tilde{A}A$ to eigenvalues $\tilde{\nu}_j$. Consider the three conditions

$$(a) \quad \sum_{j,k=1:j\neq k}^{\infty} \frac{\langle\tilde{\phi}_j, \tilde{\phi}_k\rangle_H^2}{\|\tilde{\phi}_j\|_H^2 \|\tilde{\phi}_k\|_H^2} < 1, \quad (b)\ \|\tilde{\phi}_j\|_H^2 \geq C_1\omega_j, \quad (c)\ \tilde{\nu}_j \geq C_2\kappa_j, \ \ C_1, C_2 > 0, \quad \forall j \geq 1,$$

where $\omega_j \nearrow \infty$ and $\kappa_j \searrow 0$ are two given positive sequences. Condition (a) states that the eigenfunctions are not very correlated under $\langle .,. \rangle_H$. Condition (b) gives a lower bound on the speed of divergence of the Sobolev norms of the eigenfunctions in terms of sequence ω_j. Similarly, condition (c) gives a lower bound on the speed of convergence to zero of the eigenvalues in terms of sequence κ_j. The divergence of sequence ω_j must dominate the convergence to zero of sequence κ_j for Assumption 4 (iii) to hold. Specifically, conditions (a)-(c) with $\omega_j = j^p$ and $\kappa_j = j^{-\tilde{\alpha}}$, $0 < \tilde{\alpha} < p$, imply Assumption 4 (iii) with $a = \frac{\tilde{\alpha}}{\tilde{\alpha}+p}$. The hyperbolic decay of $\tilde{\nu}_j$ corresponds to mild ill-posedness for operator $\tilde{A}A$. Moreover, conditions (a)-(c) with $\omega_j = \exp(j^2)$ and $\kappa_j = \exp(-\tilde{\alpha}j)$, $\tilde{\alpha} > 0$, imply Assumption 4 (iii) with any $a \in (0, 1/2)$. The geometric decay of $\tilde{\nu}_j$ corresponds to severe ill-posedness for operator $\tilde{A}A$ (see CFR and Kress (1999) for the terminology).

Remark 1. *In the online supplementary materials we provide an example where the orthonormal eigenfunctions of $\tilde{A}A$ in $L^2[0,1]$ are given by $\tilde{\phi}_j(x) = 1$, $\tilde{\phi}_j(x) = \sqrt{2}\cos(\pi(j-1)x)$, $j = 2, 3, \ldots$ These functions satisfy $\langle \tilde{\phi}_j, \tilde{\phi}_k \rangle_H = 1\{j = k\} \sum_{s=0}^{l} a_s (\pi j)^{2s}$. Thus, for hyperbolic decay $\tilde{\nu}_j \geq C j^{-\tilde{\alpha}}$ and finite Sobolev order $l > \tilde{\alpha}/2$, Assumption 4 (iii) holds with $a = \frac{\tilde{\alpha}}{\tilde{\alpha}+2l}$. Similarly, for geometric decay $\tilde{\nu}_j \geq C \exp(-\tilde{\alpha} j)$ and Sobolev order $l = \infty$, Assumption 4 (iii) holds with any $a \in (0, 1/2)$.*

The example in Remark 1 shows that Assumption A.4 (iii) involves an adaptation condition between the speed of decay of the spectrum of operator $\tilde{A}A$ (mild, resp. severe, ill-posednesss) and the Sobolev order l (finite, resp. infinite). This is related to condition (8.28) for operator A in Engl, Hanke and Neubauer (2000). Within our setting of assumptions for asymptotic normality, allowing for higher-order Sobolev penalties permits to accommodate various forms of ill-posedness. While the decay behaviours of the spectra of operators $\tilde{A}A$ and A^*A are expected to be tightly related, making the link explicit appears difficult in a general framework.

The regularity conditions on the eigenfunctions of A^*A are more restrictive than in Horowitz and Lee (2007), and Chen and Pouzo (2011) (see also the discussion of Assumption A.5 in Appendix 1). They are useful to establish pointwise asymptotic normality.

4.4. **Asymptotic normality.** Let us define the following quantities

$$\sigma_T^2(x) := \sum_{j=1}^{\infty} \frac{\nu_j}{(\lambda_T + \nu_j)^2} \phi_j^2(x) \text{ and } V_T(\lambda_T) := \frac{1}{T} \int \sigma_T^2(x) dx.$$

The following proposition computes the limit distribution of the Q-TiR estimator for (any sequence of) typical points. By (sequence of) typical points, we mean sets $\mathcal{X}_T \subset \mathcal{X} = [0,1]$ such that for $x \in \mathcal{X}_T$

$$\frac{V_T(\lambda_T)}{\sigma_T^2(x)/T} = O(1), \tag{4.7}$$

that is, the variance at a typical point x is not much smaller than the integrated variance.

Proposition 3: *Suppose Assumptions 1-4 and A.1-5 hold. Let h_T and λ_T be such that:* (a) $h_T \asymp T^{-\eta}$ *and* $\lambda_T \asymp T^{-\gamma}$ *with* $0 < \eta < \frac{1}{2(d_Z+2)}$, $0 < \gamma < \frac{1}{2} \min\left\{1 - \eta(d_Z+1), m\eta, \frac{1}{1+a}\right\}$; (b) $\lambda_T^{2\delta} = O(V_T(\lambda_T))$, $V_T(\lambda_T) = o(\lambda_T)$; (c) $h_T^{1/4} \frac{V_T(\lambda_T;2)}{V_T(\lambda_T)} = o(1)$ *and* $h_T^{2m} \frac{V_T(\lambda_T; 2m+\varepsilon_1)}{V_T(\lambda_T)} = o(1)$, *for some $\varepsilon_1 > 1$, where $V_T(\lambda_T;\varepsilon) := \frac{1}{T}\sum_{j=1}^{\infty} \frac{\nu_j}{(\lambda_T+\nu_j)^2} \|\phi_j\|^2 j^{\varepsilon}$. Then for any sequence $x \in \mathcal{X}_T$*

$$\sqrt{T/\sigma_T^2(x)}\left(\hat{\varphi}(x) - \varphi_0(x) - \mathcal{B}_T(x)\right) \xrightarrow{d} N(0,1).$$

Furthermore, if $\lambda_T^{2\delta} = o\left(V_T\left(\lambda_T\right)\right)$,

$$\sqrt{T/\sigma_T^2(x)}\left(\hat{\varphi}(x) - \varphi_0(x)\right) \xrightarrow{d} N(0,1).$$

The variance function $\sigma_T^2(x)$ involves the spectrum of operator A^*A and the regularization parameter λ_T. The penalty term $\lambda_T\|\varphi\|_H^2$ in the criterion defining the Q-TiR estimator implies that the inverse eigenvalues $1/\nu_j$ of the inverse of operator A^*A are ridged with $\nu_j/(\lambda_T + \nu_j)^2$. The variance formula is reminiscent of the usual asymptotic variance of the quantile regression estimator: it involves the factor $f_{V|X,Z}(0|x,z)$ (see (1.2)) through the Frechet derivative A, and the factor $\tau(1-\tau)$ through the adjoint A^*, hidden in the spectrum of A^*A. Since $\nu_j = \langle\phi_j, A^*A\phi_j\rangle_H = \|A\phi_j\|^2_{L^2(F_Z,\tau)} \leq \|A\|^2_{\mathcal{L}}\|\phi_j\|^2$, where $\|A\|_{\mathcal{L}}$ denotes the operator norm of A, we deduce that $\nu_j^{-1}\|\phi_j\|^2 \geq \|A\|_{\mathcal{L}}^{-2}$, which implies $\sum_{j=1}^{\infty}\nu_j^{-1}\|\phi_j\|^2 = \infty$. Then, by using (4.7), we get $\sigma_T^2(x) \to \infty$ as $\lambda_T \to 0$. Thus, $\sigma_T^2(x)$ summarizes the impact of ill-posedness on the nonparametric convergence rate $\sqrt{T/\sigma_T^2(x)}$. The conditions (a)-(c) on h_T and λ_T ensure that the asymptotic distribution of the nonlinear estimator $\hat{\varphi}$ is the same as the one induced by linearization. These conditions depend on the instrument dimension d_Z, the smoothness properties of the joint density of the observations via m, as well as on parameters δ and a introduced in Assumption 4. In particular, we need $\delta > 1/2 > a$. Since $\mathcal{B}_T(x) = O(\|\mathcal{B}_T\|_H) = O\left(\lambda_T^{\delta}\right)$ as shown in the proof of Proposition 3, $\lambda_T^{2\delta} = o(V_T\left(\lambda_T\right))$ is a sufficient condition for bias negligibility at a typical point. Below in Remark 3 we state an example to discuss the mutual compatibility of the conditions on h_T and λ_T for linearization and bias negligibility. Finally, we show in the supplementary materials that under a strengthening of Assumption A.5 (iii), the mean integrated square error (MISE) is asymptotically like $E[\|\hat{\varphi} - \varphi_0\|^2] = M_T(\lambda_T)\left(1 + o(1)\right)$, where $M_T(\lambda_T) := \int\left(\frac{1}{T}\sigma_T^2(x) + \mathcal{B}_T(x)^2\right)dx$.

4.5. **Estimating the asymptotic variance.** The estimation of the asymptotic variance of $\hat{\varphi}$ requires the estimation of the spectrum of operator A^*A. Let us assume the following semiparametric specification for the decay of the eigenvalues ν_j and eigenfunction values $\phi_j(x)$ at $x \in \mathcal{X}$ when $j \to \infty$.

Assumption 5: (i) *The eigenvalues are such that* $\nu_j = c_{1,j}\exp(v_j'\alpha)$ *and* (ii) *the eigenfunction values are such that* $\phi_j(x)^2 = c_{2,j}\exp(w_j\beta)\chi_j$, *where* $v_j' = -(j, \log j)$, $w_j = -\log j$, $\chi_j \geq 0$ *is an unknown periodic sequence with period* S, $\alpha = (\alpha_1, \alpha_2)'$ and β *are unknown parameters, and* $c_{1,j}, c_{2,j} > 0$ *are unknown sequences converging to strictly positive constants as* $j \to \infty$.

The specification in Assumption 5 (i) accommodates both geometric decay with $\alpha_2 = 0$ and hyperbolic decay with $\alpha_1 = 0$ in the eigenvalues (severe and mild ill-posedness for operator

A^*A). The specification in Assumption 5 (ii) accommodates both a slowly-varying trend component $\exp(w_j\beta)$ and an oscillatory component χ_j in the eigenfunction values. We can relax Assumption 5 to cover more general specifications of v_j and w_j.

