

**NONPARAMETRIC INSTRUMENTAL VARIABLE
ESTIMATION OF QUANTILE STRUCTURAL EFFECTS**

V. Chernozhukov*, P. Gagliardini[†] and O. Scaillet[‡]

This version: May 2008 [§]

(First version: December 2006)

*MIT.

[†]University of Lugano and Swiss Finance Institute.

[‡]HEC University of Geneva and Swiss Finance Institute. Corresponding author: Olivier Scaillet, HEC Genève UNI MAIL, Faculté des SES, Bd Carl Vogt 102, CH-1211 Genève 4, Switzerland. Tel.: ++ 41 22 379 88 16. Fax ++ 41 22 379 81 04. Email: Olivier.Scaillet@hec.unige.ch.

[§]The last two authors received support by the Swiss National Science Foundation through the National Center of Competence in Research: Financial Valuation and Risk Management (NCCR FINRISK). We would like to thank Jerry Hausman and Greg Sidak for generously providing us with the data on telecommunications services in the U.S. We also thank Geert Dhaene, Roger Koenker, Oliver Linton, Enno Mammen, Michael Wolf, and seminar participants at Athens University, Zurich University, Bern University, Boston University, Queen Mary, Greqam Marseille, Mannheim University, Leuven University for helpful comments.

Nonparametric Instrumental Variable Estimation of Quantile Structural Effects

Abstract

We study 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 derivative. We derive the asymptotic mean integrated square error, the rate of convergence and the pointwise asymptotic normality under a regularization parameter depending on sample size. We illustrate our theoretical findings and the small sample properties with simulation results in two numerical examples. We also discuss a data driven selection procedure of the regularization parameter via a spectral representation of the mean integrated squared error. Finally, we provide an empirical illustration to estimation of nonlinear pricing curves for telecommunications services in the U.S.

Keywords and phrases: Quantile Regression, Nonparametric Estimation, Instrumental Variable, Ill-Posed Inverse Problems, Tikhonov Regularization, Nonlinear Pricing Curve.

JEL classification: C13, C14, D12.

AMS 2000 classification: 62G08, 62G20, 62P20.

1 Introduction

This paper deals with nonparametric estimation of quantile functions that measure the structural impact of an endogenous explanatory variable X on the quantiles of the dependent variable Y . In the underlying model Y is linked to X by a structural quantile function that is strictly monotonic increasing in a nonseparable scalar disturbance, and the disturbance is independent of instrument Z (Chernozhukov and Hansen (2005), Chernozhukov, Imbens and Newey (2007)). The concept of endogenous quantile regression extends the exogenous quantile regression (QR) introduced in the seminal work of Koenker and Bassett (1978). Quantile regression has become a basic method of econometrics, with a great wealth of applications ranging from labor economics to finance, see e.g. Fitzenberger, Koenker and Machado (2001) and Koenker (2005).

In this paper we follow the route of Gagliardini and Scaillet (GS, 2006), and study a Tikhonov Regularized (TiR) estimator where regularization is achieved via a compactness-inducing penalty term - the Sobolev norm - that incorporates both the norm of functional parameter and the norm of its derivative (Groetsch (1984); see also Koenker and Mizera (2004) for related total variation penalization in exogenous quantile regression settings). The use of this penalty is important for resolving the ill-posedness of the estimation problem and for delivering good finite sample performance of the estimator. We can summarize the main results that we obtain concerning this estimator as follows. First, we formally prove the local ill-posedness of the endogenous quantile estimation problem when based solely on a minimum distance criterion, and explain how to regularize it through a Sobolev norm penal-

ization. The use of the latter leads to an effective compactification of the parameter space and in turn leads to very weak requirements on the choice of the regularization parameter, which should be of interest in a more general context. Second, we derive the asymptotic properties of the Q-TiR estimator, namely, we establish consistency under very weak conditions and derive the leading term in the asymptotic Mean Integrated Square Error (MISE), which also gives us the rates of convergence. Third, we prove the pointwise asymptotic normality under a regularization parameter depending on sample size. The last two sets of results are particularly noteworthy. Indeed, the problem and the associated difficulty of obtaining exact MISE expansions and asymptotic normality in a nonlinear setting have been pointed out in Horowitz and Lee (2007). Here we resolve these difficulties. We show how to control the errors induced by linearization of the problem under suitable smoothness assumptions. This amounts to showing the validity of a Bahadur-type representation for the functional estimator (see e.g. Koenker (2005) for Bahadur-type representations in the finite-dimensional parameter case). This in turn makes it possible to derive the explicit expression of the asymptotic MISE, instead of the upper bounds found by Horowitz and Lee (2007) for L^2 -penalized estimators, and to finally show asymptotic normality of the functional estimator. Fourth, we check validity of our assumptions in a Gaussian example. Fifth, we illustrate our theoretical findings and the small sample properties with simulation results in a separable and a nonseparable model. We also discuss a data driven selection procedure of the regularization parameter via a spectral representation of the MISE. Sixth, we provide an empirical illustration to nonlinear pricing curves for telecommunications services in the

U.S.

Our paper is related to the previous literature as follows. This paper builds on a series of fundamental papers in econometrics (Ai and Chen (2003), Darolles, Florens, and Renault (2003), Newey and Powell (2003), Hall and Horowitz (2005), Horowitz (2007), Blundell, Chen, and Kristensen (2007)) that introduce and study the ill-posed endogenous mean regression setup. This paper focuses on endogenous quantile regression, building on the insights of these previous papers. Chernozhukov, Imbens and Newey (2007), and Horowitz and Lee (2007), have also considered nonparametric instrumental variable (NIV) estimation of endogenous quantile regression. Chernozhukov, Imbens and Newey (2007) discuss identification¹, and consider consistent estimation via a constrained minimum distance criterion as in Newey and Powell (2003) and Ai and Chen (2003). Horowitz and Lee (2007) give optimal consistency rates for a L^2 -penalized estimator derived in the same spirit as the NIVR estimator of Hall and Horowitz (2005) (see Darolles, Florens and Renault (2003) for a related estimator, and the review paper by Carrasco, Florens, and Renault (2005)). Finally, in an independent work, Chen and Pouzo (2008) study semiparametric sieve estimation of conditional moment models based on possibly nonsmooth generalized residual functions. They extend the consistency and rate results of Blundell, Chen and Kristensen (2007) to cover partially linear quantile IV regression as a particular example. They also give an in-depth, unifying treatment of sieve-based and function space-based estimators and quantify when the first or the second approach dominate the asymptotic analysis of convergence rates.

¹ See Imbens and Newey (2002) and Chesher (2003) for a control function approach, and Chesher (2007) for comparison between control function and single equation IV approaches.

The results of our paper and those of Chen and Pouzo (2008) are complementary to each other. Our results also go substantively beyond the previous results in other papers on the endogenous quantile model. Indeed, the exact characterization of the MISE and pointwise normality of the nonparametric estimator are new and represent the main contribution of this paper. Other contributions of this paper, such as the formal proof of ill-posedness in a nonlinear setting and the consistency results under weak conditions using the compactness-inducing Sobolev penalty, are also of independent interest. Finally, the function space-based estimator with Sobolev penalty itself is new for our problem and should be relevant also for applications other than Engel curve and nonlinear pricing schedule analysis.

We organize the rest of the paper as follows. In Section 2, we formally prove the local ill-posedness of the estimation problem, and clarify the importance of including a derivative term in a penalization approach. In Section 3, we explain how we can resolve the ill-posedness by introducing a Sobolev penalty, and define our function space-based estimator. In Section 4, we prove consistency. In Section 5, we derive the exact MISE expression, the rates of convergence and the asymptotic normality of the estimator. Also, in Section 5, we verify the regularity conditions and illustrate the results of this paper in a Gaussian example. In Section 6, we provide computational experiments, and present our empirical illustration. In the Appendix, we gather the technical assumptions and proofs. Finally, we place all omitted proofs of technical Lemmas and detailed computations in the Gaussian example in a Technical Report, which is available online at our web pages.

2 Ill-posedness in nonseparable models

Let us consider the nonseparable model $Y = g(X, U)$ of Chernozhukov, Imbens and Newey (2007), where the error U is independent of the instrument Z , and has a uniform distribution $U \sim \mathcal{U}(0, 1)$. 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 variable Y also has compact support \mathcal{Y} in $[0, 1]$.² The variable Z has support $\mathcal{Z} \subset \mathbb{R}^{d_Z}$. 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 functional parameter φ_0 belongs to a subset Θ of the Sobolev space $H^2[0, 1]$, i.e., the completion of the linear space $\{\varphi \in C^1[0, 1] \mid \nabla\varphi \in L^2[0, 1]\}$ with respect to the scalar product $\langle \varphi, \psi \rangle_H := \langle \varphi, \psi \rangle + \langle \nabla\varphi, \nabla\psi \rangle$, where $\langle \varphi, \psi \rangle = \int_{\mathcal{X}} \varphi(x)\psi(x)dx$. The Sobolev space $H^2[0, 1]$ is an Hilbert space w.r.t. the scalar product $\langle \varphi, \psi \rangle_H$, and the corresponding Sobolev norm is denoted by $\|\varphi\|_H = \langle \varphi, \varphi \rangle_H^{1/2}$. We use the L^2 norm $\|\varphi\| = \langle \varphi, \varphi \rangle^{1/2}$ as consistency norm, and we assume that Θ is bounded w.r.t. $\|\cdot\|$.

The function φ_0 satisfies the conditional quantile restriction

$$P[Y \leq \varphi_0(X) \mid Z] = P[g(X, U) \leq g(X, \tau) \mid Z] = P[U \leq \tau \mid Z] = \tau, \quad (1)$$

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

$$Y = \varphi_0(X) + V, \quad P[V \leq 0 \mid Z] = \tau. \quad (2)$$

From (1), the quantile structural effect is the solution to a nonlinear functional equation

$$\mathcal{A}(\varphi_0) = \tau, \quad (3)$$

² In the nonparametric nonseparable setting, a compact support for both X and Y can be achieved by transformation of the model w.l.o.g.. We assume this to simplify the proofs.

where the operator \mathcal{A} is defined by $\mathcal{A}(\varphi)(z) = \int_{\mathcal{X}} F_{Y|X,Z}(\varphi(x)|x, z) f_{X|Z}(x|z) dx$, $z \in \mathcal{Z}$, 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, we can rewrite the operator \mathcal{A} in terms of the conditional c.d.f. $F_{U|X,Z}$ of U given X, Z as $\mathcal{A}(\varphi)(z) = \int_{\mathcal{X}} F_{U|X,Z}(g^{-1}(x, \varphi(x))|x, z) f_{X|Z}(x|z) dx$, where $g^{-1}(x, \cdot)$ denotes the generalized inverse of function $g(x, \cdot)$ w.r.t. its second argument.³

We assume identification.

Assumption 1: *The function φ_0 is identified.*

Sufficient conditions ensuring Assumption 1 locally around φ_0 are given in Chernozhukov, Imbens and Newey (2007).