Remark 2. *For the geometric case of the example in Remark 1, we have* $\tilde{\nu}_j = \tilde{c}_1 e^{-\alpha_1(j-1)}$, *with* $\tilde{c}_1, \alpha_1 > 0$. *For the Sobolev order* $l = 1$, *we also have* $\nu_j = \frac{c_1}{1+\pi^2(j-1)^2} e^{-\alpha_1(j-1)}$ *and* $\phi_1(x)^2 = 1$, $\phi_j(x)^2 = \frac{c_2}{1+\pi^2(j-1)^2}\cos^2(\pi(j-1)x)$, $j > 1$, *with* $c_1, c_2 > 0$. *Such an example is compatible with Assumption 5 where* $\alpha_2 = \beta = 2$ *and* $\chi_j = \cos^2(\pi(j-1)x)$, *if* x *is a rational number.*

Let $\bar{\nu}_j$ and $\bar{\phi}_j$, for $j \in \mathbb{N}$, be the eigenvalues and eigenfunctions of operator $\bar{A}^*\bar{A}$, normalized such that $\left\|\bar{\phi}_j\right\|_H = 1$, where $\bar{A} = D\hat{\mathcal{A}}\left(\bar{\varphi}\right)$ is based on a pilot estimator $\bar{\varphi}$. The pilot estimator $\bar{\varphi}$ is a Q-TiR estimator with regularization parameter $\bar{\lambda}_T$ and bandwidth $\bar{h}_T$ satisfying the assumptions of Proposition 3. Let $n_T \leq N_T$ be integers growing with T. Define $\hat{\nu}_j = \bar{\nu}_j$ for $1 \leq j \leq n_T$, and $\hat{\nu}_j = \bar{\nu}_{n_T}\exp((v_j - v_{n_T})'\hat{\alpha})$, for $n_T < j \leq N_T$, where the estimator $\hat{\alpha}$ is computed through Ordinary Least Squares (OLS) by regressing $\log\left(\bar{\nu}_j/\bar{\nu}_{n_T}\right)$ on $v_j - v_{n_T}$ for $j = n_T/2, \cdots, n_T - 1$. Similarly, define $\hat{\phi}_j(x)^2 = \bar{\phi}_j(x)^2$ for $1 \leq j \leq n_T$, and $\hat{\phi}_j(x)^2 = \bar{\phi}_{S,n_T}(x)^2\exp((w_j - w_{n_T})\hat{\beta})\hat{\chi}_{j\text{mod}S}$, for $n_T < j \leq N_T$. Here, $\bar{\phi}_{S,j}(x)^2 := \sum_{k=0}^{S-1}\bar{\phi}_{j-k}(x)^2$, the estimator $\hat{\beta}$ is computed through OLS by regressing $\log\left(\bar{\phi}_{S,j}(x)^2/\bar{\phi}_{S,n_T}(x)^2\right)$ on $w_j - w_{n_T}$ for $j = n_T/2, \cdots, n_T - 1$, and $\hat{\chi}_j = \frac{2S}{n_T}\sum_{k: n_{T/2} \leq k < n_T, k = j\text{mod}S}\bar{\phi}_k(x)^2/\bar{\phi}_{S,k}(x)^2$, for $j = 1, ..., S$, where $k = j\text{mod}S$ if $k - j$ is an integer multiple of S. The estimator of the variance function is $\hat{\sigma}_T^2(x) = \sum_{j=1}^{N_T}\frac{\hat{\nu}_j}{(\hat{\nu}_j + \lambda_T)^2}\hat{\phi}_j(x)^2$.

We have introduced the parameter n_T since the nonparametric estimates of the spectrum of A^*A may be unprecise in relative terms for j close to the truncation parameter N_T. The nonparametric estimates are replaced by extrapolated estimates between n_T and N_T. The extrapolation procedure exploits the supposed decay behaviour of the spectrum in Assumption 5. For the eigenfunction values, we estimate and extrapolate the parametric trend component by using that the filtered spectral coefficients $\phi_{S,j}(x)^2 := \sum_{k=0}^{S-1}\phi_{j-k}(x)^2$ approaches $\exp(w_j\beta)$ for $j \to \infty$, up to a scale constant. Then, we estimate the periodic component by averaging the detrended square eigenfunction values over all lags j with same phase of the cycle.

Proposition 4: *Denote* $\Delta\nu_j := \min_{k \leq j}\left(\nu_{k-1} - \nu_k\right)$. *Let* N_T *be such that*

$$\sum_{j=N_T+1}^{\infty}\nu_j\left\|\phi_j\right\|^2 = o\left(T\lambda_T^2 V_T\left(\lambda_T\right)\right), \tag{4.8}$$

and n_T *such that* $N_T = O(n_T)$ *and*

$$\frac{1}{\exp(w_{n_T}\beta)\Delta\nu_{n_T}}\left(\frac{1}{Th_T^2} + h_T^{2m} + M_T\left(\bar{\lambda}_T\right)\right)^{1/2} = O_p\left(T^{-b}\right), \tag{4.9}$$

for $b > 0$. *Assume that* $\sigma^2_{*,T}(x)/\sigma^2_T(x) = O(1)$, *where* $\sigma^2_{*,T}(x) = \sum_{j=1}^{\infty} \frac{\nu_j}{(\nu_j+\lambda_T)^2} c_{2,j} \exp(w_j\beta)$. *Then, under Assumptions 1-5 and A.1-5,* $\frac{\sigma^2_T(x)}{\hat{\sigma}^2_T(x)} \xrightarrow{p} 1$.

Condition (4.8) ensures that the truncation bias is negligible. Condition (4.9) ensures that $\bar{\nu}_j$ and $\bar{\phi}_j(x)^2$ are consistent in relative terms uniformly over $1 \le j \le n_T$. It involves the asymptotic MISE $M_T(\bar{\lambda}_T)$ of the pilot estimator. The condition $N_T = O(n_T)$ and the extrapolation procedure yield relative consistency of $\hat{\nu}_j$ and $\hat{\phi}_j(x)^2$ uniformly over $1 \le j \le N_T$.

Remark 3. *Let the eigenvalues and eigenfunctions satisfy Assumption 5 with* $\alpha_1, \beta > 0$, $\alpha_2 \ge 0$ *and* $\|\phi_j\|^2 \asymp j^{-\beta}$ *(see Remarks 1 and 2). Let us verify that there exist admissible values of* γ, η, $\bar{\gamma}$, $\bar{\eta}$, *so that the conditions on* λ_T, h_T, $\bar{\lambda}_T$, $\bar{h}_T$, n_T, N_T, *to get Propositions 3 and 4 are all compatible. In the supplementary materials we prove that* $V_T(\lambda_T;\varepsilon) \asymp \frac{1}{T\lambda_T[\log(1/\lambda_T)]^{\beta-\varepsilon}}$ *for* $\varepsilon \ge 0$, *and* $\sigma^2_T(x), \sigma^2_{*,T}(x) \asymp \frac{1}{\lambda_T[\log(1/\lambda_T)]^{\beta}}$. *If* $\frac{2}{m(1+2\delta)} < \bar{\eta} < \frac{1}{d_Z+1}\frac{2\delta-1}{2\delta+1}$, $\frac{1}{1+2\delta} < \bar{\gamma} < \frac{1}{2}\min\left\{1-(d_Z+1)\bar{\eta}, m\bar{\eta}, \frac{1}{1+a}\right\}$, *the pilot estimator satisfies the assumptions of Proposition 3. If* $\frac{2}{m(1+2\delta)} < \eta < \frac{1}{d_Z+1}\frac{2\delta-1}{2\delta+1}$, $\frac{1}{1+2\delta} < \gamma < \frac{1}{2}\min\left\{1-(d_Z+1)\eta, m\eta, \frac{1}{1+a}\right\}$, *estimator* $\hat{\varphi}$ *is asymptotically normal with vanishing bias. The conditions on* γ, η, $\bar{\gamma}$, $\bar{\eta}$ *are compatible if* $m > 2(d_Z+1)$ *and* $\delta > \frac{1}{2} + \max\left\{\frac{d_Z+1}{m}, a\right\}$. *Since* $\sum_{j=n+1}^{\infty} \nu_j \|\phi_j\|^2 \le Ce^{-\alpha_1 n}$, $n \in \mathbb{N}$, *and* $\Delta\nu_j = \nu_{j-1} - \nu_j \asymp j^{-\alpha_2}e^{-\alpha_1 j}$, *conditions (4.8) and (4.9) are satisfied for* $n_T = c_1 \log T$ *and* $N_T = c_2 \log T$ *such that* $c_1 < \frac{1}{\alpha_1}\min\left\{\frac{1-2\eta}{2}, m\eta, \frac{1-\bar{\gamma}}{2}, \delta\bar{\gamma}\right\}$ *and* $c_2 > \frac{1}{\alpha_1}\gamma$. *Then* $\sqrt{T/\hat{\sigma}^2_T(x)}\,(\hat{\varphi}(x) - \varphi_0(x)) \xrightarrow{d} N(0,1)$, *from which asymptotically valid pointwise confidence intervals can be computed.*

5. Monte-Carlo results and empirical illustration

5.1. **Monte-Carlo results.** We consider an experiment, following Newey and Powell (2003), where the errors U_1^*, U_2^* and the instrument Z follow a trivariate normal distribution, with zero means, unit variances and a correlation coefficient of 0.5 between U_1^* and U_2^*. We take $X^* = Z + U_2^*$, and build $X = \Phi(X^*)$, $Y = \sin(\pi X) + U_1^*$, and $U = \Phi(U_1^*)$. The median condition is $P[Y - \varphi_0(X) \le 0 \mid Z] = .5$, where the functional parameter is $\varphi_0(x) = \sin(\pi x)$, $x \in [0,1]$. We consider sample size $T = 1000$ and a classical Sobolev penalty of order $l \in \{1,2\}$. For $l = 1$ we consider $\bar{\lambda} = \lambda \in \{.00002, .00005, .0001\}$, while for $l = 2$ we consider $\bar{\lambda} = \lambda \in \{.0000003, .000001, .000003\}$. We use Gaussian kernels and select the bandwidth with the standard rule of thumb (Silverman (1986)). To compute $\hat{\varphi}$ and $\bar{\varphi}$ we opt for a numerical series approximation based on standardized shifted Chebyshev polynomials of the first kind, and a user-supplied analytical gradient and Hessian optimization procedure. We report results using 16 polynomials (order 0 to 15); results using the first 8, or more, polynomials are nearly identical. The variance estimator $\hat{\sigma}^2_T(x)$ is computed with $N_T = n_T = 4$. Figure 1 displays the QQ-plots of the finite sample distributions of $\sqrt{T/\hat{\sigma}_T^2(x)}\,(\hat{\varphi}(x)-\varphi_0(x))$, $x=.5$, for $l=1$, $\bar{\lambda}=\lambda=.00005$ (left), and $l=2$, $\bar{\lambda}=\lambda=.000001$ (right), built on 1000 replications. The finite sample distributions are close to the standard normal distribution. For the selected values of $\bar{\lambda},\lambda$, the regularization bias is rather small. Table I reports the finite sample coverage of pointwise confidence intervals for $\varphi_0(x)$, $x\in\{.10,.25,.50,.75,.90\}$ using the different values of l, $\bar{\lambda}$, λ. Let us first consider the results for $l=1$ (left panel). The finite sample coverage is close to the nominal coverage at 90%, 95%, 99% for $x=.5$ and $\bar{\lambda}=\lambda=.00005$. At $x=.5$ we observe overrejection for $\bar{\lambda}=\lambda=.0001$ because of regularization bias, and underrejection for $\bar{\lambda}=\lambda=.00002$ because of a too small number of terms in the estimated variance. For instance, with $N_T=n_T=5$ and at $x=.5$, the finite sample coverage at 90%, 95%, 99% is .964, .985, .998 for $l=1$, $\bar{\lambda}=\lambda=.00002$, and .947, .974, .999 for $l=2$, $\bar{\lambda}=\lambda=.0000003$. For $x=.10,.25,.75,.90$ and the considered values of $\bar{\lambda},\lambda$, we typically observe some overrejection. Finally, the results for $l=2$ are qualitatively similar to those for $l=1$.