Equation (3) implies the conditional moment restriction

$$m(\varphi_0, z) := E[1\{Y - \varphi_0(X) \leq 0\} - \tau | Z = z] = \mathcal{A}(\varphi_0)(z) - \tau = 0, \quad z \in \mathcal{Z}.$$

We consider a minimum distance approach for the estimation of the quantile structural effect φ_0 . The limit criterion is

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

where F_Z denotes the marginal distribution of Z and $L^2(F_Z, \tau)$ denotes the L^2 space w.r.t.

³ Since Y has a compact support, the strictly monotonic increasing function $u \rightarrow g(x, u)$ has a bounded range $[\underline{\gamma}, \bar{\gamma}]$, where $\underline{\gamma}$ and $\bar{\gamma}$ may depend on x , for any $x \in [0, 1]$. Thus, the inverse of $g(x, \cdot)$ is not defined at $y \notin [\underline{\gamma}, \bar{\gamma}]$. The generalized inverse $g^{-1}(x, \cdot)$ is such that $g^{-1}(x, y) = 1$ for $y > \bar{\gamma}$ and $g^{-1}(x, y) = 0$ for $y < \underline{\gamma}$. Consequently, g^{-1} is such that $g(x, u) \leq y$ if and only if $u \leq g^{-1}(x, y)$, for any $x \in \mathcal{X}$, $u \in [0, 1]$ and $y \in \mathbb{R}$.

measure $F_Z/(\tau(1-\tau))$.⁴ By the identification assumption, φ_0 is the unique minimizer of Q_∞ on Θ .

The minimum distance problem is locally ill-posed if, for any $r > 0$ small enough, there exist $\varepsilon \in (0, r)$ and a sequence $(\varphi_n) \subset B_r(\varphi_0) := \{\varphi \in L^2[0, 1] : \|\varphi - \varphi_0\| < r\}$ such that $\|\varphi_n - \varphi_0\| \geq \varepsilon$ and $Q_\infty(\varphi_n) \rightarrow 0$ (see e.g. Definition 1.1 in Hofmann and Scherzer (1998)). Under a stronger condition than Assumption 1, namely local injectivity of \mathcal{A} , this 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). This follows from $Q_\infty(\varphi) = \|\mathcal{A}(\varphi) - \mathcal{A}(\varphi_0)\|_{L^2(F_Z, \tau)}^2$.

Proposition 1: *Under Assumptions 1 and A.3 (i)-(iii): (a) the problem is locally ill-posed; (b) any sequence $(\varphi_n) \subset B_r(\varphi_0)$ such that $\|\varphi_n - \varphi_0\| \geq \varepsilon$ for $r > \varepsilon > 0$ and $Q_\infty(\varphi_n) \rightarrow 0$ satisfies $\limsup_{n \rightarrow \infty} \|\nabla \varphi_n\| = +\infty$.*

Part (a) of Proposition 1 establishes the local ill-posedness of the minimum distance estimation problem under suitable boundedness and smoothness assumptions on $g(x, \tau)$, $f_{X|Z}$, and $F_{U|X, Z}$. Contrary to what we might expect, 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)). Several counterexamples are available in the literature (Schock (2002)). The proof in Appendix 2 relies on a constructive approach and gives explicit sequences (φ_n) generating ill-posedness. Part (b)

⁴ The constant weighting factor $1/(\tau(1-\tau))$ is irrelevant for minimization of Q_∞ . However, it matters for the normalization of the regularization parameter in the penalized criterion (see footnote 7) and yields an expression of the asymptotic MISE (see Proposition 3) formally similar to the one for the regression case given in GS. The weighting factor $1/(\tau(1-\tau))$ is the inverse of the conditional variance $V[1\{Y - \varphi_0(X) \leq 0\} - \tau|Z]$ of the moment function.

of Proposition 1 provides a theoretical underpinning for a penalization approach based on the Sobolev norm which includes the penalty term $\|\nabla\varphi\|$ in its definition.

3 The Q-TiR estimator

We address ill-posedness by Tikhonov regularization (Tikhonov (1963a,b); see also Kress (1999), Chapter 16). We consider a penalized criterion $Q_T(\varphi) + \lambda_T \|\varphi\|_H^2$, where $Q_T(\varphi)$ is an empirical counterpart of $Q_\infty(\varphi)$ defined by

$$Q_T(\varphi) := \frac{1}{T\tau(1-\tau)} \sum_{t=1}^T \hat{m}(\varphi, Z_t)^2 I_t, \quad (4)$$

and λ_T is a sequence of strictly positive regularization parameters vanishing as sample size grows. In (4) we estimate the conditional moment nonparametrically with

$$\hat{m}(\varphi, z) := \int_{\mathcal{X}} \hat{f}_{X|Z}(x|z) \hat{F}_{Y|X,Z}(\varphi(x)|x, z) dx - \tau =: \hat{\mathcal{A}}(\varphi)(z) - \tau, \quad z \in \mathcal{Z}, \quad (5)$$

where $\hat{f}_{X|Z}$ and $\hat{F}_{Y|X,Z}$ denote kernel estimators of $f_{X|Z}$ and $F_{Y|X,Z}$ with kernel K and bandwidth h_T . Indicator $I_t = 1 \left\{ Z_t \in \mathcal{Z}_T, \hat{f}_Z(Z_t) \geq (\log T)^{-1} \right\}$ is a trimming factor based on the sequence of sets $\mathcal{Z}_T \subset \mathcal{Z}$ (see Assumption A.7) which controls for small values of kernel estimator \hat{f}_Z of f_Z .

Definition 1: *The Q-TiR estimator is defined by*

$$\hat{\varphi} := \arg \inf_{\varphi \in \Theta} Q_T(\varphi) + \lambda_T \|\varphi\|_H^2 =: L_T(\varphi), \quad (6)$$

where $Q_T(\varphi)$ is as in (4), and λ_T is a stochastic sequence with $\lambda_T > 0$ and $\lambda_T \rightarrow 0$, P -a.s..

We prove in Appendix 3.1 that the estimator exists. Term $\lambda_T \|\varphi\|_H^2$ in (6) penalizes highly oscillating components of the estimated function induced by ill-posedness, and restores its consistency.

4 Consistency of the Q-TiR estimator

The next result establishes consistency of the Q-TiR estimator with stochastic regularization parameter.

Proposition 2: *Suppose λ_T is such that $\lambda_T > 0$, $\lambda_T \rightarrow 0$, P -a.s. and $\frac{(\log T)^2}{\lambda_T} \left(\frac{\log T}{Th_T^{d_Z+1}} + h_T^{2m} \right) = O_p(1)$, where $m \geq 2$ is the order of differentiability of the joint density of (X, Y, Z) . Then, under Assumptions 1, A.1-A.3, A.7 and A.10, the Q-TiR estimator $\hat{\varphi}$ is consistent, namely $\|\hat{\varphi} - \varphi_0\| \xrightarrow{p} 0$.*

The proof of Proposition 2 in Appendix 3.2 relies on two results. First, the Sobolev penalty implies that the sequence of estimates $\hat{\varphi}$ for $T \in \mathbb{N}$ is tight in $(L^2[0, 1], \|\cdot\|)$. 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 show a suitable uniform convergence result for Q_T on Θ . We obtain such a result with an infinite-dimensional and possibly non-totally bounded parameter set by exploiting the specific expression of $\hat{m}(\varphi, z)$ given in (5). We are able to reduce the sup over Θ to a sup over a bounded subset of a finite-dimensional space. Combining the tightness and uniform convergence results allows us to conclude on consistency by an argument similar to the one for finite-dimensional or well-posed settings.

5 Asymptotic distribution of the Q-TiR estimator

In the rest of the paper we assume a deterministic regularization parameter λ_T .

5.1 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 following operators from $L^2[0, 1]$ to $L^2(F_Z, \tau)$

$$A\psi(z) = \int f_{X,Y|Z}(x, \varphi_0(x)|z)\psi(x) dx, \quad (7)$$

and

$$\hat{A}\psi(z) = \int \hat{f}_{X,Y|Z}(x, \hat{\varphi}(x)|z)\psi(x) dx, \quad (8)$$

where $z \in \mathcal{Z}$, $\psi \in L^2[0, 1]$. These operators correspond to the Frechet derivative $A := D\mathcal{A}(\varphi_0)$ of operator \mathcal{A} at φ_0 , and to the Frechet derivative $\hat{A} := D\hat{\mathcal{A}}(\hat{\varphi})$ of operator $\hat{\mathcal{A}}$ at $\hat{\varphi}$, respectively (see Appendix 4.1). Under Assumption A.6, operator A is compact. In Appendix 4.2 we show that the Q-TiR estimator satisfies w.p.a. 1 the first-order condition

$$0 = \left. \frac{d}{d\varepsilon} L_T(\hat{\varphi} + \varepsilon\psi) \right|_{\varepsilon=0} = 2 \left\langle \hat{A}^* \left(\hat{\mathcal{A}}(\hat{\varphi}) - \tau \right) + \lambda_T \hat{\varphi}, \psi \right\rangle_H, \quad (9)$$

for any $\psi \in H^2[0, 1]$, where $\hat{A}^* = \mathcal{D}^{-1} \tilde{\hat{A}}$. Operator $\tilde{\hat{A}}$ is defined by $\tilde{\hat{A}}\psi(x) = \frac{1}{T\tau(1-\tau)} \sum_{t=1}^T I_t \hat{f}_{X,Y|Z}(x, \hat{\varphi}(x)|Z_t) \psi(Z_t)$ and \mathcal{D}^{-1} denotes the inverse of operator $\mathcal{D} : H_0^2[0, 1] \rightarrow L^2[0, 1]$ with $\mathcal{D} := 1 - \nabla^2$ and $H_0^2[0, 1] := \{\varphi \in H^2[0, 1] : \nabla\varphi(0) = \nabla\varphi(1) = 0\}$.

Operators \hat{A}^* and \tilde{A} are the empirical counterparts of A^* and \tilde{A} , which are the adjoint operators of A w.r.t. the Sobolev and L^2 scalar products on $H^2[0, 1]$, respectively, and are linked by $A^* = \mathcal{D}^{-1}\tilde{A}$ (see GS).⁵ From (9) $\hat{\varphi}$ satisfies the nonlinear integro-differential equation⁶

$$\hat{A}^* \left(\hat{\mathcal{A}}(\hat{\varphi}) - \tau \right) + \lambda_T \hat{\varphi} = 0. \quad (10)$$

5.2 Highlighting the nonlinearity issue

We can rewrite Equation (10) by using the second-order expansion $\hat{\mathcal{A}}(\hat{\varphi}) = \hat{\mathcal{A}}(\varphi_0) + \hat{A}_0 \Delta \hat{\varphi} + \hat{R}$, where $\hat{A}_0 \psi(z) = \int \hat{f}_{X,Y|Z}(x, \varphi_0(x)|z) \psi(x) dx$ and $\hat{R} := \hat{R}(\hat{\varphi}, \varphi_0)$ is the second order residual term (see Lemma A.4). Then, after rearranging, we get

$$\Delta \hat{\varphi} = \Delta \hat{\psi} + \hat{\mathcal{K}}_T(\Delta \hat{\varphi}), \quad (11)$$

where $\Delta \hat{\psi} = \hat{\psi} - \varphi_0$, with $\hat{\psi} := \left(\lambda_T + \hat{A}_0^* \hat{A}_0 \right)^{-1} \hat{A}_0^* \hat{r} - \left(\lambda_T + \hat{A}_0^* \hat{A}_0 \right)^{-1} \left(\hat{A}^* - \hat{A}_0^* \right) \left(\hat{\mathcal{A}}(\hat{\varphi}) - \tau \right) =: \hat{\psi}_1 + \hat{\psi}_2$ and $\hat{r} = \hat{A}_0 \varphi_0 + \tau - \hat{\mathcal{A}}(\varphi_0)$, $\hat{\mathcal{K}}_T(\Delta \hat{\varphi}) := - \left(\lambda_T + \hat{A}_0^* \hat{A}_0 \right)^{-1} \hat{A}_0^* \hat{R}$, and \hat{A}_0^* is defined as \hat{A}^* , but with φ_0 substituted for $\hat{\varphi}$.

This representation is instrumental in distinguishing the different contributing terms:

(i) The interpretation of $\hat{\psi}_1$ is as a linearized solution obtained from replacing the nonlinear equation $\hat{\mathcal{A}}(\varphi) = \tau$ with the linear integral type I equation $\hat{A}_0 \varphi \simeq \hat{r}$, and from applying

Tikhonov regularization to the linear proxy.

⁵ The boundary conditions $\nabla \varphi(0) = \nabla \varphi(1) = 0$ in the definition of $H_0^2[0, 1]$ are not restrictive since they concern the estimate $\hat{\varphi}$, whose properties are studied in L^2 or pointwise, but not the true function φ_0 . Propositions 3-5 below hold independently whether φ_0 satisfy these boundary conditions or not (see also Kress (1999), Theorem 16.20).

⁶ See e.g. Linton and Mammen (2005), (2006), Gagliardini and Gouriéroux (2007), and the survey by Carrasco, Florens and Renault (2005) for examples of estimation problems leading to linear integral equations of Type II.

(ii) Impact of nonlinearity is two-fold. On the one hand, we face the second order term \hat{R} in $\hat{\mathcal{K}}_T(\Delta\hat{\varphi})$. It is induced by the expansion of the nonlinear operator $\hat{\mathcal{A}}$. On the other hand, we face $\hat{A}^* - \hat{A}_0^*$ in $\hat{\psi}_2$. It is induced by the use of the estimate $\hat{\varphi}$ in the Frechet derivative \hat{A} in (8).

The key difference between our ill-posed setting and standard finite-dimensional parametric estimation problems, or well-posed functional estimation problems, concerns the behaviour (and the complex control) of the nonlinearity terms $\hat{\mathcal{K}}_T(\Delta\hat{\varphi})$ and $\hat{\psi}_2$. We prove in Appendix 4.3 (i) that $\hat{\mathcal{K}}_T(\Delta\hat{\varphi})$ satisfies a quadratic bound

$$\left\| \hat{\mathcal{K}}_T(\Delta\hat{\varphi}) \right\| \leq \frac{C}{\sqrt{\lambda_T}} \|\Delta\hat{\varphi}\|^2, \quad (12)$$

w.p.a. 1, with a suitable constant C . From the RHS of (12) we see that the coefficient of the quadratic bound is not fixed but rather is proportional to $1/\sqrt{\lambda_T}$. It diverges as the sample size increases. Hence, the usual argument that the quadratic nonlinearity term is negligible w.r.t. to the first-order term no matter the convergence rate of the latter, does not apply. Still, we can derive an asymptotic expansion for the MISE of the estimator $\hat{\varphi}$ in terms of $\Delta\hat{\psi}$ (see Appendix 4.3 (ii)-(iii)) via Equation (11):

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

Then, we need a suitable condition on the choice of the regularization parameter λ_T to ensure that $E \left[\|\Delta\hat{\psi}\|^2 \right]$ is the dominant term in such an expansion, and that the nonlinearity term $\hat{\psi}_2$ in $\Delta\hat{\psi}$ is negligible. This is made precise in the next section. Finally, $\Delta\hat{\psi}$ can be

decomposed as (see Appendix 4.4):

$$\Delta \hat{\psi} = (\lambda_T + A^*A)^{-1} A^* \hat{\zeta} + [(\lambda_T + A^*A)^{-1} A^*A - 1] \varphi_0 + \mathcal{R}_T, \quad (13)$$

where $\hat{\zeta} = \int \int (\tau - 1 \{y \leq \varphi_0(x)\}) \frac{\Delta \hat{f}_{X,Y,Z}(x, y, z)}{f_Z(z)} dy dx$ and the expression of the remainder term \mathcal{R}_T is given in (49) below. This corresponds to a Bahadur-type representation of the Q-TiR estimator (see e.g. Koenker (2005) for Bahadur-type representations in the finite-dimensional parameter case).

5.3 Mean Integrated Square Error

Proposition 3: *Suppose that Assumptions 1 and A hold. Let $h_T \rightarrow 0$ such that*

$$\frac{(\log T)^2}{Th_T^{2(d_Z+1)}} = O(1). \quad (14)$$

Further, let $\lambda_T \rightarrow 0$ such that for some $\varepsilon > 0$

$$\frac{(\log T)^2}{Th_T^{d_Z+1}} + h_T^m = o(\lambda_T b(\lambda_T)), \quad \frac{1}{Th_T^{\max\{d_Z, 2\}}} + h_T^{2m} = O(\lambda_T^{2+\varepsilon}), \quad (15)$$

and

$$\frac{1}{T} \sum_{j=1}^{\infty} \frac{\nu_j}{(\nu_j + \lambda_T)^2} \|\phi_j\|^2 + b(\lambda_T)^2 = o(\lambda_T), \quad (16)$$

*where $b(\lambda_T) = \|(\lambda_T + A^*A)^{-1} A^*A\varphi_0 - \varphi_0\|$, and ν_j and ϕ_j are the eigenvalues and eigenfunctions of A^*A , with $\|\phi_j\|_H = 1$. Then, up to negligible terms,*

$$E[\|\Delta \hat{\varphi}\|^2] = \frac{1}{T} \sum_{j=1}^{\infty} \frac{\nu_j}{(\nu_j + \lambda_T)^2} \|\phi_j\|^2 + b(\lambda_T)^2 =: V_T(\lambda_T) + b(\lambda_T)^2 =: M_T(\lambda_T). \quad (17)$$

Proof: Appendix 4.4.

The sufficient conditions (14)-(16) on h_T and λ_T ensure that the MISE of the estimator $\hat{\varphi}$ is asymptotically equivalent to the MISE of the linearization, and asymptotically equal to $M_T(\lambda_T)$ given in (17). The possibility to find bandwidth and regularization parameter sequences satisfying these restrictions is determined by an interplay between the smoothness properties of the parameter φ_0 , through the bias function $b(\lambda)$, and the severity of ill-posedness, through the decay behaviour of ν_j and $\|\phi_j\|^2$ (see Proposition 4 below). The set of Assumptions A in Appendix 1 used to prove Proposition 3 includes regularity conditions on the eigenfunctions of operator A^*A (Assumptions A.11-A.13), which are more restrictive (see also the discussion in Appendix 1) than the conditions used e.g. in Horowitz and Lee (2007), and Chen and Pouzo (2008). These stronger assumptions are required to derive the sharp asymptotic expansion of the MISE, which is a stronger result than the upper bounds on MISE rates derived in previous papers. Furthermore these stronger assumptions are also required to establish later the pointwise asymptotic normality result, which is new for the nonlinear ill-posed setting.

The expression of the asymptotic MISE in (17) is similar to the formula derived in GS for the MISE of a TiR estimator for NIVR, with operator A in (7) replacing the conditional expectation operator of X given Z . In $M_T(\lambda_T)$ we distinguish a regularization bias component $b(\lambda_T)^2$, and a variance component $V_T(\lambda_T)$. The bias $b(\lambda_T)$ converges to zero as $\lambda_T \rightarrow 0$. The formula of $V_T(\lambda_T)$ is reminiscent of the usual asymptotic variance of the quantile regression estimator: it involves the factor $f_{V|X,Z}(0|x, z)$ (see (2)) through the Frechet derivative A ,

and the factor $\tau(1 - \tau)$ through the adjoint A^* , hidden in the spectrum of A^*A .⁷

The optimal regularization sequence λ_T^* is defined by minimizing the asymptotic MISE $E[\|\Delta\hat{\varphi}\|^2]$ w.r.t. λ_T . The optimal MISE is denoted by M_T^* . The optimal sequence λ_T^* also corresponds to the minimizer of $M_T(\lambda_T)$ in (17), whenever the latter satisfies Conditions (14)-(16) for the validity of the asymptotic expansion. In the next section we illustrate the interplay between the smoothness of φ_0 and the severity of ill-posedness in a Gaussian example of nonparametric IV median regression. The assumptions in Horowitz and Lee (2007) do not cover the Gaussian case, which is known to yield a severely ill-posed estimation problem in NIVR.

5.4 A Gaussian example

Let us assume that variables X, U^*, Z admit a jointly normal distribution

$$\begin{pmatrix} X \\ U^* \\ Z \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho & \varrho \\ \rho & 1 & 0 \\ \varrho & 0 & 1 \end{pmatrix} \right),$$

with $\rho^2 + \varrho^2 < 1$, and define $U = \Phi(U^*) \sim \mathcal{U}(0, 1)$, where the function Φ denotes the c.d.f. of a standard Gaussian variable.⁸ We consider a separable specification $Y = \bar{\varphi}_0(X) + U^*$ and the case of a median regression, $\tau = 1/2$.

⁷ If the limit minimum distance criterion is not rescaled by factor $\tau(1-\tau)$, that is $Q_\infty(\varphi) = E[m(\varphi, Z)^2]$, the asymptotic MISE in (17) becomes $E[\|\Delta\hat{\varphi}\|^2] = \frac{\tau(1-\tau)}{T} \sum_{j=1}^{\infty} \frac{\nu_j}{(\nu_j + \lambda_T)^2} \|\phi_j\|^2 + b(\lambda_T)^2$ after appropriate redefinition of adjoint A^* and regularization parameter λ_T .

⁸ In this section, the variable X is defined on $\mathcal{X} = \mathbb{R}$. We can map it to $[0, 1]$ using a transformation with Φ . Other functions on \mathcal{X} , norms on \mathcal{X} and operators are transformed accordingly. We do not make this transformation explicit to simplify the notation.