5.2. **Empirical illustration.** This section presents an empirical illustration with U.S. long-distance call data extracted from the sample of Hausman and Sidak (2004). They investigate nonlinear price schedules chosen by consumers of message toll service offered by long-distance interexchange carriers. We estimate median structural effects for nonlinear pricing curves based on the conditional quantile condition $P\left[Y\leq\varphi_0\left(X\right)\mid Z\right]=.5$, with $X=\Phi\left(min\right)$ and $Z=\Phi\left(Inc\right)$. Variable Y is the price per minute in dollars, min is the standardized amount of use in minutes, and Inc is the standardized logarithm of annual income. We look at clients of a leading long-distance interexchange carrier with age between 30 and 45. To study the effect of education on the chosen nonlinear price schedule we divide the sample into people with at most 12 years of education ($T=978$), and people with more than 12 years of education ($T=435$).

We set $\bar{\lambda}=.0001$, and start the optimization algorithm with the NIVR estimates of GS. The specification test of Gagliardini and Scaillet (2007) does not reject the null hypothesis of the correct specification of the moment restriction used in estimating the mean pricing curve at the 5% significance level (p-value $=.32,.77$). We apply a reduction factor .8 to the regularization parameters selected by the heuristic data-driven procedure of GS run on the linearization. For the two education categories, we get $\lambda=.154,.034$ under a Sobolev penalty with $l=1$, and $\lambda=.088,.001$ under a Sobolev penalty with $l=2$. We present the empirical results with sixteen polynomials. They remain virtually unchanged when increasing gradually the number of polynomials from eight to sixteen. There is a stabilization of the value of the optimized objective function, of the loadings in the numerical series approximation, and of the data-driven regularization parameter. Higher order polynomials receive loadings which are closer and closer to zero. This suggests that we can limit ourselves to a small number of polynomials in this application.

Figure 2 plots the estimated NIV median structural effect, pointwise asymptotic confidence intervals at 95%, and the linear IVQR estimate for the two education categories. The upper panel for $l = 1$ and the lower panel for $l = 2$ show that estimated NIVQR and IVQR structural effects are close, and their patterns differ across the two education categories. As in Hausman and Sidak (2004) we observe that less educated customers pay more than better educated customers when the number of minutes of use increases. A possible explanation is that the latter exploit better the tariff options for long-distance calls available at those ranges.

Figure 3 is a picture "à la box-plot" where we represent the estimated quartile structural effects and the estimated mean structural effect with $l = 1$. The box-plot interpretation comes from the conditional probability of the shaded area being asymptotically $P\left[g\left(X, 0.25\right) \leq Y \leq g\left(X, 0.75\right) | Z = z\right] = .5$, for any given value z of the instrument. Vertical sections of the shaded areas correspond to measures of dispersion. For both education categories, the dispersion is smaller at high usage of the service, likely because of the effort of people to find the most convenient tariffs.

References

[1] Adams, R., and J. Fournier (2003): Sobolev Spaces. Boston: Academic Press.

[2] Ai, C., and X. Chen (2003): Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions, Econometrica, 71, 1795-1843.

[3] Alt, H. (1992): Linear Functional Analysis. Berlin: Springer Verlag.

[4] Andrews, D. (1994): Empirical Process Methods in Econometrics, in the Handbook of Econometrics, Vol. 4, ed. by R. Engle and D. McFadden. Amsterdam: North Holland, 2247-2294.

[5] Blundell, R., X. Chen, and D. Kristensen (2007): Semi-Nonparametric IV Estimation of Shape Invariant Engel Curves, Econometrica, 75, 1613-1669.

[6] Bosq, D. (1998): Nonparametric Statistics for Stochastic Processes. Berlin: Springer Verlag.

[7] Bosq, D. (2000): Linear Processes in Function Spaces. Theory and Applications. Berlin: Springer Verlag.

[8] Carrasco, M., J.-P. Florens, and E. Renault (2007): Linear Inverse Problems in Structural Econometrics: Estimation Based on Spectral Decomposition and Regularization, in the Handbook of Econometrics, Vol. 6, ed. by J. Heckman and E. Leamer. Amsterdam: North Holland, 5633-5751.

[9] Chen, X., and S. Ludvigson (2009): Land of Addicts? An Empirical Investigation of Habit-Based Asset Pricing Models, Journal of Applied Econometrics, 24, 1057-1093.

[10] Chen, X., and D. Pouzo (2009): Efficient Estimation of Semiparametric Conditional Moment Models with Possibly Nonsmooth Residuals, Journal of Econometrics, 152, 46-60.

[11] Chen, X., and D. Pouzo (2011): Estimation of Nonparametric Conditional Moment Models with Possibly Nonsmooth Moments, forthcoming in Econometrica.

[12] Chernozhukov, V., and C. Hansen (2005): An IV Model of Quantile Treatment Effect, Econometrica, 73, 245-261.

[13] Chernozhukov, V., G. Imbens, and W. Newey (2007): Instrumental Variable Estimation of Nonseparable Models, Journal of Econometrics, 139, 4-14.

[14] Darolles, S., Y. Fan, J.-P. Florens, and E. Renault (2011): Nonparametric Instrumental Regression, forthcoming in Econometrica.

[15] Dubinskij, J. (1986): Sobolev Spaces of Infinite Order and Differential Equations. Dordrecht: Kluwer academic publishers.

[16] Egger, H., and H. Engl (2005): Tikhonov Regularization Applied to the Inverse Problem of Option Pricing: Convergence Analysis and Rates, Inverse Problems, 21, 1027-1045.

[17] Engl, H., M. Hanke, and A. Neubauer (2000): Regularization of Inverse Problems. Dordrecht: Kluwer academic publishers.

[18] Engl, H., K. Kunisch, and A. Neubauer (1989): Convergence Rates for Tikhonov Regularization of Nonlinear Ill-posed Problems, Inverse problems, 5, 523-540.

[19] Gagliardini, P., and O. Scaillet (2007): A Specification Test for Nonparametric Instrumental Variable Regression, Swiss Finance Institute Discussion Paper 2007.13.

[20] Gagliardini, P., and O. Scaillet (2011a): Tikhonov Regularization for Nonparametric Instrumental Variable Estimators, forthcoming in Journal of Econometrics.

[21] Gagliardini, P., and O. Scaillet (2011b): "Nonparametric Instrumental Variable Estimation of Structural Quantile Effects, Supplementary material: Proofs and Example", Econometrica Supplementary Material.

[22] Groetsch, C. (1984): The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind. Boston: Pitman Advanced Publishing Program.

[23] Hall, P., and J. Horowitz (2005): Nonparametric Methods for Inference in the Presence of Instrumental Variables, Annals of Statistics, 33, 2904-2929.

[24] Hall, P., and M. Hosseini-Nasab (2006): On Properties of Functional Principal Components Analysis, Journal of the Royal Statistical Society, Series B, 68, 109-126.

[25] Hansen, B. (2008): Uniform Convergence Rates for Kernel Estimation with Dependent Data, Econometric Theory, 24, 726-748.

[26] Hausman, J., and G. Sidak (2004): Why Do the Poor and the Less-Educated Pay More for Long-Distance Calls?, Contributions to Economic Analysis and Policy, 3, Article 3.

[27] Hofmann, B., and O. Scherzer (1998): Local Ill-posedness and Source Conditions of Operator Equations in Hilbert Spaces, Inverse Problems, 14, 1189-1206.

[28] Horowitz, J. (2007): Asymptotic Normality of a Nonparametric Instrumental Variables Estimator, International Economic Review, 48, 1329-1349.

[29] Horowitz, J., and S. Lee (2007): Nonparametric Instrumental Variables Estimation of a Quantile Regression Model, Econometrica, 75, 1191-1208.

[30] Koenker, R. (2005): Quantile Regression, Econometric Society Monographs, 38. Cambridge: Cambridge University Press.

[31] Kress, R. (1999): Linear Integral Equations. New York: Springer.

[32] Newey, W., and J. Powell (2003): Instrumental Variable Estimation of Nonparametric Models, Econometrica, 71, 1565-1578.

[33] Reed, M., and B. Simon (1980): Functional Analysis. San Diego: Academic Press.

[34] Schock, E. (2002): Non-linear Ill-posed Equations: Counter-examples, Inverse problems, 18, 715-717.

[35] Silverman, B. (1986): Density Estimation for Statistics and Data Analysis. London: Chapman and Hall.

[36] Tikhonov, A. N. (1963a): On the Solution of Incorrectly Formulated Problems and the Regularization Method, Soviet Math. Doklady, 4, 1035-1038 (English Translation).

[37] Tikhonov, A. N. (1963b): Regularization of Incorrectly Posed Problems, Soviet Math. Doklady, 4, 1624-1627 (English Translation).

[38] White, H., and J. Wooldridge (1991): Some Results on Sieve Estimation with Dependent Observations, in Nonparametric and Semiparametric Methods in Econometrics and Statistics, ed. by W. Barnett, J. Powell and G. Tauchen. New York: Cambridge University Press, 459-493.

Appendix A. Regularity Conditions and Proofs

A.1. **Assumptions A.** Below we list the additional technical regularity conditions. In particular, we invoke A.1-3 for proofs of local ill-posedness and consistency and A.4-5 for asymptotic normality. For a function f of variable s in $\mathbb{R}^{d_s}$ and a multi-index $\alpha \in \mathbb{N}^{d_s}$, we denote $\nabla^\alpha f := \nabla_{s_1}^{\alpha_1} \cdots \nabla_{s_{d_s}}^{\alpha_{d_s}} f$, $|\alpha| := \sum_{i=1}^{d_s} \alpha_i$, $\|f\|_\infty := \sup_s |f(s)|$ and $\|Df\|_\infty := \sum_{\alpha:|\alpha|=1} \|\nabla^\alpha f\|_\infty$.

A.1: (i) $\{(X_l, Y_l, Z_l^*) : l = 1, ..., T^*\}$ *is a sample of i.i.d. observations of random variable* (X, Y, Z^*) *admitting a density* f_{X,Y,Z^*} *on the support* $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}^* \subset \mathbb{R}^d$, *where* $\mathcal{X} = [0,1]$, $\mathcal{Y} = [0,1]$, $\mathcal{Z}^* \subset \mathbb{R}^{d_Z}$, $d = 2 + d_Z$. (ii) *The density* f_{X,Y,Z^*} *is in class* $C^m(\mathbb{R}^d)$, *with* $m \geq 2$, *and* $\nabla^\alpha f_{X,Y,Z^*}$ *is uniformly continuous and bounded, for any* $\alpha \in \mathbb{N}^d$ *with* $|\alpha| = m$. (iii) *The random variable* (X, Y, Z) *is such that* $(X, Y, Z) = (X, Y, Z^*)$ *if* $Z^* \in \mathcal{Z}$, *where* $\mathcal{Z} = [0,1]^{d_Z}$ *is interior to* $\mathcal{Z}^*$, *and the density* f_Z *of* Z *is such that* $\inf_{z\in\mathcal{Z}} f_Z(z) > 0$.