From (7), the Frechet derivative operator is given by

$$\begin{aligned} A\psi(z) &= \phi(\Phi^{-1}(\tau)) \int_{\mathbb{R}} f_{X|Z,U}(x|z, \tau) \psi(x) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \frac{1}{\sqrt{1-\rho^2-\varrho^2}} \phi\left(\frac{x-\varrho z}{\sqrt{1-\rho^2-\varrho^2}}\right) \psi(x) dx, \end{aligned} \quad (18)$$

where ϕ denotes the p.d.f. of a standard Gaussian variable. It is a conditional expectation operator for the distribution of X given Z and $U = \tau$ (up to a multiplicative constant).

The spectral decomposition of A^*A admits simple expressions when the norms $\|\cdot\|$ and $\|\cdot\|_H$ on $[0, 1]$ correspond to norms $L^2(F_{X|U=\tau})$ and $H^2(F_{X|U=\tau})$ on \mathbb{R} . We derive it by adapting the standard result for the spectral decomposition of the conditional expectation operator of normal variables (see e.g. Carrasco, Florens, and Renault (2005)) to the case where the adjoint A^* is defined w.r.t. the Sobolev norm. We get:

$$\phi_j(x) = \sqrt{\frac{c^2}{c^2+j}} H_j\left(\frac{x}{c}\right), \quad \nu_j = \frac{2}{\pi} \frac{c^2}{c^2+j} \xi^{2j}, \quad j = 0, 1, \dots \quad (19)$$

where $c = \frac{\varrho}{\xi}$ and $\xi \in (0, 1)$ is such that $\frac{1-\xi^2}{\xi^2} = \frac{1-\rho^2-\varrho^2}{\varrho^2}$, and $H_j(\cdot)$ for $j = 0, 1, \dots$ denote the Hermite polynomials. Further, $\|\phi_j\|^2 = \frac{c^2}{c^2+j}$. Thus, the eigenvalues feature geometric decay $\nu_j \asymp \xi^{2j}$, and the eigenfunctions norms feature hyperbolic decay $\|\phi_j\|^2 \asymp j^{-1}$. The bias is given by

$$b(\lambda)^2 = \lambda^2 \sum_{j=1}^{\infty} \frac{d_j^2}{(\lambda + \nu_j)^2}, \quad d_j = \int_{\mathbb{R}} \bar{\varphi}_0(cx) H_j(x) \phi(x) dx. \quad (20)$$

If function $\bar{\varphi}_0$ is a polynomial (of any degree), then $b(\lambda) \asymp \lambda$ as $\lambda \rightarrow 0$.⁹

⁹ The notation $a(\lambda) \asymp b(\lambda)$ means that functions $a(\lambda)$ and $b(\lambda)$ are equivalent as $\lambda \rightarrow 0$ up to logarithmic terms, i.e. $c_1 [\log(1/\lambda)]^{c_3} \leq a(\lambda)/b(\lambda) \leq c_2 [\log(1/\lambda)]^{c_4}$ for some constants $0 \leq c_1 \leq c_2$ and $c_3 \leq c_4$.

We particularize the result in Proposition 3 for a spectrum of operator A^*A featuring decays as in (19), and a bias function behaving like a power of λ . For expository purpose, we consider $d_Z = 1$.

Proposition 4: *Suppose that Assumptions 1 and A hold. Further, suppose that $\nu_j \asymp e^{-\alpha j}$, $\|\phi_j\|^2 \asymp j^{-\beta}$, as $j \rightarrow \infty$, and $b(\lambda) \asymp \lambda^\delta$ as $\lambda \rightarrow 0$, for some $\alpha, \beta > 0$ and $\delta \in (0, 1]$.*

(i) *If*

$$1/2 < \delta \leq 1, \quad (21)$$

$$h_T \asymp T^{-\eta} \text{ with } \eta = \frac{1 + \delta}{2(1 + \delta + m)}, \text{ and}$$

$$\lambda_T \asymp T^{-\gamma} \text{ with } \gamma < \frac{m}{2(1 + \delta + m)}, \quad (22)$$

then

$$M_T(\lambda_T) \asymp \frac{1}{T\lambda_T [\log(1/\lambda_T)]^\beta} + \lambda_T^{2\delta}. \quad (23)$$

(ii) *If*

$$\frac{m + 2}{2(m + 1)} \leq \delta \leq 1, \quad (24)$$

and h_T is as above, then the optimal regularization parameter λ_T^ is such that:*

$$\lambda_T^* \asymp T^{-\gamma} \text{ with } \gamma = \frac{1}{1 + 2\delta}, \quad (25)$$

and the optimal MISE is

$$M_T^* \asymp T^{-\frac{2\delta}{1+2\delta}}. \quad (26)$$

Condition (21) guarantees that there exists a suitable window of regularization parameters (given in (22)), for which the asymptotic expansion of the MISE in (23) is valid.

Condition (22) is an upper bound on the rate of convergence of λ_T . The choice for the bandwidth h_T is in order to maximize the window of admissible regularization parameters in (22). Under the stronger Condition (24), we can derive the sequence of optimal regularization parameters and the optimal MISE, given in (25) and (26), respectively. Note that the minimal admissible regularity $\delta = \frac{m+2}{2(m+1)}$ depends on the order of differentiability of the joint density $f_{X,Y,Z}$, such that it approaches the limit $1/2$ as m increases.

Condition (24) is a joint constraint on the severity of ill-posedness and on the smoothness of function $\bar{\varphi}_0$. Given the geometric decay of the eigenvalues in (19), $\nu_j \asymp e^{-\alpha j}$ with $\alpha = 2 \log(1/\xi)$, a necessary condition is that the coefficients d_j of function $\bar{\varphi}_0(cx)$ in the Hermite basis share a geometric decay.¹⁰ If $d_j^2 \asymp e^{-\mu j}$ for some $\mu > 0$, we can show that:

$$b(\lambda) \asymp \lambda^\delta, \quad \delta = \min \left\{ \frac{\mu}{2\alpha}, 1 \right\}. \quad (27)$$

Condition (24) becomes

$$\mu \geq \alpha \frac{m+2}{m+1}. \quad (28)$$

If the function $\bar{\varphi}_0$ is such that $\|\nabla^j \bar{\varphi}_0\| \leq \kappa^j$, $j \in \mathbb{N}$, for a given $\kappa > 0$, we can prove that $d_j^2 \leq e^{-\mu j}$ for any $\mu > 0$ and large j , and (28) is satisfied.

The analytical study of the spectral decomposition of operator A^*A with general norms $\|\cdot\|$ and $\|\cdot\|_H$ and across probability levels $\tau \in (0, 1)$ is more complex. However, from the expression of the Frechet derivative in (18), we can make a couple of general remarks for the separable Gaussian case: (i) If function $\bar{\varphi}_0$ is such that $|\bar{\varphi}_0(-x)| = |\bar{\varphi}_0(x)|$, $x \in \mathbb{R}$,

¹⁰ To see this, suppose that $d_j^2 \geq cj^{-\mu}$ for some $c, \mu > 0$. Using $\nu_j \leq e^{-\alpha j}$, $\alpha = 2 \log(1/\xi)$, and defining $j(\lambda) \in \mathbb{N}$ such that $e^{-\alpha j(\lambda)} \asymp \lambda$ as $\lambda \rightarrow 0$, we have from (20) that $b(\lambda)^2 \geq Cj(\lambda)^{-\mu} \asymp [\log(1/\lambda)]^{-\mu}$.

the asymptotic MISE is symmetric w.r.t. $\tau \rightarrow 1 - \tau$. (ii) Two functions $\bar{\varphi}_0$ and $\tilde{\varphi}_0$ that differ by a constant, have the same regularity in the sense that they yield regularization bias functions with same behaviour as $\lambda \rightarrow 0$. (iii) By changing $\tau \in (0, 1)$, the regularity of $\varphi_0(x) = \bar{\varphi}_0(x) + \Phi^{-1}(\tau)$ is invariant and the impact on the MISE is through the density $f_{X|Z,U}(x|z, \tau)$, and thus through the spectrum of A^*A , only.

5.5 Asymptotic normality

In the next proposition we establish pointwise asymptotic normality of the Q-TiR estimator.

This result is induced by the Bahadur-type representation (13).

Proposition 5: *Suppose that Assumptions 1 and A hold, $\frac{(\log T)^2}{Th_T^{d_Z+1}} + h_T^m = O\left(\lambda_T^{1+\varepsilon/2}b(\lambda_T)\right)$,*

$$\frac{(\log T)^2}{Th_T^{2(1+d_Z)}} = O(1), \frac{1}{Th_T^{\max\{d_Z, 2\}}} + h_T^{2m} = O(\lambda_T^{2+\varepsilon}), T\lambda_T^3 = O(1), M_T(\lambda_T) = O(\lambda_T^{1+\varepsilon}),$$

$$\frac{M_T(\lambda_T)}{\sigma_T^2(x)/T} = o(\lambda_T^{-\varepsilon}), \quad (29)$$

for a $\varepsilon > 0$, where $\sigma_T^2(x) := \sum_{j=1}^{\infty} \frac{\nu_j}{(\lambda_T + \nu_j)^2} \phi_j^2(x)$. Further, suppose that for a $\bar{\varepsilon} > 0$ we have

$$\frac{1}{T^{1/3}} \frac{1}{\sigma_T^2(x)} \sum_{j=1}^{\infty} \frac{\nu_j}{(\lambda_T + \nu_j)^2} \phi_j^2(x) \|g_j\|_3^2 j^{1+\bar{\varepsilon}} = o(1), \quad (30)$$

where $\|g_j\|_3 := E[g_j(X, Y, Z)^3]^{1/3}$, $g_j(x, y, z) := \frac{1}{\tau(1-\tau)} (A\phi_j)(z)' (1\{y \leq \varphi_0(x)\} -$

$\tau)/\sqrt{\nu_j}$, and $\sqrt{h_T} \frac{1}{\sigma_T^2(x)} \sum_{j=1}^{\infty} \frac{\nu_j}{(\lambda_T + \nu_j)^2} = o(1)$. Then the Q-TiR estimator is asymptotically normal:

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

where $\mathcal{B}_T(x) = (\lambda_T + A^*A)^{-1} A^*A\varphi_0(x) - \varphi_0(x)$.

Proof: See Appendix 5.

Condition (29) requires that the rate of convergence of the variance $\sigma_T^2(x)/T$ at $x \in \mathcal{X}$ is not too large compared to the global rate of convergence of the MISE. Condition (30) is used to apply a Lyapunov CLT. When $\|g_j\|_3^2 j^{1+\varepsilon}$ diverges with j , Condition (30) is an upper bound on the rate of convergence of λ_T . Under an assumption of geometric spectrum for the eigenvalues ν_j , and hyperbolic behavior for the eigenfunction values $\phi_j^2(x)$ and for $\|g_j\|_3$, the arguments in the proof of Proposition 4 imply that (30) is satisfied whenever $\lambda_T \geq cT^{-\gamma}$ for some $c, \gamma > 0$. Proposition 5 shows that a pointwise nondegenerate limit distribution exists, a usual prerequisite for applying bootstrap.

6 Monte-Carlo results and empirical illustration

6.1 Computation of the estimator

To compute the estimate $\hat{\varphi}$, defined on a subset of the function space $H^2[0, 1]$, we use a numerical series approximation. We rely on standardized shifted Chebyshev polynomials of the first kind (see Section 22 of Abramowitz and Stegun (1970) for their mathematical properties). For example, if we take orders 0 to 5 this yields six coefficients ($k = 6$) to be estimated in the approximation $\varphi(x) \simeq \sum_{j=0}^{k-1} \theta_j P_j(x) =: \theta' P(x)$, where $P_0(x) = T_0^*(x)/\sqrt{\pi}$, $P_j(x) = T_j^*(x)/\sqrt{\pi/2}$, $j \neq 0$. The shifted Chebyshev polynomials of the first kind are $T_0^*(x) = 1$, $T_1^*(x) = -1 + 2x$, $T_2^*(x) = 1 - 8x + 8x^2$, $T_3^*(x) = -1 + 18x - 48x^2 + 32x^3$, $T_4^*(x) = 1 - 32x + 160x^2 - 256x^3 + 128x^4$, $T_5^*(x) = -1 + 50x - 400x^2 + 1120x^3 - 1280x^4 + 512x^5$. The squared Sobolev norm is approximated by $\|\varphi\|_H^2 = \int_0^1 \varphi^2 + \int_0^1 (\nabla\varphi)^2 \simeq \sum_{i=0}^{k-1} \sum_{j=0}^{k-1} \theta_i \theta_j \int_0^1 (P_i P_j + \nabla P_i \nabla P_j) =: \theta' D \theta$. The coefficients in this quadratic form are explicitly computed with a symbolic calculus package. The squared

L_2 norm $\|\varphi\|^2$ is approximated similarly by $\theta' B \theta$, say. Such simple and exact forms ease implementation ¹¹, and improve on computational speed. The convexity in θ (quadratic penalty) helps the numerical stability of the estimation procedure.

Estimated conditional probabilities $\hat{P}[Y \leq \varphi(X)|Z = z_t]$ used in the criterion are based on a Gaussian kernel smoother, and are approximated by

$$\frac{\sum_{l=1}^T IK\left(-\frac{y_l - \theta' P(x_l)}{h_Y}\right) K\left(\frac{z_l - z_t}{h_Z}\right)}{\sum_{l=1}^T K\left(\frac{z_l - z_t}{h_Z}\right)},$$

where $IK(x) = \int_{-\infty}^x K(u) du$ corresponds to the integrated kernel of K . This smoothing is asymptotically equivalent to the one described in Equation (5). We use it because of its numerical tractability: we avoid numerical integration without sacrificing differentiability (in a classical sense) of the approximated empirical criterion $Q_T(\theta' P)$ with respect to θ . Individual bandwidths are selected via the standard rule of thumb (Silverman (1986)). Analytical expressions for derivatives are computed and are implemented in a user-supplied analytical gradient and Hessian optimization procedure. In the separable case we take the NIVR estimates of GS on the same regularization parameter grid as starting values. In the nonseparable case we start from the true coefficients of the projection of φ_0 on the polynomial basis. If we use NIVR estimates instead we need to extend convergence time by a factor of at least five to achieve the same accuracy when $\tau \neq 1/2$.

The $k \times k$ matrix corresponding to operator $\hat{A}^* \hat{A}$ on the subspace spanned by the finite-

¹¹ The Gauss programs developed for this section and the empirical application are available on request from the authors.

dimensional basis of functions $\{P_j : j = 0, \dots, k-1\}$ in $H^2[0, 1]$ is given by $\langle P_i, \hat{A}^* \hat{A} P_j \rangle_H = \frac{1}{T\tau(1-\tau)} \sum_{t=1}^T (\hat{A} P_i)(Z_t) (\hat{A} P_j)(Z_t) = \frac{1}{T} (\hat{P}' \hat{P})_{i+1, j+1}$, $i, j = 0, \dots, k-1$, where \hat{P} is the $T \times k$ matrix with rows $\hat{P}(Z_t)' = (\tau(1-\tau))^{-1/2} \int P(x)' \hat{f}_{X,Y|Z}(x, \hat{\varphi}(x)|Z_t) dx$, $t = 1, \dots, T$. Hence we can use the suggestion of GS, which consists in estimating the asymptotic spectral representation (17) to select the regularization parameter. We need a first-step Q-TiR estimator $\bar{\theta}' P$ of φ_0 based on a pilot regularization parameter $\bar{\lambda}$. Then we can perform the spectral decomposition of the matrix $D^{-1} \hat{P}' \hat{P} / T$ to get the eigenvalues $\hat{\nu}_j$ and the eigenvectors \hat{w}_j , normalized to $\hat{w}_j' D \hat{w}_j = 1$, $j = 1, \dots, k$. Finally we can minimize an estimate of the MISE

$$\begin{aligned} \bar{M}(\lambda) = & \frac{1}{T} \sum_{j=1}^k \frac{\hat{\nu}_j}{(\lambda + \hat{\nu}_j)^2} \hat{w}_j' B \hat{w}_j \\ & + \bar{\theta}' \left[\frac{1}{T} \hat{P}' \hat{P} \left(\lambda D + \frac{1}{T} \hat{P}' \hat{P} \right)^{-1} - I \right] B \left[\left(\lambda D + \frac{1}{T} \hat{P}' \hat{P} \right)^{-1} \frac{1}{T} \hat{P}' \hat{P} - I \right] \bar{\theta}, \end{aligned}$$

w.r.t. λ to get the optimal regularization parameter $\hat{\lambda}$, and compute the second-step Q-TiR estimator with $\hat{\theta}$ using the regularization parameter $\hat{\lambda}$.

6.2 Monte-Carlo results

Following Newey and Powell (2003), the errors U_1^*, U_2^* and the instrument Z are jointly normally distributed, 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 $X = \Phi(X^*)$ and $U = \Phi(U_1^*)$.¹²

First we examine a separable design. **Case 1** is $Y = \sin(\pi X) + U_1^*$. The variable X is endogenous. The quantile condition is $P[Y - \varphi_0(X) \leq 0 | Z] = \tau$, where the functional

¹² This data generating process is a Gaussian model of the type considered in Section 5.4, but with the variable X explicitly transformed to live in $\mathcal{X} = [0, 1]$.

parameter is $\varphi_0(x) = \sin(\pi x) + \Phi^{-1}(\tau)$, $x \in [0, 1]$, for a given $\tau \in (0, 1)$. The chosen function resembles a concave Engel curve as in GS. **Case 2** is $Y = \Phi(3(X + U - 1))$, and the quantile structural effect is $\varphi_0(x) = \Phi(3(x + \tau - 1))$, $x \in [0, 1]$, for a given $\tau \in (0, 1)$. This is a nonseparable specification. We get a monotone increasing convex function in x for small τ , an S -shape for moderate τ , and a monotone increasing concave function for large τ . We climb towards the upper part of the normal c.d.f. by increasing τ , and gradually change the curvature of the function.

We consider sample size $T = 1000$. In Table 1 we compare the optimal asymptotic MISE and the optimal finite sample MISE of the Q-TiR estimator with $k = 6$ on 1000 replications. The asymptotic features are computed by Monte-Carlo integration using (17) and (18). The minimum of the finite sample MISE and of the asymptotic MISE are close for all probability levels. For $\tau = 1/2$ they are equal to .0133 and .0147, and correspond to a regularization parameter equal to .0006 and .0009. For other probability levels $\tau \in \{.10, .25, .75, .90\}$ we have an increase of the optimal finite sample and asymptotic MISE because of (equally) increasing bias and variance. This is a consequence of a vanishing density function, as we move into the tails, together with the invariant regularity of function φ_0 (Remark (iii) in Section 5.4). Symmetry with respect to $\tau = 1/2$ is explained by the symmetry of the Frechet derivative (Remark (i)). Optimal regularization parameters are (symmetrically) slightly decreasing as we depart from $\tau = 1/2$.

Table 1 also gathers results for estimation with the data driven procedure. We use $\bar{\lambda} = .0001$ as pilot regularization parameter. Other values such as .00005 or .0002 leave the results

qualitatively unchanged. We report the average and quartiles of the selected lambda over 1000 simulations, and the average ISE when we use the optimal data driven regularization parameter at each simulation. The selection procedure tends to slightly overpenalize in average, and the selected lambdas are positively skewed. However impact on the MISE of the data driven Q-TiR estimator is low since the average ISE are of the same magnitude as the optimal finite sample MISE. This is true across probability levels τ .

When we move to the nonseparable case, Table 2 shows that we need to penalize more. The order of magnitude of the optimal regularization parameter is 10^{-2} or 10^{-3} instead of 10^{-4} . We also observe a larger difference between the asymptotic and finite sample values of the MISE in relative terms, probably explained by the increased complexity of the model. However, the differences in absolute terms are comparable in Case 1 and Case 2. The performance of the data driven procedure is also comparable, and the selection rule continues to deliver good results.

6.3 Empirical illustration

This section presents an empirical illustration with U.S. long-distance call data¹³ 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. Their econometric methodology relies on parametric IVR and parametric IVQR using logarithm of annual income as instrument. We estimate quantile structural effects for non-

¹³ See e.g. Miravete (2002) and Economides, Seim and Viard (2008) for empirical studies on local telephone service demand.

linear pricing curves based on the conditional quantile condition $P[Y \leq \varphi_0(X) | Z] = \tau$, with $X = \Phi(X^*)$ and $Z = \Phi(Z^*)$. Variable Y denotes the price per minute in dollars, X^* denotes the standardized amount of use in minutes, and Z^* denotes the standardized logarithm of annual income. We consider the quartile structural effects, i.e., $\tau = \{.25, .5, .75\}$, and 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 ($T = 435$). The estimation procedure is as in the previous Monte-Carlo study and uses data-driven regularization parameters for each quantile structural effect and education category. We present our empirical results with eight polynomials ($k = 8$). We have checked that estimation results remain virtually unchanged when increasing gradually the number of polynomials up to sixteen. Setting in advance k large may raise numerical convergence issues when optimizing a nonlinear criterion. We have observed 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. We have also observed that 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. We use a pilot regularization parameter $\bar{\lambda} = .0001$ to get a first step estimator of φ_0 , and start the optimization algorithm with the data-driven NIVR estimates. To build pointwise confidence bands we use a nonparametric bootstrap procedure following Blundell, Chen and Kristensen (2007).