A.2: *The kernel* K *on* $\mathbb{R}^d$ *is such that* (i) $\int K(u)du = 1$ *and* K *is bounded*; (ii) K *has compact support;* (iii) K *is differentiable, with bounded derivatives*; (iv) $\int u^\alpha K(u)du = 0$ *for any* $\alpha \in \mathbb{N}^d$ *with* $|\alpha| < m$.

A.3: (i) *The function* $\tau \mapsto g(x, \tau)$ *is strictly monotonic increasing and continuous, for almost any* $x \in (0,1)$, *and* $\sup_{x,\tau} |g(x,\tau)| < \infty$, $\sup_{x,\tau} |\nabla_x g(x,\tau)| < \infty$; (ii) $\|Df_{X|Z}\|_\infty < \infty$; (iii) $\|DF_{U|X,Z}\|_\infty < \infty$.

A.4: (i) *There exists* $h > 0$ *such that function* $q(s) := \sum_{\alpha:|\alpha|\leq m} \sup_{v\in B_h(s)} |\nabla^\alpha f_{X,Y,Z}(v)|$, $s \in \mathcal{S} := \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, *is integrable and satisfies* $\int_{\mathcal{S}} \frac{q(s)^2}{f_{X,Y,Z}(s)} ds < \infty$, *where* $B_h(s)$ *denotes the ball in* $\mathbb{R}^d$ *of radius* h *centered at* s; (ii) $\|f_{X,Y|Z}\|_\infty < \infty$; (iii) $\|\nabla_y f_{X,Y|Z}\|_\infty < \infty$.

A.5: (i) $\sum_{j,k=1,j\neq k}^\infty \frac{\langle\phi_j,\phi_k\rangle^2}{\|\phi_j\|^2\|\phi_k\|^2} < \infty$; (ii) *The functions* $\psi_j(z) := \frac{1}{\sqrt{\nu_j}} (A\phi_j)(z)$, $j \in \mathbb{N}$, *satisfy* $\sup_{j\in\mathbb{N}} E\left[|\psi_j(Z)|^{\bar{s}}\right]^{1/\bar{s}} < \infty$, *for* $\bar{s} \geq 4$; (iii) *The functions* ψ_j *are in class* $C^m(\mathbb{R}^{d_Z})$ *such that* $E\left[|\nabla^\alpha \psi_j(Z)|^{\bar{s}}\right]^{1/\bar{s}} = O\left(j^{|\alpha|}\right)$ *for any* $\alpha \in \mathbb{N}^{d_Z}$ *with* $|\alpha| \leq m$.

In Assumption A.1 (i), the compact support of X and Y is used for technical reasons. Assuming univariate X simplifies the exposition. Assumptions A.1 (ii) and A.2 are classical conditions in kernel density estimation concerning smoothness of the density and of the kernel.

In particular, when $m > 2$, K is a higher order kernel. Moreover, we assume a compact support for the kernel K to simplify the set of regularity conditions. In Assumption A.1 (iii), variable Z is obtained by truncating Z^* on the compact set $\mathcal{Z}$, and the density f_Z of Z is bounded from below away from 0 on the support $\mathcal{Z}$. The corresponding observations are Z_t, $t = 1, ..., T$, where $T \leq T^*$. We get the estimator $\hat{f}_{X,Y,Z}$ of the density $f_{X,Y,Z}$ from the kernel estimator $\hat{f}_{Y,X,Z^*}(x, y, z) = \frac{1}{T^* h_T^d} \sum_{l=1}^{T^*} K\left((X_l - x)/h_T\right) K\left((Y_l - y)/h_T\right) K\left((Z_l^* - z)/h_T\right)$ of density f_{X,Y,Z^*} by normalization, namely $\hat{f}_{Y,X,Z} = \hat{f}_{Y,X,Z^*}/\int_{\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}} \hat{f}_{Y,X,Z^*} = \hat{f}_{Y,X,Z^*}/\int_{\mathcal{Z}} \hat{f}_{Z^*}$. This trick is used in the proofs to control for small values of the estimator of the density of Z appearing in denominators and to avoid edge effects. Alternative approaches to address these technical issues consist in using trimming (see e.g. Hansen (2008)), boundary kernels (see e.g. Hall and Horowitz (2005) for the use of such kernels in NIVR) or density weighting (see e.g. Horowitz and Lee (2007)). Assumption A.3 (i) is a boundedness and smoothness condition on function $g(x, \tau)$ w.r.t. both its arguments. Assumptions A.3 (ii) and (iii) concern boundedness and smoothness of the p.d.f. of X given Z, and the c.d.f. of U given X, Z, respectively.

Assumption A.4 concerns the joint density $f_{X,Y,Z}$ and the conditional density $f_{X,Y|Z}$. Specifically, Assumption A.4 (i) imposes an integrability condition on a suitable measure of local variation of density $f_{X,Y,Z}$ and its derivatives. This assumption is used in the proof of Lemma A.10 to bound higher order terms in the asymptotic expansion of the estimator coming from kernel estimation bias. Assumptions A.4 (ii)-(iii) are used to show that $\mathcal{A}$ is Frechet differentiable, with compact Frechet derivative. These assumptions can be rewritten in terms of densities $f_{U|X,Z}$, $f_{X|Z}$ and function g. The formulation as in Assumptions A.4 (ii)-(iii) is closer to the use in the proofs, and simplifies the exposition. Assumption A.4 (iii) also implies a Lipschitz behaviour of the Frechet derivative operator $D\mathcal{A}(\varphi)$ w.r.t. φ in a neighborhood of the true function (Assumption ii) in Theorem 10.4 of Engl, Hanke and Neubauer (2000)). Finally, Assumption A.5 concerns the singular system $\left\{\sqrt{\nu_j}, \phi_j, \psi_j; j \in \mathbb{N}\right\}$ of operator A (Kress (1999), p. 278). Assumption A.5 (i) requires that the $\langle ., . \rangle_H$-orthonormal basis of eigenfunctions of A^*A satisfy a summability condition w.r.t. $\langle ., . \rangle$. This assumption eases the derivation of the upper bound of the MISE in Lemma A.7. Assumptions A.5 (ii) and (iii) ask for the existence of bounds for moments of derivatives of functions ψ_j, $j \in \mathbb{N}$. Functions ψ_j, $j \in \mathbb{N}$, are an orthonormal system in $L^2(F_Z, \tau)$. These assumptions control for terms of the type $\int \psi_j(z) \left[1\left\{y \leq \varphi_0(x)\right\} - \tau\right] \hat{f}_{X,Y,Z}(s) ds$, uniformly in $j \in \mathbb{N}$, in the proof of Lemmas A.7, A.8 and A.11.

A.2. **Proof of Proposition 1.** In Step 1 we show local ill-posedness of the nonseparable setting (part (a)), and in Step 2 we prove that sequences generating ill-posedness exhibit diverging L^2-norm of their first derivative (part (b)). We place all omitted proofs to the supplementary materials.

Step 1. (Proof of part (a)) We use the next Lemma A.1, which is a local version of Proposition 10.1 in Engl, Hanke and Neubauer (2000).

Lemma A.1. *Suppose that: (i) Operator $\mathcal{A}$ is compact. (ii) For any $r > 0$ small enough, there exists a sequence $(\varphi_n) \subset B_r(\varphi_0)$ s.t. $\varphi_n \nrightarrow \varphi_0$ and $\mathcal{A}(\varphi_n) \overset{w}{\to} \mathcal{A}(\varphi_0)$ where $\overset{w}{\to}$ denotes weak convergence. Then the minimum distance problem is locally ill-posed.*

Condition (i) in Lemma A.1 follows from the next Lemma A.2, which is proved using a result in Alt (1992).

Lemma A.2. *Under Assumptions A.3 (ii) and (iii), operator $\mathcal{A}$ is compact.*

Let us now verify Condition (ii) in Lemma A.1. Define $\psi(x) := \sin(2\pi x)$ and $\psi_n(x) := \varepsilon\psi(nx)$, $x \in \mathcal{X}$, where $0 < \varepsilon < \min\{\tau, 1-\tau\}$. Further, let $\varphi_n(x) := g(x, \tau + \psi_n(x))$, $x \in \mathcal{X}$. Then, we deduce that $\|\varphi_n - \varphi_0\|^2 = \int_{\mathcal{X}} [g(x, \tau + \varepsilon\psi(nx)) - g(x,\tau)]^2 dx$. Split the integral w.r.t. x over the partition $((k-1)/n, k/n]$, $k = 1, ..., n$, of $(0,1]$. It follows that

$$\begin{aligned}\|\varphi_n - \varphi_0\|^2 &= \sum_{k=1}^{n} \int_{(k-1)/n}^{k/n} [g(x, \tau + \varepsilon\psi(nx)) - g(x,\tau)]^2 dx \\ &= \sum_{k=1}^{n} \frac{1}{n} \int_0^1 \left[g\left(\frac{k-1}{n} + \frac{y}{n}, \tau + \varepsilon\psi(y)\right) - g\left(\frac{k-1}{n} + \frac{y}{n}, \tau\right)\right]^2 dy,\end{aligned}$$

using the periodicity of ψ. Using Assumption A.3 (i),

$$\|\varphi_n - \varphi_0\|^2 = \sum_{k=1}^{n} \frac{1}{n} \int_0^1 \left[g\left(\frac{k-1}{n}, \tau + \varepsilon\psi(y)\right) - g\left(\frac{k-1}{n}, \tau\right)\right]^2 dy + O(1/n).$$

The first term is a converging Riemann sum, and we get $\|\varphi_n - \varphi_0\|^2 \to I_\varepsilon := \int_{\mathcal{X}} \int_0^1 [g(x, \tau + \varepsilon\psi(y)) - g(x,\tau)]^2 dy dx$ as $n \to \infty$. Then, $I_\varepsilon > 0$, and $I_\varepsilon \to 0$ as $\varepsilon \to 0$ by the dominated convergence Theorem. Thus, for $\varepsilon > 0$ sufficiently small, we have $(\varphi_n) \subset B_r(\varphi_0)$ and $\varphi_n \nrightarrow \varphi_0$. For $\bar{q} \in L^2(F_Z, \tau)$ we have,

$$\langle \bar{q}, \mathcal{A}(\varphi_n)\rangle_{L^2(F_Z,\tau)} = \frac{1}{\tau(1-\tau)} \int \bar{q}(z) f_Z(z) \int_{\mathcal{X}} f_{X|Z}(x|z) F_{U|X,Z}(\tau + \psi_n(x)|x,z)\, dxdz.$$