Figure 1 plots the estimated median structural effect and bootstrap pointwise confidence bands at 95% with 1000 replications for the two education categories. Figure 2 is a picture "à la box-plot" where we represent the estimated quantile structural effects at $\tau = \{.25, .5, .75\}$ and the estimated mean structural effect (NIVR estimate ¹⁴) for the two education categories. The box-plot interpretation is as follows. For any given value z of the instrument, the conditional probability of the shaded area is asymptotically $P[g(X, 0.25) \leq Y \leq g(X, 0.75) | Z = z] = .5$. Both figures show that the estimated structural effects are nonlinear, 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. Vertical sections of the shaded areas correspond to measures of dispersion. For high usage the dispersion is smaller for better educated people, and vice versa for low usage. Thus, better educated people seem to make a strong effort to find the more convenient tariffs only in the case of high usage of the service. Regularization parameters range from 10^{-2} to 10^{-1} . This exemplifies the need to opt for a data-driven procedure, and not a fixed value for all categories and probability levels τ . In Figure 3 we report the difference between the estimated quantile structural effect at $\tau = .75$ and at $\tau = .25$, i.e., the estimated interquartile range $\hat{g}(x, .75) - \hat{g}(x, .25)$ of the structural effect, together with bootstrap pointwise confi-

¹⁴ 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 for the less educated category and p -value = .77 for the better educated category).

dence bands at .95% with 1000 replications. If the model were separable as in the Gaussian example (Section 5.4) and Case 1 (Section 6.2) we would have a straight line. The plots point to a separable pricing curve for less educated people, and a nonseparable pricing curve for better educated people. A formal test of separability is left for future research.

References

- Abramowitz, M. and I. Stegun (1970): *Handbook of Mathematical Functions*, Dover Publications, New York.
- Adams, R. (1975): *Sobolev Spaces*, Academic Press, Boston.
- Ai, C. and X. Chen (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions", *Econometrica*, 71, 1795-1843.
- Alt, H. (1992): *Lineare Funktional Analysis*, Springer Verlag, Berlin.
- Andrews, D. (1994): "Empirical Process Methods in Econometrics", in the *Handbook of Econometrics*, Vol. 4, Engle, R. and D. McFadden, Eds., North Holland, 2247-2294.
- Blundell, R., Chen, X. and D. Kristensen (2007): "Semi-Nonparametric IV Estimation of Shape Invariant Engel Curves", *Econometrica*, 75, 1613-1669.
- Carrasco, M., Florens, J.-P. and E. Renault (2005): "Linear Inverse Problems in Structural Econometrics: Estimation Based on Spectral Decomposition and Regularization", *Handbook of Econometrics*, forthcoming.
- Chen, X. and D. Pouzo (2008): "Efficient Estimation of Semiparametric Conditional Moment Models with Possibly Nonsmooth Moments", Cowles Foundation Discussion Paper 1640.
- Chernozhukov, V. and C. Hansen (2005): "An IV Model of Quantile Treatment Effect", *Econometrica*, 73, 245-261.
- Chernozhukov, V., Imbens, G. and W. Newey (2007): "Instrumental Variable Estimation of Nonseparable Models", *Journal of Econometrics*, 139, 4-14.
- Chesher, A. (2003): "Identification in Nonseparable Models", *Econometrica*, 71, 1405-1441.
- Chesher, A. (2007): "Identification of Nonadditive Structural Functions", in *Advances in Economics and Econometrics, Theory and Applications: 9th World Congress of the Econometric Society*, Blundell, R., Persson, T. and W. Newey, Eds., Volume 3, Cambridge University Press, Cambridge.
- Darolles, S., Florens, J.-P. and E. Renault (2003): "Nonparametric Instrumental Regression", Working Paper.
- Economides, N., Seim, K. and B. Viard (2008): "Quantifying the Benefits of Entry into Local Phone Service", forthcoming in *RAND Journal of Economics*.

- Engl, H., Hanke, M. and A. Neubauer (2000): *Regularization of Inverse Problems*, Kluwer academic publishers, Dordrecht.
- Engl, H., Kunisch, K. and A. Neubauer (1989): "Convergence Rates for Tikhonov Regularization of Non-linear Ill-posed Problems", *Inverse problems*, 5, 523-540.
- Fitzenberger, B., Koenker, R. and A. Machado (2001): "Introduction", *Empirical Economics*, 26, 1-5.
- Gagliardini, P. and C. Gouriéroux (2007): "An Efficient Nonparametric Estimator for Models with Nonlinear Dependence", *Journal of Econometrics*, 137, 189-229.
- Gagliardini, P. and O. Scaillet (2006): "Tikhonov Regularization for Nonparametric Instrumental Variable Estimators", Working Paper.
- Gagliardini, P. and O. Scaillet (2007): "A Specification Test for Nonparametric Instrumental Variable Regression", Working Paper.
- Groetsch, C. (1984): *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*, Pitman Advanced Publishing Program, Boston.
- Hall, P. and J. Horowitz (2005): "Nonparametric Methods for Inference in the Presence of Instrumental Variables", *Annals of Statistics*, 33, 2904-2929.
- Hansen, B. (2007): "Uniform Convergence Rates for Kernel Estimation with Dependent Data", forthcoming in *Econometric Theory*.
- 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.
- Hofmann, B. and O. Scherzer (1998): "Local Ill-posedness and Source Conditions of Operator Equations in Hilbert Spaces", *Inverse Problems*, 14, 1189-1206.
- Horowitz, J. (2007): "Asymptotic Normality of a Nonparametric Instrumental Variables Estimator", *International Economic Review*, 48, 1329-1349.
- Horowitz, J. and S. Lee (2007): "Nonparametric Instrumental Variables Estimation of a Quantile Regression Model", *Econometrica*, 75, 1191-1208.
- Imbens, G. and W. Newey (2002): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity", Working Paper.
- Koenker, R. (2005): *Quantile Regression*, Econometric Society Monographs, 38, Cambridge University Press, Cambridge.
- Koenker, R. and G. Bassett (1978): "Regression Quantiles", *Econometrica*, 46, 33-50.

- Koenker, R. and I. Mizera (2004): "Penalized Triograms: Total Variation Regularization for Bivariate Smoothing", *Journal of the Royal Statistical Society, Series B*, 66, 145-163.
- Kress, R. (1999): *Linear Integral Equations*, Springer, New York.
- Linton, O. and E. Mammen (2005): "Estimating Semiparametric ARCH(∞) Models by Kernel Smoothing Methods", *Econometrica*, 73, 771-836.
- Linton, O. and E. Mammen (2006): "Nonparametric Transformation to White Noise", forthcoming in *Journal of Econometrics*.
- Miravete, E. (2002): "Estimating Demand for Local Telephone Service with Asymmetric Information and Optional Calling Plans", *Review of Economic Studies*, 69, 943-971.
- Newey, W. and J. Powell (2003): "Instrumental Variable Estimation of Nonparametric Models", *Econometrica*, 71, 1565-1578.
- Reed, M. and B. Simon (1980): *Functional Analysis*, Academic Press, San Diego.
- Schock, E. (2002): "Non-linear Ill-posed Equations: Counter-examples", *Inverse problems*, 18, 715-717.
- Silverman, B. (1986): *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.
- Tikhonov, A. N. (1963a): "On the Solution of Incorrectly Formulated Problems and the Regularization Method", *Soviet Math. Doklady*, 4, 1035-1038 (English Translation).
- Tikhonov, A. N. (1963b): "Regularization of Incorrectly Posed Problems", *Soviet Math. Doklady*, 4, 1624-1627 (English Translation).
- White, H. and J. Wooldridge (1991): "Some Results on Sieve Estimation with Dependent Observations", in *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, Proceedings of the Fifth International Symposium in Economic Theory and Econometrics, Cambridge University Press.

Appendices

In Appendix 1, we list the regularity conditions and provide their detailed discussion. In Appendix 2, we prove Proposition 1 on local ill-posedness of the nonparametric estimation of the endogenous quantile problem. In Appendix 3, we establish Proposition 2 on consistency of the Q-TiR estimator, and provide the auxiliary result of its existence. In Appendix 4, we prove Proposition 3 on the explicit expression of the asymptotic MISE, and provide the auxiliary steps leading to it. In Appendix 5 we show Proposition 5 on asymptotic normality of the Q-TiR estimator.

Appendix 1: List of regularity conditions

A.1: $\{(X_t, Y_t, Z_t) : t = 1, \dots, T\}$ is an i.i.d. sample from a distribution admitting a density $f_{X,Y,Z}$ with convex support $\mathcal{S} = \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \subset \mathbb{R}^d$, $\mathcal{X} = [0, 1]$, $\mathcal{Y} = [0, 1]$, $d = 2 + d_Z$.

A.2: The density $f_{X,Y,Z}$ of (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 $|\alpha| := \sum_{i=1}^d \alpha_i = m$.

A.3: (i) 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) $\sup_{x, z} f_{X|Z}(x|z) < \infty$, $\sup_{x, z} |\nabla_x f_{X|Z}(x|z)| < \infty$ and $\sup_{x, z} |\nabla_z f_{X|Z}(x|z)| < \infty$; (iii) $\sup_{u, x, z} |\nabla_x F_{U|X,Z}(u|x, z)| < \infty$ and $\sup_{u, x, z} |\nabla_z F_{U|X,Z}(u|x, z)| < \infty$; (iv) $\sup_{u, x, z} f_{U|X,Z}(u|x, z) < \infty$.

A.4: There exists $h > 0$ such that function $q(s) := \sup_{v \in B_h(s)} |\nabla f_{X,Y,Z}(v)|$, $s \in \mathcal{S}$, is integrable

w.r.t. Lebesgue measure on \mathcal{S} , where $B_h(s)$ denotes the ball in \mathbb{R}^d of radius h around s .

A.5: There exists $h > 0$ such that function $q_\alpha(s) := \sup_{v \in B_h(s)} |\nabla^\alpha f_{X,Y,Z}(v)|$, $s \in \mathcal{S}$, satisfies

$$\int_{\mathcal{S}} \frac{q_\alpha(s)^2}{f_{X,Y,Z}(s)} ds < \infty, \text{ for any } \alpha \in \mathbb{N}^d \text{ with } |\alpha| = m.$$

A.6: (i) $\sup_{x,z} f_{X,Y|Z}(x, \varphi_0(x)|z) < \infty$; (ii) $\sup_{x,y,z} |\nabla_y f_{X,Y|Z}(x, y|z)| < \infty$.

A.7: Set $\mathcal{Z}_T \subset \mathcal{Z}$ is such that $\sup_{z \in \mathcal{Z}_T} |z| = O(T^b)$ for $b < \infty$, $P(Z \in \mathcal{Z}_T^c) = O(T^{-\bar{b}})$ for any $\bar{b} > 0$ and $\inf_{z \in \mathcal{Z}_T} f_Z(z) \geq 2(\log T)^{-1}$.

A.8: We have: $\frac{20}{\tau(1-\tau)} \sup_{x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}} |\nabla_y f_{X,Y|Z}(x, y|z)|^2 < \frac{1}{\|\varphi_0\|_H^2} \inf_{\psi: \|\psi\|_H \leq 2\|\varphi_0\|_H} \frac{\langle \psi, A^* A \psi \rangle_H}{\|\psi\|^2}$.

A.9: Function φ_0 is an interior point of Θ w.r.t. $\|\cdot\|$.

A.10: 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 $|\alpha| < m$, where m is as in Assumption A.2.

A.11: The $\langle \cdot, \cdot \rangle_H$ -orthonormal basis of eigenfunctions $\{\phi_j : j \in \mathbb{N}\}$ of operator $A^* A$ satisfies (i) $\sum_{j=1}^{\infty} \|\phi_j\| < \infty$; (ii) $\sum_{j,l=1, j \neq l}^{\infty} \frac{\langle \phi_j, \phi_l \rangle^2}{\|\phi_j\|^2 \|\phi_l\|^2} < \infty$.

A.12: Functions $\psi_j(z) := \frac{1}{\sqrt{\nu_j}} (A\phi_j)(z)$, $j \in \mathbb{N}$, satisfy $\sup_{j \in \mathbb{N}} E[\omega(Z) \psi_j(Z)^2] < \infty$ and

$$\sup_{j \in \mathbb{N}} E\left[|\psi_j(Z)|^{\bar{s}}\right] < \infty, \text{ for } \bar{s} > 2, \text{ where } \omega(z) := \frac{\int q(w, z) dw}{f_Z(z)} \text{ for } w := (x, y) \text{ and } q \text{ as in}$$

Assumption A.4.

A.13: Functions ψ_j are in class $C^m(\mathbb{R}^{dz})$ such that (i) $\sup_{j \in \mathbb{N}} E\left[|\nabla \psi_j(Z)|^2\right] < \infty$

and $\sup_{j \in \mathbb{N}} E \left[\omega(Z) |\nabla \psi_j(Z)|^2 \right] < \infty$; (ii) $\sup_{j \in \mathbb{N}} E \left[|\nabla^\alpha \psi_j(Z)|^2 \right] < \infty$ and $\sup_{j \in \mathbb{N}} E \left[|\nabla^\alpha \psi_j(Z - \zeta) - \nabla^\alpha \psi_j(Z)|^2 \right] \rightarrow 0$ as $|\zeta| \rightarrow 0$, for any $\alpha \in \mathbb{N}^{dz}$ with $|\alpha| = m$.

In Assumption A.1, the compact support of X and Y is used for technical reasons. Mapping in $[0, 1]$ can be achieved by simple linear or nonlinear monotone transformations. Assuming univariate X simplifies the exposition. Assumptions A.2 and A.10 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. It is possible to reformulate the list of technical assumptions in terms of high-level conditions concerning the nonparametric estimation of $f_{X,Y,Z}$. This facilitates the extension to other types of smoothing methods, but requires listing some technical conditions to get the results on the MISE and asymptotic normality, which are difficult to interpret at the generic level. This also complicates the separation of the regularity conditions on the nonparametric smoothing parameter and the regularization parameter.

Assumptions A.3 (i)-(iii) are used to prove local ill-posedness (Proposition 1). Specifically, Assumption A.3 (i) is a boundedness and smoothness condition on function $g(x, \tau)$ w.r.t. both its arguments. In particular, it implies that the structural quantile effect $\varphi_0 \in H^2[0, 1]$, for any $\tau \in (0, 1)$. Assumptions A.3 (ii) and (iii) concern boundedness and smoothness of the densities of X given Z , and of U given X, Z , respectively. Assumption A.3 (iv) also concerns the density of U given X, Z and is used in the proof of consistency

(Proposition 2). The set \mathcal{Z}_T in Assumption A.7 is used to introduce a trimming for small values of $f_Z(z)$. Assumptions A.1-A.3, A.7 and A.10 imply the consistency of the Q-TiR estimator.

The remaining assumptions are used to derive the exact asymptotic expansion of the MISE and prove asymptotic normality. Specifically, Assumptions A.4 and A.5 impose integrability conditions on suitable measures of local variation of density $f_{X,Y,Z}$. These assumptions are used in the proof of Lemmas A.11 and A.12 to bound higher order terms in the asymptotic expansion of the MISE coming from kernel estimation bias. Assumption A.6 is used to show that \mathcal{A} is Frechet differentiable, with compact Frechet derivative. This assumption can be rewritten in terms of densities $f_{U|X,Z}$, $f_{X|Z}$ and function g . The formulation as in A.6 is closer to the use in the proofs, and simplifies the exposition. In Assumption A.8, $\sup_{x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}} |\nabla_y f_{X,Y|Z}(x, y|z)|$ is involved in the bound of the quadratic term in the expansion of \mathcal{A} (see Lemma A.3 in Appendix 4), and controls for the amount of nonlinearity in the conditional moment restrictions. Thus, Assumption A.8 is a joint restriction on the amount of nonlinearity, on the severity of ill-posedness, and on the Sobolev norm of the functional parameter. In particular, for given joint density of X, Y, Z , Assumption A.8 is satisfied if $\|\varphi_0\|_H$ is small enough. Assumption A.8 is used to bound some residual terms in the asymptotic expansion of the estimator, involving probabilities of large deviations (see Lemma A.10 in Appendix 4). Assumption A.9 is used to derive the first-order condition satisfied by estimator $\hat{\varphi}$. Finally, the last three Assumptions A.11-A.13 concern the spectrum of operator A^*A . Assumption A.11 (i) is used to simplify the proof of Lemmas A.14, A.16 and A.17 in

order to get asymptotic normality. Assumption A.11 (ii) requires that the eigenfunctions of operator A^*A , which are orthogonal w.r.t. $\langle \cdot, \cdot \rangle_H$, satisfy a summability condition w.r.t. $\langle \cdot, \cdot \rangle$. Under this Assumption, the asymptotic expansion of the MISE in Proposition 3 involves a single sum, and not a double sum, over the spectrum. Assumptions A.12 and A.13 ask for the existence of a uniform bound for moments of derivatives of functions $\psi_j(z) = \frac{1}{\sqrt{\nu_j}} (A\phi_j)(z)$, $j \in \mathbb{N}$, both under density f_Z and under the density defined by function q in Assumption A.4. Functions ψ_j satisfy $E[\psi_j(Z)^2] = 1$, $j \in \mathbb{N}$. Assumptions A.12 and A.13 are met whenever the functions ψ_j and their derivatives do not exhibit too heavy tails. These assumptions are used to control terms of the type $\int \psi_j(z) [1\{y \leq \varphi_0(x)\} - \tau] \hat{f}_{X,Y,Z}(r) dr$, uniformly in $j \in \mathbb{N}$, in Lemma A.11 and in the proof of Lemma A.14.

Appendix 2: Proof of Proposition 1

Here we show local ill-posedness of the nonseparable setting (part (a)), and we prove that sequences generating ill-posedness exhibit diverging L^2 -norm of their first derivative (part (b)). To prove part (a), we need the following Lemma A.1, which is a local version of Proposition 10.1 in Engl, Hanke and Neubauer (2000).

Lemma A.1: *Suppose the following conditions are satisfied: (a) Operator \mathcal{A} is compact. (b) For any $r > 0$ small enough, there exists a sequence $(\varphi_n) \subset B_r(\varphi_0)$ s.t. $\varphi_n \not\rightarrow \varphi_0$ and $\mathcal{A}(\varphi_n) \xrightarrow{w} \mathcal{A}(\varphi_0)$ where \xrightarrow{w} denotes weak convergence. Then the minimum distance problem is locally ill-posed.*

Proof of Proposition 1 (a): We establish that the conditions (a) and (b) in Lemma A.1 are satisfied. (a) We have to show that \mathcal{A} maps closed sets into relatively compact sets. Let $S \subset L^2[0, 1]$ be bounded. We have to prove that the closure of $\mathcal{A}(S) \subset L^2(F_Z, \tau)$ is compact. We can equivalently use $\|\cdot\|_{L^2(F_Z, \tau)}$ or $\|\cdot\|_{L^2(F_Z)}$. Proposition 2.24 in Alt (1992) states that $\mathcal{A}(S)$ is relatively compact if and only if:

$$\sup_{\varphi \in S} \|\mathcal{A}(\varphi)\|_{L^2(F_Z)} < \infty, \quad (31)$$

$$\sup_{\varphi \in S} \|\mathcal{A}(\varphi)(\cdot + h) - \mathcal{A}(\varphi)\|_{L^2(F_Z)} \rightarrow 0, \quad \text{as } |h| \rightarrow 0, \quad (32)$$

and

$$\sup_{\varphi \in S} \|\mathcal{A}(\varphi) \cdot \chi_{\mathbb{R}^{d_Z} \setminus B_R(0)}\|_{L^2(F_Z)} \rightarrow 0, \quad \text{as } R \nearrow \infty, \quad (33)$$

where $\chi_{\mathbb{R}^{d_Z} \setminus B_R(0)}(z) := 1 \{z \in \mathbb{R}^{d_Z} \setminus B_R(0)\}$, and $B_R(0)$ is a ball in \mathbb{R}^{d_Z} of radius R around 0. To prove (31), notice that for any z

$$|\mathcal{A}(\varphi)(z)| = \int_{\mathcal{X}} f_{X|Z}(x|z) F_{U|X,Z}(g^{-1}(x, \varphi(x)) | x, z) dx \leq \int_{\mathcal{X}} f_{X|Z}(x|z) dx = 1. \quad (34)$$

Thus $\|\mathcal{A}(\varphi)\|_{L^2(F_Z)} \leq 1$, for any $\varphi \in L^2[0, 1]$, and (31) follows. To prove (32) we use

$$\begin{aligned} & |\mathcal{A}(\varphi)(z+h) - \mathcal{A}(\varphi)(z)| \\ & \leq \int_{\mathcal{X}} |f_{X|Z}(x|z+h) - f_{X|Z}(x|z)| F_{U|X,Z}(g^{-1}(x, \varphi(x)) | x, z+h) dx \\ & \quad + \int_{\mathcal{X}} f_{X|Z}(x|z) |F_{U|X,Z}(g^{-1}(x, \varphi(x)) | x, z+h) - F_{U|X,Z}(g^{-1}(x, \varphi(x)) | x, z)| dx \\ & \leq C |h|, \end{aligned}$$

where $C := \sup_{u, x, z} |\nabla_z F_{U|X,Z}(u|x, z)| f_{X|Z}(x|z) + \sup_{x, z} |\nabla_z f_{X|Z}(x|z)| < \infty$ from Assumptions

A.3 (ii) and (iii). Thus we get

$$\|\mathcal{A}(\varphi)(\cdot + h) - \mathcal{A}(\varphi)\|_{L^2(F_Z)} \leq C |h| \rightarrow 0 \text{ as } h \rightarrow 0, \quad (35)$$

uniformly in $\varphi \in L^2[0, 1]$. Thus, (32) is proved. Finally, from (34) we get that for $\varphi \in L^2[0, 1]$

$$\|\mathcal{A}(\varphi) \cdot \chi_{\mathbb{R}^{d_Z} \setminus B_R(0)}\|_{L^2(F_Z)}^2 \leq \int_{\mathbb{R}^{d_Z} \setminus B_R(0)} f_Z(z) dz \rightarrow 0 \text{ as } R \nearrow \infty.$$

This implies (33) and that \mathcal{A} is compact.