Thus, we have to show

$$J_n := \int \bar{q}(z) f_Z(z) \int_{\mathcal{X}} f_{X|Z}(x|z) F_{U|X,Z}(\tau + \psi_n(x)|x,z)\, dxdz \to \tau \int \bar{q}(z) f_Z(z) dz. \tag{A.1}$$

To this end, in J_n we split the integral w.r.t. x over the partition $((k-1)/n, k/n]$ with $k = 1, ..., n$ and get

$$J_n = \sum_{k=1}^{n} \frac{1}{n} \int_0^1 \int \bar{q}(z) f_Z(z) f_{X|Z}\left(\frac{k-1}{n} + \frac{1}{n} y | z\right) F_{U|X,Z}\left(\tau + \varepsilon\psi(y) | \frac{k-1}{n} + \frac{1}{n} y, z\right) dzdy$$

after a change of variable and using periodicity of function ψ. Then we have

$$J_n = \sum_{k=1}^{n} \frac{1}{n} \int \bar{q}(z) f_Z(z) f_{X|Z}\left(\frac{k-1}{n}|z\right) \int_0^1 F_{U|X,Z}\left(\tau + \varepsilon\psi(y) | \frac{k-1}{n}, z\right) dydz + I_{1,n}, \quad (A.2)$$

where $|I_{1,n}| \leq \sum_{k=1}^{n} \frac{1}{n^2} \int_0^1 \int \bar{q}(z) f_Z(z) \sup_{u,x,z} |\nabla_x H(u,x|z)| \, ydzdy = O(1/n)$ and $H(u,x|z) := F_{U|X,Z}(u|x,z) f_{X|Z}(x|z)$, with $\sup_{u,x,z} |\nabla_x H(u,x|z)| < \infty$ from Assumptions A.3 (ii)-(iii). Since the Riemann sum in (A.2) converges to the corresponding integral, we get

$$J_n \to \int_{\mathcal{X}} \int \bar{q}(z) f_Z(z) f_{X|Z}(x|z) \int_0^1 F_{U|X,Z}(\tau + \varepsilon\psi(y) | x, z) \, dydzdx =: J.$$

Using that $\int_{\mathcal{X}} f_{X|Z}(x|z) F_{U|X,Z}(u|x,z) dx = P[U \leq u | Z = z] = u$ by the independence of U and Z, and the uniform distribution of U, we get $J = \tau \int \bar{q}(z) f_Z(z) dz + \varepsilon \int \bar{q}(z) f_Z(z) \int_0^1 \psi(y) \, dy = \tau \int \bar{q}(z) f_Z(z) dz$, and (A.1) follows.

Step 2. (Proof of part (b)) The proof is by contradiction. Suppose that there exists $B < \infty$ such that $\|\nabla\varphi_n\| \leq B$ for any n large enough. Since Θ is bounded, and by the compact embedding theorem (see Adams and Fournier (2003)), set $\{\varphi \in \Theta : \|\nabla\varphi\| \leq B\}$ is compact w.r.t. the norm $\|.\|$. Therefore, there exists a subsequence (φ_{m_n}) which converges in norm $\|.\|$ to $\varphi^* \in \Theta$, say. Since Q_∞ is continuous, we get $Q_\infty(\varphi_{m_n}) \to Q_\infty(\varphi^*)$, and thus $Q_\infty(\varphi^*) = 0$. By identification (Assumption 1 (i)), we deduce $\varphi^* = \varphi_0$, and the subsequence (φ_{m_n}) converges to φ_0. But this is impossible, since $\|\varphi_{m_n} - \varphi_0\| \geq \varepsilon > 0$.

A.3. **Proof of Proposition 2.** We establish existence of the Q-TiR estimator in the supplementary materials by using a result in Reed, Simon (1980). To prove consistency, the next Lemma A.3 establishes (uniform) convergence of the minimum distance criterion $Q_T(\varphi)$ by using results in Andrews (1994), Bosq (1998) and Hansen (2008).

Lemma A.3. *Under Assumptions A.1, A.2, A.3 (ii)-(iii): (i)* $Q_T(\varphi_0) - Q_\infty(\varphi_0) = O_p(a_T)$, *where* $a_T := \dfrac{\log T}{T h_T^{d_Z+1}} + h_T^{2m}$; *(ii)* $\sup_{\varphi\in\Theta} |Q_T(\varphi) - Q_\infty(\varphi)| = O_p\left(\sqrt{a_T} + \frac{1}{\sqrt{T}}\right) = o_p(1)$ *for* $a_T = o(1)$.

By Lemma A.3 (i) and the condition on λ_T, we have

$$0 \leq Q_T(\hat{\varphi}) + \lambda_T \|\hat{\varphi}\|_H^2 \leq Q_T(\varphi_0) + \lambda_T \|\varphi_0\|_H^2 = O_p(a_T + \lambda_T) = O_p(\lambda_T). \quad (A.3)$$

By $Q_T \geq 0$, this implies that $\lambda_T \|\hat{\varphi}\|_H^2 = O_p(\lambda_T)$, that is, $\|\hat{\varphi}\|_H^2 = O_p(1)$. Thus, by the compact embedding theorem, the sequence of minimizers $\hat{\varphi}$ is tight in $(L^2[0,1], \|\cdot\|)$. Namely, for any $\delta > 0$, there exists a compact subset K_δ of $(L^2[0,1] \cap \Theta, \|\cdot\|)$, such that $P[\hat{\varphi} \in K_\delta] \geq 1 - \delta$ for all sufficiently large sample sizes.

Next we have that for any $\varepsilon > 0$ and $\delta > 0$, and any T sufficiently large,

$$\begin{aligned} P[\hat{\varphi} \notin B_\varepsilon(\varphi_0)] &\leq P[\{\hat{\varphi} \notin B_\varepsilon(\varphi_0)\} \cap \{\hat{\varphi} \in K_\delta\}] + P[\hat{\varphi} \notin K_\delta] \\ &\leq P[\{\hat{\varphi} \notin B_\varepsilon(\varphi_0)\} \cap \{\hat{\varphi} \in K_\delta\}] + \delta \\ &\leq P[\inf_{\varphi \in K_\delta \cap \Theta \setminus B_\varepsilon(\varphi_0)} Q_T(\varphi) + \lambda_T \|\varphi\|_H^2 \leq Q_T(\hat{\varphi}) + \lambda_T \|\hat{\varphi}\|_H^2] + \delta. \end{aligned}$$

Using bound (A.3) and Lemma A.3 (ii), we get:

$$\begin{aligned} P[\hat{\varphi} \notin B_\varepsilon(\varphi_0)] &\leq P[\inf_{\varphi \in K_\delta \cap \Theta \setminus B_\varepsilon(\varphi_0)} Q_\infty(\varphi) + \lambda_T \|\varphi\|_H^2 + o_p(1) \leq O_p(\lambda_T)] + \delta \\ &\leq P[\inf_{\varphi \in K_\delta \cap \Theta \setminus B_\varepsilon(\varphi_0)} Q_\infty(\varphi) + o_p(1) \leq O_p(\lambda_T)] + \delta \\ &\leq P[\inf_{\varphi \in K_\delta \cap \Theta \setminus B_\varepsilon(\varphi_0)} Q_\infty(\varphi) \leq o_p(1)] + \delta. \end{aligned}$$

Now let $\kappa_{\delta,\varepsilon} := \inf_{\varphi \in K_\delta \cap \Theta \setminus B_\varepsilon(\varphi_0)} Q_\infty(\varphi)$. By compactness of K_δ, continuity of Q_∞ and identification, we have $\kappa_{\delta,\varepsilon} = Q_\infty(\varphi^*_{\delta,\varepsilon}) > 0$ for some $\varphi^*_{\delta,\varepsilon} \in K_\delta \cap \Theta \setminus B_\varepsilon(\varphi_0)$. Thus

$$P[\hat{\varphi} \notin B_\varepsilon(\varphi_0)] \leq P[\kappa_{\delta,\varepsilon} \leq o_p(1)] + \delta \to \delta, \text{ as } T \to \infty.$$

Since δ can be made arbitrary small, we conclude that $P[\hat{\varphi} \notin B_\varepsilon(\varphi_0)] \to 0$. Since $\varepsilon > 0$ is arbitrary, consistency follows.

A.4. **Proof of Proposition 3.** The steps are as follows: 1) getting the first-order condition, 2) deriving a Bahadur representation and an asymptotic expansion of the MISE, 3) proving asymptotic normality and 4) showing bias negligibility.

Step 1. (First-order condition) The following lemma provides the Frechet derivative of operator $\hat{\mathcal{A}}$.

Lemma A.4. *Under Assumption A.2, the Frechet derivative of $\hat{\mathcal{A}}$ at $\bar{\varphi}$ is the linear operator $\bar{A} := D\hat{\mathcal{A}}(\bar{\varphi})$ defined by $\bar{A}\varphi(z) = \int \hat{f}_{X,Y|Z}(x, \bar{\varphi}(x)|z)\varphi(x)\,dx, \quad z \in \mathcal{Z}$, for $\varphi \in L^2[0,1]$. Moreover we have $\hat{\mathcal{A}}(\varphi) = \hat{\mathcal{A}}(\bar{\varphi}) + \bar{A}(\varphi - \bar{\varphi}) + \hat{R}(\varphi, \bar{\varphi})$, where $\hat{R}(\varphi, \bar{\varphi})$ is such that P-a.s., $\left\|\hat{R}(\varphi, \bar{\varphi})\right\|_{L^2(\hat{F}_{Z,\tau})} \leq \frac{1}{2\sqrt{\tau(1-\tau)}}\hat{c}\,\|\varphi - \bar{\varphi}\|^2$, and $\hat{c} := \sup_{x \in \mathcal{X}, y \in \mathbb{R}, z \in \mathcal{Z}} |\nabla_y \hat{f}_{X,Y|Z}(x, y|z)|$.*

By Assumption 1 (ii), let $r > 0$ be such that $B_r(\varphi_0) \cap H^l[0,1]$ is contained in Θ. When $\|\Delta\hat{\varphi}\| < r$ we have: $\forall \varphi \in H^l[0,1]$, $\exists \rho = \rho(\varphi) > 0 : \hat{\varphi} + \varepsilon\varphi \in \Theta$ for any $\varepsilon \in \mathbb{R}$ s.t. $|\varepsilon| < \rho$. By Lemma A.4, the estimator $\hat{\varphi}$ satisfies the first order condition (4.1). We show in the supplementary materials that $P[\|\Delta\hat{\varphi}\| \geq r] = O\left(T^{-\bar{b}}\right)$, for any $\bar{b} > 0$.

Step 2. (Bahadur representation and asymptotic expansion of the MISE) We first rewrite (4.2) as

$$\Delta\hat{\varphi} = \Delta\hat{\psi} + \hat{\mathcal{K}}_T(\Delta\hat{\varphi}), \tag{A.4}$$

where $\Delta\hat{\psi} := \hat{\psi} - \varphi_0$, with $\hat{\psi} := \left(\lambda_T + \hat{A}_0^* \hat{A}_0\right)^{-1} \hat{A}_0^* \hat{r} - \left(\lambda_T + \hat{A}_0^* \hat{A}_0\right)^{-1} \left(\hat{A}^* - \hat{A}_0^*\right) \left(\hat{\mathcal{A}}\left(\hat{\varphi}\right) - \tau\right)$ $=: \hat{\psi}_1 + \hat{\psi}_2$ and $\hat{r} := \tau + \hat{A}_0 \varphi_0 - \widehat{\mathcal{A}}(\varphi_0)$, and $\hat{\mathcal{K}}_T\left(\Delta\hat{\varphi}\right)$ defined in Section 4.2. The interpretation of $\hat{\psi}_1$ is as a linearized solution obtained from applying a Tikhonov regularization to the linear proxy $\hat{A}_0 \varphi = \hat{r}$. Impact of nonlinearity is two-fold. We face the second-order term $\hat{R}$ in $\hat{\mathcal{K}}_T\left(\Delta\hat{\varphi}\right)$ because of the expansion, and we face $\hat{A}^* - \hat{A}_0^*$ in $\hat{\psi}_2$ because of $\hat{\varphi}$ in $\hat{A}$. Now, we decompose $\Delta\hat{\psi}$ as:

$$\Delta\hat{\psi} = \left(\lambda_T + A^* A\right)^{-1} A^* \hat{\zeta} + \mathcal{B}_T + \mathcal{R}_T, \tag{A.5}$$

$$\mathcal{R}_T = \left[\left(\lambda_T + \hat{A}_0^* \hat{A}_0\right)^{-1} - \left(\lambda_T + A^* A\right)^{-1}\right] A^* \hat{\zeta} + \left[\left(\lambda_T + \hat{A}_0^* \hat{A}_0\right)^{-1} \hat{A}_0^* \hat{A}_0 - \left(\lambda_T + A^* A\right)^{-1} A^* A\right] \varphi_0$$
$$+ \left(\lambda_T + \hat{A}_0^* \hat{A}_0\right)^{-1} \left(\hat{A}_0^* \left(\hat{\zeta} - \hat{q}\right) - A^* \hat{\zeta}\right) - \left(\lambda_T + \hat{A}_0^* \hat{A}_0\right)^{-1} \left(\hat{A}^* - \hat{A}_0^*\right) \left(\hat{\mathcal{A}}\left(\hat{\varphi}\right) - \tau\right), \tag{A.6}$$

where $\hat{q} := \hat{\mathcal{A}}\left(\varphi_0\right) - \tau + \hat{\zeta}$. The Bahadur-type representation (4.3) follows from (A.4) and (A.5).

We show by using Lemma A.5 that the nonlinearity term $\hat{\mathcal{K}}_T\left(\Delta\hat{\varphi}\right)$ in Equation (A.4) satisfies a quadratic bound w.p.a. 1.

Lemma A.5. *Under Assumptions A.1-A.3, A.4 (ii)-(iii) and* $\eta < \frac{1}{d_Z+4}$, *for any* $\bar{b} > 0$ *and* $C > \frac{1}{2\sqrt{\tau(1-\tau)}} \sup_{x,y,z} \left|\nabla_y f_{X,Y|Z}(x,y|z)\right|$*:*

$$P\left[\left\|\hat{\mathcal{K}}_T\left(\Delta\hat{\varphi}\right)\right\| > \frac{C}{\sqrt{\lambda_T}} \left\|\Delta\hat{\varphi}\right\|^2\right] = O\left(T^{-\bar{b}}\right).$$

Next, by exploiting (A.4) and the quadratic nature of $\hat{\mathcal{K}}_T\left(\Delta\hat{\varphi}\right)$, we get an asymptotic expansion of the MISE $E\left[\left\|\Delta\hat{\varphi}\right\|^2\right]$ in terms of λ_T and the expectations of powers of $\left\|\Delta\hat{\psi}\right\|$.

Lemma A.6. *Under Assumptions 1-4, A.1-A.3, A.4 (ii)-(iii),* $\eta < \frac{1}{4+d_Z}$ *and provided that* $\gamma < \min\left\{\frac{1-\eta(d_Z+1)}{1+2a}, \frac{2m\eta}{1+2a}, \frac{1}{2(1+a)}\right\}$ *we have that for any* $\bar{b} > 0$

$$E\left[\left\|\Delta\hat{\varphi}\right\|^2\right] = E\left[\left\|\Delta\hat{\psi}\right\|^2\right] + O\left(\frac{1}{\sqrt{\lambda_T}} E\left[\left\|\Delta\hat{\psi}\right\|^3\right]\right) + O\left(T^{-\bar{b}}\right).$$

To compute moments of $\left\|\Delta\hat{\psi}\right\|$, we use the decomposition (A.5) and get the next upper bound for the asymptotic MISE.

Lemma A.7. *Under Assumptions 1-4, A.1-5, and* $\eta < \frac{1}{2(d_Z+2)}$, $\gamma < \frac{1}{2} \min\left\{1 - \eta\left(d_Z + 1\right), m\eta, \frac{1}{1+a}\right\}$, $V_T\left(\lambda_T\right) = o\left(\lambda_T\right)$, $h_T^{1/4} \frac{V_T(\lambda_T;2)}{V_T(\lambda_T)} = o(1)$*:* $E\left[\left\|\Delta\hat{\varphi}\right\|^2\right] = O\left(M_T(\lambda_T)\right)$*, where* $M_T(\lambda_T) = \int \left(\frac{1}{T}\sigma_T^2(x) + \mathcal{B}_T(x)^2\right) dx$.

STEP 3. (Asymptotic normality) From the Bahadur representation (4.3), we have

$$\begin{aligned}\sqrt{T/\sigma_T^2(x)}\left(\hat{\varphi}\left(x\right)-\varphi_0\left(x\right)\right) &= \sqrt{T/\sigma_T^2(x)}\left(\lambda_T+A^*A\right)^{-1}A^*\left(\hat{\zeta}-E\hat{\zeta}\right)(x)\\ &\quad+\sqrt{T/\sigma_T^2(x)}\mathcal{B}_T(x)+\sqrt{T/\sigma_T^2(x)}\hat{\mathcal{K}}_T\left(\Delta\hat{\varphi}\right)(x)\\ &\quad+\sqrt{T/\sigma_T^2(x)}\left(\lambda_T+A^*A\right)^{-1}A^*E\hat{\zeta}\left(x\right)+\sqrt{T/\sigma_T^2(x)}\mathcal{R}_T(x)\\ &=: \text{(I)}+\text{(II)}+\text{(III)}+\text{(IV)}+\text{(V)}.\end{aligned}$$

STEP 3(A). (Asymptotic normality of I) Since $\{\phi_j : j\in\mathbb{N}\}$ is an orthonormal basis w.r.t. $\langle.,.\rangle_H$, we can write:

$$\begin{aligned}\left(\lambda_T+A^*A\right)^{-1}A^*\left(\hat{\zeta}-E\hat{\zeta}\right)(x) &= \sum_{j=1}^{\infty}\left\langle\phi_j,\left(\lambda_T+A^*A\right)^{-1}A^*\left(\hat{\zeta}-E\hat{\zeta}\right)\right\rangle_H\phi_j(x)\\ &= \sum_{j=1}^{\infty}\frac{1}{\lambda_T+\nu_j}\left\langle A\phi_j,\hat{\zeta}-E\hat{\zeta}\right\rangle_{L^2(F_Z,\tau)}\phi_j(x),\end{aligned}$$

for almost any $x\in[0,1]$. Then, we get

$$\sqrt{T/\sigma_T^2(x)}\left(\lambda_T+A^*A\right)^{-1}A^*\left(\hat{\zeta}-E\hat{\zeta}\right)(x)=\sum_{j=1}^{\infty}w_{j,T}(x)Z_{j,T},$$

where $Z_{j,T}:=\left\langle\psi_j,\sqrt{T}\left(\hat{\zeta}-E\hat{\zeta}\right)\right\rangle_{L^2(F_Z,\tau)}$ and $w_{j,T}(x):=\frac{\sqrt{\nu_j}}{\lambda_T+\nu_j}\phi_j\left(x\right)/\sigma_T(x)$, $j=1,2,\cdots$. Note that $\sum_{j=1}^{\infty}w_{j,T}(x)^2=1$. Let $g_j(r):=\frac{1}{\tau(1-\tau)}\psi_j(z)1_{\varphi_0}\left(w\right)$ where $1_{\varphi_0}\left(w\right):=1\{y\leq\varphi_0(x)\}-\tau$, $w=(y,x)$.

Lemma A.8. *Under Assumptions A.1, A.2, A.5 (iii),* $\gamma<m\eta$, $\frac{V_T(\lambda_T)}{\sigma_T^2(x)/T}=O(1)$, $\sqrt{h_T}\frac{V_T(\lambda_T;\varepsilon_1)}{V_T(\lambda_T)}=o(1)$, $h_T^{2m}\frac{V_T(\lambda_T;2m+\varepsilon_1)}{V_T(\lambda_T)}=o(1)$, $\varepsilon_1>1$: $\sum_{j=1}^{\infty}w_{j,T}(x)Z_{j,T}=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Y_{tT}+o_p(1)$, *where* $Y_{tT}:=\sum_{j=1}^{\infty}w_{j,T}(x)g_j(R_t)$, $R_t=(X_t,Y_t,Z_t)$.

From Lemma A.8 it is sufficient to prove that $T^{-1/2}\sum_{t=1}^{T}Y_{tT}$ is asymptotically $N(0,1)$ distributed. Note that $E\left[g_j(R)\right]=\frac{1}{\tau(1-\tau)}\frac{1}{\sqrt{\nu_j}}E\left[\left(A\phi_j\right)(Z)\,E\left[1_{\varphi_0}\left(W\right)|Z\right]\right]=0$, and $Cov\left[g_j(R),g_k(R)\right]=\frac{1}{\sqrt{\nu_j}\sqrt{\nu_k}}\left\langle\phi_j,A^*A\phi_k\right\rangle_H=\delta_{j,k}$. Thus $E\left[Y_{tT}\right]=0$ and $V\left[Y_{tT}\right]=1$. By the Lyapunov CLT, it is sufficient to show that

$$\frac{1}{T^{1/2}}E\left[\left|Y_{tT}\right|^3\right]\to 0,\qquad T\to\infty. \tag{A.7}$$

To this goal, using the triangular inequality and denoting $\left\|g_j\right\|_3:=E\left[\left|g_j(R)\right|^3\right]^{1/3}$, we get $E\left[\left|Y_{tT}\right|^3\right]^{1/3}\leq\sum_{j=1}^{\infty}\left|w_{j,T}(x)\right|\left\|g_j\right\|_3\leq\frac{C}{\sigma_T(x)}\sum_{j=1}^{\infty}\frac{\sqrt{\nu_j}}{\lambda_T+\nu_j}\left|\phi_j\left(x\right)\right|$, where $C:=\frac{1}{\tau(1-\tau)}\sup_{j\in\mathbb{N}}E\left[\left|\psi_j\left(Z\right)\right|^3\right]^{1/3}<\infty$ from Assumption A.5 (ii). From the Cauchy-Schwarz inequality we have $\sum_{j=1}^{\infty}\frac{\sqrt{\nu_j}}{\lambda_T+\nu_j}\left|\phi_j(x)\right| \leq \left(\sum_{j=1}^{\infty}\frac{\nu_j}{(\lambda_T+\nu_j)^2}\phi_j(x)^2 j^{\varepsilon_2}\right)^{1/2}\left(\sum_{j=1}^{\infty}\frac{1}{j^{\varepsilon_2}}\right)^{1/2}$, and $\sum_{j=1}^{\infty}\frac{1}{j^{\varepsilon_2}} < \infty$, for any $\varepsilon_2 > 1$. From (4.7) we get $\frac{1}{T^{1/2}}E\left[\left|Y_{tT}\right|^3\right] = O\left(\left[\frac{1}{T^{1/3}}\frac{\delta_T(x)}{V_T(\lambda_T)}\right]^{3/2}\right)$, where $\delta_T(x) := \frac{1}{T}\sum_{j=1}^{\infty}\frac{\nu_j}{(\lambda_T+\nu_j)^2}\phi_j(x)^2 j^{\varepsilon_2}$. Condition (A.7) follows from $\int\frac{1}{T^{1/3}}\frac{\delta_T(x)}{V_T(\lambda_T)}dx = \frac{1}{T^{1/3}}\frac{V_T(\lambda_T;\varepsilon_2)}{V_T(\lambda_T)} = O\left(h_T^{1/4}\frac{V_T(\lambda_T;\varepsilon_2)}{V_T(\lambda_T)}\right) = o(1)$.

Step 3(b). (Negligibility of III) We use the following lemma.

Lemma A.9. *Under Assumptions A.1-A.3, A.4 (ii)-(iii),* $\eta < \frac{1}{d_Z+4}$*:* $\hat{\mathcal{K}}_T(\Delta\hat{\varphi})(x) = O_p\left(\frac{1}{\sqrt{\lambda_T}}\|\Delta\hat{\varphi}\|^2\right)$.

By Lemmas A.7 and A.9, we get $\hat{\mathcal{K}}_T(\Delta\hat{\varphi})(x) = O_p\left(\frac{1}{\sqrt{\lambda_T}}M_T(\lambda_T)\right)$. Since $\int\mathcal{B}_T(x)^2dx = O\left(\lambda_T^{2\delta}\right)$ (see Step 4) and $\lambda_T^{2\delta} = O(V_T(\lambda_T))$, we have $M_T(\lambda_T) = O(V_T(\lambda_T))$. Then, we get $\sqrt{T/\sigma_T^2(x)}\hat{\mathcal{K}}_T(\Delta\hat{\varphi})(x) = O_p\left(\sqrt{\frac{V_T(\lambda_T)}{\lambda_T}}\right) = o_p(1)$ from (4.7) and $V_T(\lambda_T) = o(\lambda_T)$.

Step 3(c). (Negligibility of IV and V) Here we rely on the following lemmas.

Lemma A.10. *Under Assumptions A.2, A.4 (i),* $\gamma < \frac{m\eta}{2}$, $\lambda_T^{2\delta} = O(V_T(\lambda_T))$, $\frac{V_T(\lambda_T)}{\sigma_T^2(x)/T} = O(1)$: $\sqrt{T/\sigma_T^2(x)}\left(\lambda_T + A^*A\right)^{-1}A^*E\hat{\zeta}(x) = o(1)$.

Lemma A.11. *Under Assumptions A.1-5,* $\lambda_T^{2\delta} = O(V_T(\lambda_T))$, $\frac{V_T(\lambda_T)}{\sigma_T^2(x)/T} = O(1)$, $V_T(\lambda_T) = o(\lambda_T)$, $\eta < \frac{1}{2(d_Z+2)}$, $\gamma < \frac{1}{2}\min\left\{1-\eta(d_Z+1), m\eta, \frac{1}{1+a}\right\}$: $\sqrt{T/\sigma_T^2(x)}\mathcal{R}_T(x) = o_p(1)$.

Step 4. (Bias negligibility) Similarly to the proof of Proposition 3.11 in CFR, we have

$$\|\mathcal{B}_T\|_H^2 = \sum_{j=1}^{\infty}\frac{\lambda_T^2\langle\phi_j,\varphi_0\rangle_H^2}{(\lambda_T+\nu_j)^2} = \lambda_T^{2\delta}\sum_{j=1}^{\infty}\frac{\lambda_T^{2-2\delta}\nu_j^{2\delta}}{(\lambda_T+\nu_j)^2}\frac{\langle\phi_j,\varphi_0\rangle_H^2}{\nu_j^{2\delta}} \leq \lambda_T^{2\delta}\sum_{j=1}^{\infty}\frac{\langle\phi_j,\varphi_0\rangle_H^2}{\nu_j^{2\delta}} = O\left(\lambda_T^{2\delta}\right),$$

from Assumption 4 (i). This proves (4.5). Moreover, from Lemma C.1 in the supplementary materials, we have $\mathcal{B}_T(x) \leq 2\|\mathcal{B}_T\|_H$, for any $x \in [0,1]$. Thus, from (4.7) we get $\sqrt{T/\sigma_T^2(x)}\mathcal{B}_T(x) = O\left(\sqrt{\lambda_T^{2\delta}/V_T(\lambda_T)}\right) = o(1)$ for any typical point x, if $\lambda_T^{2\delta} = o(V_T(\lambda_T))$.

A.5. **Proof of Proposition 4.** Condition $\frac{\sigma_T^2(x)}{\hat{\sigma}_T^2(x)} \xrightarrow{p} 1$ is equivalent to $\frac{\left|\hat{\sigma}_T^2(x)-\sigma_T^2(x)\right|}{\sigma_T^2(x)} = o_p(1)$. Let us introduce the truncated series $\sigma_{0,T}^2(x) = \sum_{j=1}^{N_T}\frac{\nu_j}{(\nu_j+\lambda_T)^2}\phi_j(x)^2$, and the residual $\sigma_{1,T}^2(x) = \sum_{j=N_T+1}^{\infty}\frac{\nu_j}{(\nu_j+\lambda_T)^2}\phi_j(x)^2$. We have $\frac{\left|\hat{\sigma}_T^2(x)-\sigma_T^2(x)\right|}{\sigma_T^2(x)} \leq \frac{\left|\hat{\sigma}_T^2(x)-\sigma_{0,T}^2(x)\right|}{\sigma_T^2(x)} + \frac{\sigma_{1,T}^2(x)}{\sigma_T^2(x)}$. In Step 1 we show that the residual in the truncated series is negligible: $\xi_T := \frac{\sigma_{1,T}^2(x)}{\sigma_T^2(x)} = o(1)$. In Step 2 we show that the truncated estimator is consistent for the truncated series: $\delta_T := \frac{\left|\hat{\sigma}_T^2(x)-\sigma_{0,T}^2(x)\right|}{\sigma_T^2(x)} = o_p(1)$. In Step 3 we show the relative consistency of the estimated spectrum.

STEP 1. (Condition on the cut-off) Using $\int_0^1 \sigma_{1,T}^2(x)dx = \sum_{j=N_T+1}^{\infty} \frac{\nu_j}{(\nu_j+\lambda_T)^2} \|\phi_j\|^2 \leq \lambda_T^{-2} \sum_{j=N_T+1}^{\infty} \nu_j \|\phi_j\|^2$ and Condition (4.8), we get $\frac{\int_0^1 \sigma_{1,T}^2(x)dx/T}{V_T(\lambda_T)} = o(1)$. This implies $\frac{\sigma_{1,T}^2(x)/T}{V_T(\lambda_T)} = o(1)$ for almost all $x \in [0,1]$. By using $\frac{V_T(\lambda_T)}{\sigma_T^2(x)/T} = O(1)$, it follows that $\xi_T = o(1)$.

STEP 2. (Consistent estimation of the truncated series) We use the next Lemma.

Lemma A.12. *Let* $\epsilon_{1,T} := \sup_{1\leq j\leq N_T} \frac{|\hat{\nu}_j - \nu_j|}{\nu_j}$ *and* $\epsilon_{2,T} := \sup_{1\leq j\leq N_T} \frac{\left|\hat{\phi}_j(x)^2 - \phi_j(x)^2\right|}{\xi_j^*}$, *where* $\xi_j^* = c_{2,j}\exp(w_j\beta)$. *Then* $\delta_T \leq C\epsilon_{2,T}(1+\epsilon_{1,T}) + \epsilon_{1,T} + \sqrt{1+\delta_T - \xi_T}\frac{\epsilon_{1,T}}{\sqrt{1-\epsilon_{1,T}}}\left(1+\sqrt{C\epsilon_{2,T}}\right)$, *for a constant* $C > 0$.

Since $\xi_T = o(1)$ from Step 1, we get $\delta_T = o_p(1)$ if we show $\epsilon_{1,T} = o_p(1)$ and $\epsilon_{2,T} = o_p(1)$.

STEP 3(A). (Relative consistency of $\hat{\nu}_j$ and $\hat{\phi}_j(x)^2$ uniformly over $1 \leq j \leq n_T$) Let us first show that $\sup_{1\leq j\leq n_T} \frac{|\bar{\nu}_j - \nu_j|}{\nu_j} = O_p(T^{-b})$ and $\sup_{1\leq j\leq n_T} \frac{\left|\bar{\phi}_j(x)^2 - \phi_j(x)^2\right|}{\xi_j^*} = O_p(T^{-b})$, for $b > 0$. These conditions are implied by

$$\frac{1}{\nu_{n_T}} \sup_{1\leq j\leq n_T} |\bar{\nu}_j - \nu_j| = o_p(T^{-b}), \quad \frac{1}{\xi_{n_T}^*} \sup_{1\leq j\leq n_T} \left\|\bar{\phi}_j - \phi_j\right\| = o_p(T^{-b}), \tag{A.8}$$

for $b > 0$. To prove (A.8) we use the next Lemma A.13, which follows from Lemmas 4.2 and 4.3 in Bosq (2000); see also Theorem 1 in Hall and Hosseini-Nasab (2006) for a spectral decomposition in $L^2[0,1]$.

Lemma A.13. *For any* j: *(i)* $|\bar{\nu}_j - \nu_j| \leq \left\|\hat{D}\right\|_H$, *and (ii)* $\left\|\bar{\phi}_j - \phi_j\right\| \leq \frac{2\sqrt{2}}{\Delta\nu_j}\left\|\hat{D}\right\|_H$, *where* $\hat{D} := \bar{A}^*\bar{A} - A^*A$, *and* $\left\|\hat{D}\right\|_H := \sup_{\varphi\in H^l[0,1]:\|\varphi\|_H=1}\left\|\hat{D}\varphi\right\|_H$ *denotes the operator norm in* $H^l[0,1]$.

From Lemma A.13, an upper bound on the rate of convergence of $|\bar{\nu}_j - \nu_j|$ and $\left\|\bar{\phi}_j - \phi_j\right\|$ can be deduced from the rate of convergence of $\left\|\hat{D}\right\|_H$.

Lemma A.14. *Under Assumptions A.1-5:* $\left\|\hat{D}\right\|_H = O_p\left(\frac{1}{\sqrt{Th_T^2}} + h_T^m + \|\bar{\varphi} - \varphi_0\|\right)$.

From Lemmas A.13 and A.14, and $\|\bar{\varphi} - \varphi_0\| = O_p\left(M_T\left(\bar{\lambda}_T\right)^{1/2}\right)$ from Lemma A.7, we get $\sup_{1\leq j\leq n_T}|\bar{\nu}_j - \nu_j| = O_p(\kappa_T)$ and $\sup_{1\leq j\leq n_T}\left\|\bar{\phi}_j - \phi_j\right\| = O_p\left(\frac{\kappa_T}{\Delta\nu_{n_T}}\right)$, where $\kappa_T := \frac{1}{\sqrt{Th_T^2}} + h_T^m + M_T\left(\bar{\lambda}_T\right)^{1/2}$. Thus, (A.8) follows from condition (4.9) and $\Delta\nu_{n_T} \leq \nu_{n_T-1}$.

STEP 3(B). (Relative consistency of $\hat{\nu}_j$ and $\hat{\phi}_j(x)^2$ uniformly over $1 \leq j \leq N_T$) Finally, we prove in the next Lemma A.15 that the uniform convergence of $\hat{\nu}_j$ and $\hat{\phi}_j$ can be extended to $j \leq N_T$ by the extrapolation procedure.

Lemma A.15. *Under Assumptions 5 and A.1-5, and if* $N_T = O(n_T)$: *(i)* $\sup_{n_T<j\leq N_T}\frac{|\hat{\nu}_j - \nu_j|}{\nu_j} = o_p(1)$, *(ii)* $\sup_{n_T<j\leq N_T}\frac{\left|\hat{\phi}_j(x)^2 - \phi_j(x)^2\right|}{\xi_j^*} = o_p(1)$.

	$l=1$					$l=2$				
	$\bar{\lambda}=\lambda=.00002$					$\bar{\lambda}=\lambda=.0000003$				
	$x=0.1$	$x=.25$	$x=.5$	$x=.75$	$x=.9$	$x=0.1$	$x=.25$	$x=.5$	$x=.75$	$x=.9$
90%	.909	.952	.754	.947	.868	.928	.986	.718	.990	.917
95%	.956	.973	.850	.978	.937	.967	.993	.807	.998	.953
99%	.989	.996	.953	.993	.979	.994	.997	.915	1	.989
	$\bar{\lambda}=\lambda=.00005$					$\bar{\lambda}=\lambda=.000001$				
	$x=0.1$	$x=.25$	$x=.5$	$x=.75$	$x=.9$	$x=0.1$	$x=.25$	$x=.5$	$x=.75$	$x=.9$
90%	.964	.968	.879	.956	.947	.989	.985	.922	.973	.986
95%	.989	.989	.935	.981	.974	.996	.997	.971	.996	.996
99%	.996	1	.983	.996	.998	.999	.999	.994	.999	1
	$\bar{\lambda}=\lambda=.0001$					$\bar{\lambda}=\lambda=.000003$				
	$x=0.1$	$x=.25$	$x=.5$	$x=.75$	$x=.9$	$x=0.1$	$x=.25$	$x=.5$	$x=.75$	$x=.9$
90%	.985	.958	.954	.951	.977	.999	.943	.992	.927	1
95%	.995	.980	.984	.979	.994	.999	.978	.999	.964	1
99%	.999	.992	.998	.998	1	1	.995	1	.995	1

TABLE I: Finite-sample coverage probabilities of confidence intervals for $\varphi_0(x)$.

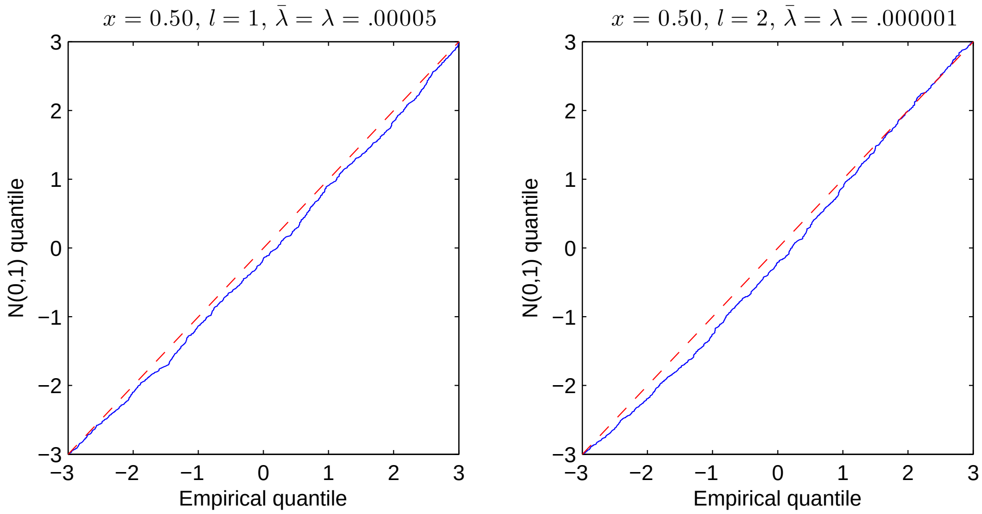

FIGURE 1. QQ-plot of $\sqrt{T/\hat{\sigma}_T^2(x)}\,(\hat{\varphi}(x) - \varphi_0(x))$.

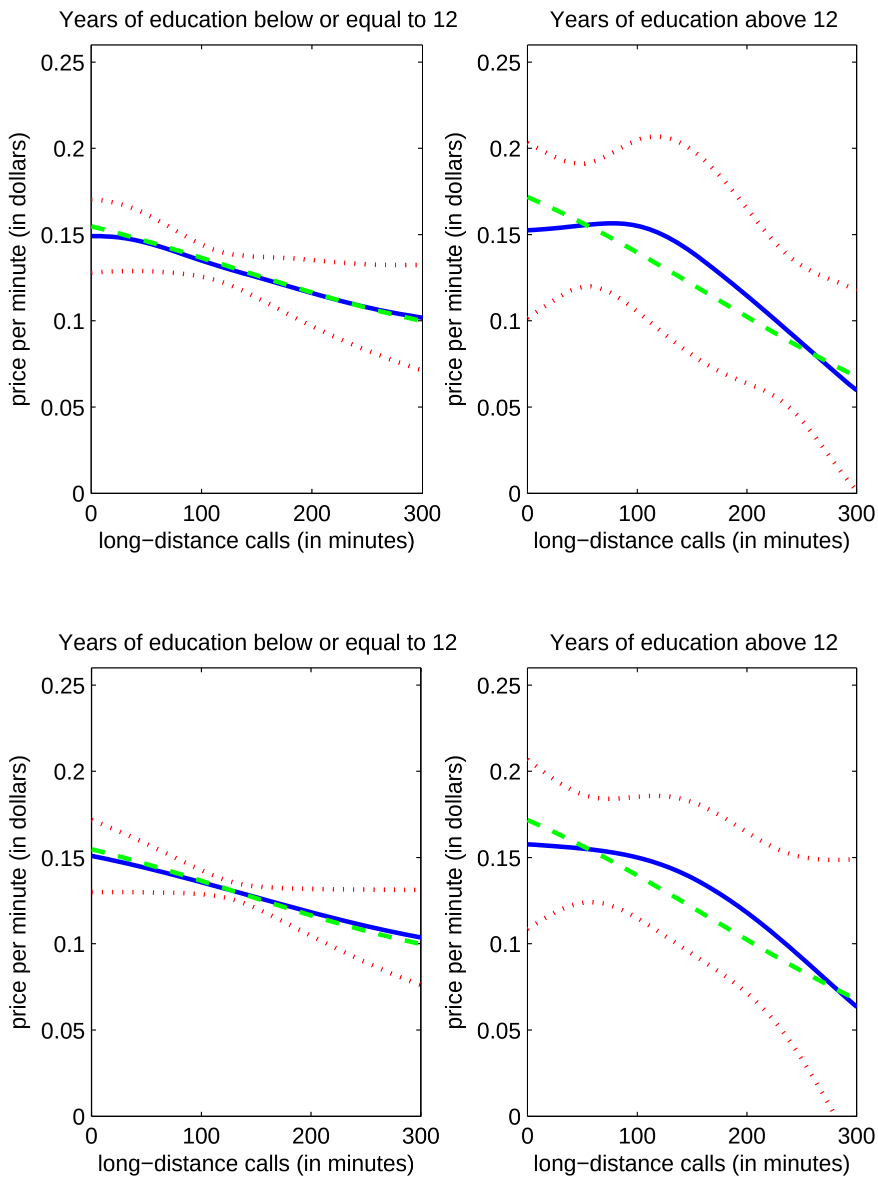

FIGURE 2. Estimated median structural effect for the two education categories: NIVQR (solid line), IVQR (dashed line), 95% confidence intervals (dotted lines). The upper panel uses a Sobolev penalty of order $l = 1$, the lower panel uses $l = 2$.

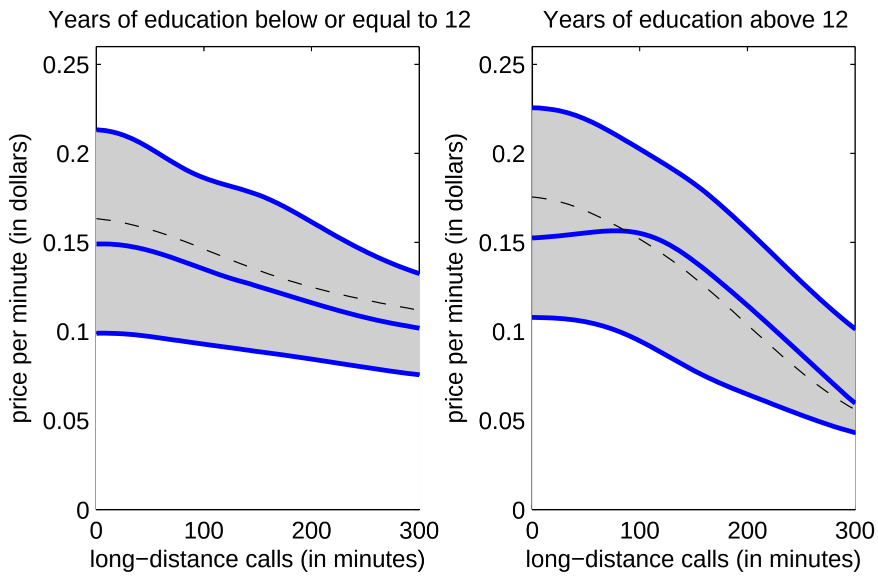

FIGURE 3. Estimated quartile structural effects (solid lines) and mean structural effects (dashed lines) for the two education categories, with a Sobolev penalty of order $l = 1$.

UNIVERSITY OF LUGANO AND SWISS FINANCE INSTITUTE;
UNIVERSITY OF GENEVA AND SWISS FINANCE INSTITUTE