(b) 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 $((k-1)/n, k/n)$ with $k = 1, \dots, n$. 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

$$\|\varphi_n - \varphi_0\|^2 \rightarrow \int_{\mathcal{X}} \int_0^1 [g(x, \tau + \varepsilon\psi(y)) - g(x, \tau)]^2 dy dx.$$

The RHS is strictly larger than zero, and converges to zero as $\varepsilon \rightarrow 0$ by the dominated convergence Theorem. Thus, for $\varepsilon > 0$ sufficiently small, we have $(\varphi_n) \subset B_r(\varphi_0)$ and $\varphi_n \not\rightarrow \varphi_0$.

Moreover, for $\bar{q} \in L^2(F_Z, \tau)$ we have

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

Thus, we have to show

$$J_n := \int \bar{q}(z) f_Z(z) \int_{\mathcal{X}} f_{X|Z}(x|z) F_{U|X,Z}(\tau + \psi_n(x) | x, z) dx dz \rightarrow \tau \int \bar{q}(z) f_Z(z) dz. \quad (36)$$

To this end, split the integral w.r.t. x over the partition $((k-1)/n, k/n)$ with $k = 1, \dots, n$

$$\begin{aligned} J_n &= \sum_{k=1}^n \int_{(k-1)/n}^{k/n} \int \bar{q}(z) f_Z(z) f_{X|Z}(x|z) F_{U|X,Z}(\tau + \varepsilon \psi(nx) | x, z) dz dx \\ &= \sum_{k=1}^n \frac{1}{n} \int_0^1 \int \bar{q}(z) f_Z(z) f_{X|Z}\left(\frac{k-1}{n} + \frac{1}{n}y | z\right) F_{U|X,Z}\left(\tau + \varepsilon \psi(y) | \frac{k-1}{n} + \frac{1}{n}y, z\right) dz dy \end{aligned}$$

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

$$J_n = \sum_{k=1}^n \frac{1}{n} \int \bar{q}(z) f_Z(z) f_{X|Z}\left(\frac{k-1}{n} | z\right) \int_0^1 F_{U|X,Z}\left(\tau + \varepsilon \psi(y) | \frac{k-1}{n}, z\right) dy dz + I_{1,n}, \quad (37)$$

where

$$|I_{1,n}| \leq \sum_{k=1}^n \frac{1}{n^2} \int_0^1 \int \bar{q}(z) f_Z(z) \sup_{u,x,z} |\nabla_x H(u, x|z)| y dz dy = O(1/n),$$

and $H(u, x|z) := F_{U|X,Z}(u|x, z) f_{X|Z}(x|z)$, $\sup_{u,x,z} |\nabla_x H(u, x|z)| < \infty$ from Assumptions A.3 (ii)-(iii). Since the Riemann sum in (37) converges to the corresponding integral, we get

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

Using that $\int_{\mathcal{X}} f_{X|Z}(x|z) F_{U|X,Z}(u|x, z) dx = P[U \leq u|z] = u$ 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 (36) follows. ■

Proof of Proposition 1 (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 (1975)), set $\{\varphi \in \Theta : \|\nabla\varphi\| \leq B\}$ is compact w.r.t. the norm $\|\cdot\|$. Therefore, there exists a subsequence (φ_{m_n}) which converges in norm $\|\cdot\|$ to $\varphi^* \in \Theta$, say. Since Q_∞ is continuous, we get $Q_\infty(\varphi_{m_n}) \rightarrow Q_\infty(\varphi^*)$, and thus $Q_\infty(\varphi^*) = 0$. By identification (Assumption 1), we deduce $\varphi^* = \varphi_0$, and the subsequence (φ_{m_n}) converges to φ_0 . But this is impossible, since $\|\varphi_{m_n} - \varphi_0\| \geq \varepsilon > 0$. ■

Appendix 3: Proof of Proposition 2

We establish existence of the Q-TiR estimator in A.3.1 before showing its consistency in A.3.2.

A.3.1 Existence

Since $Q_T(\varphi) = \frac{1}{T\tau(1-\tau)} \sum_{t=1}^T \hat{m}(\varphi, Z_t)^2 I_t$ is positive, a function $\hat{\varphi} \in \Theta$ minimizes $Q_T(\varphi) + \lambda_T \|\varphi\|_H^2$ if and only if

$$\hat{\varphi} = \arg \inf_{\varphi \in \Theta} Q_T(\varphi) + \lambda_T \|\varphi\|_H^2, \quad \text{s.t.} \quad \lambda_T \|\varphi\|_H^2 \leq L_T(\varphi_0). \quad (38)$$

The solution $\hat{\varphi}$ in (38) exists P -a.s. since (i) mapping $\varphi \rightarrow \|\varphi\|_H^2$ is lower semicontinuous on $H^2[0, 1]$ w.r.t. the norm $\|\cdot\|$ (see Reed and Simon (1980), p. 358) and mapping $\varphi \rightarrow Q_T(\varphi)$ is continuous on Θ w.r.t. the norm $\|\cdot\|$, P -a.s., for any T ; (ii) set $\{\varphi \in \Theta : \|\varphi\|_H^2 \leq \bar{L}\}$ is compact w.r.t. the norm $\|\cdot\|$, for any constant $0 < \bar{L} < \infty$ (compact embedding theorem; see Adams (1975)).

The continuity of $Q_T(\varphi)$, P -a.s., follows from the mapping $\varphi \rightarrow \hat{m}(\varphi, z)$ being continuous

for almost any $z \in \mathcal{Z}$, P -a.s.. The latter holds since for any $\varphi_1, \varphi_2 \in \Theta$,

$$\begin{aligned} & \left| \hat{f}_{X|Z}(x|z) \right| \left| \hat{F}_{Y|X,Z}(\varphi_1(x)|x, z) - \hat{F}_{Y|X,Z}(\varphi_2(x)|x, z) \right| \\ = & \left| \int_{\varphi_2(x)}^{\varphi_1(x)} \hat{f}_{X,Y|Z}(x, y|z) dy \right| \leq \left(\sup_{x \in [0,1], y \in \mathbb{R}} \left| \hat{f}_{X,Y|Z}(x, y|z) \right| \right) |\varphi_1(x) - \varphi_2(x)|, \end{aligned}$$

and thus, by the Cauchy-Schwarz inequality,

$$\begin{aligned} |\hat{m}(\varphi_1, z) - \hat{m}(\varphi_2, z)| & \leq \int_{\mathcal{X}} \left| \hat{f}_{X|Z}(x|z) \right| \left| \hat{F}_{Y|X,Z}(\varphi_1(x)|x, z) - \hat{F}_{Y|X,Z}(\varphi_2(x)|x, z) \right| dx \\ & \leq \bar{C}_T \|\varphi_1 - \varphi_2\|, \end{aligned}$$

for almost any $z \in \mathcal{Z}$, P -a.s., where $\bar{C}_T := \sup_{x \in [0,1], y \in \mathbb{R}} \left| \hat{f}_{X,Y|Z}(x, y|z) \right| < \infty$ for almost any $z \in \mathcal{Z}$, P -a.s.

A.3.2 Consistency

The next Lemma A.2 is used below in the proof of Proposition 2 (consistency). Lemma A.2

(i) is proved in the Technical Report by extending an argument in Hansen (2007). Lemma

A.2 (ii) and (iii) establish (uniform) convergence of the minimum distance criterion $Q_T(\varphi)$.

Lemma A.2: *Under Assumptions A.1, A.2, A.3 (ii)-(iv), A.7, and A.10:*

$$\begin{aligned} \text{(i)} \quad & \sup_{x \in [0,1], y \in \mathbb{R}, z \in \mathcal{Z}_T} \left| \hat{f}_{X|Z}(x|z) \hat{F}_{Y|X,Z}(y|x, z) - f_{X|Z}(x|z) F_{Y|X,Z}(y|x, z) \right|^2 = O_p(a_T), \text{ where } a_T := \\ & (\log T)^2 \left(\frac{\log T}{T h_T^{d_Z+1}} + h_T^{2m} \right); \text{ (ii) } Q_T(\varphi_0) - Q_\infty(\varphi_0) = O_p(a_T); \text{ (iii) } \sup_{\varphi \in \Theta} |Q_T(\varphi) - Q_\infty(\varphi)| = \\ & O_p \left(\sqrt{a_T} + \frac{1}{\sqrt{T}} \right) = o_p(1) \text{ for } a_T = o(1). \end{aligned}$$

Proof of Lemma A.2: (ii) We have $Q_T(\varphi_0) - Q_\infty(\varphi_0) = \frac{1}{T\tau(1-\tau)} \sum_{t=1}^T \Delta \hat{m}(\varphi_0, Z_t)^2 I_t$,

where $\Delta\hat{m}(\varphi, \cdot) := \hat{m}(\varphi, \cdot) - m(\varphi, \cdot)$. Furthermore,

$$\begin{aligned} |\Delta\hat{m}(\varphi, \cdot)| &\leq \int_{\mathcal{X}} \left| \hat{f}_{X|Z}(x|\cdot)\hat{F}_{Y|X,Z}(\varphi(x)|x, \cdot) - f_{X|Z}(x|\cdot)F_{Y|X,Z}(\varphi(x)|x, \cdot) \right| dx \\ &\leq \sup_{x \in [0,1], y \in \mathbb{R}} \left| \hat{f}_{X|Z}(x|\cdot)\hat{F}_{Y|X,Z}(y|x, \cdot) - f_{X|Z}(x|\cdot)F_{Y|X,Z}(y|x, \cdot) \right|, \end{aligned} \quad (39)$$

uniformly in $\varphi \in \Theta$. Then, (ii) follows from (i).

(iii) Using $\hat{m}(\varphi, \cdot) = \Delta\hat{m}(\varphi, \cdot) + m(\varphi, \cdot)$, we have

$$\begin{aligned} Q_T(\varphi) - Q_\infty(\varphi) &= \frac{1}{T\tau(1-\tau)} \sum_{t=1}^T \Delta\hat{m}(\varphi, Z_t)^2 I_t + \left\{ \frac{1}{T\tau(1-\tau)} \sum_{t=1}^T m(\varphi, Z_t)^2 I_t - Q_\infty(\varphi) \right\} \\ &\quad + 2 \frac{1}{T\tau(1-\tau)} \sum_{t=1}^T \Delta\hat{m}(\varphi, Z_t) m(\varphi, Z_t) I_t. \end{aligned}$$

From (39) and (i), the first term in the RHS is $O_p(a_T)$, uniformly in $\varphi \in \Theta$. By Cauchy-Schwarz inequality, the third term in the RHS is $O_p\left(\sqrt{a_T} \left(\frac{1}{T\tau(1-\tau)} \sum_{t=1}^T m(\varphi, Z_t)^2 I_t\right)^{1/2}\right)$, uniformly in $\varphi \in \Theta$. Thus, the conclusion follows if we show that

$$\sup_{\varphi \in \Theta} \left| \frac{1}{T\tau(1-\tau)} \sum_{t=1}^T m(\varphi, Z_t)^2 I_t - Q_\infty(\varphi) \right| = O_p\left(\frac{1}{\sqrt{T}}\right). \text{ We have:}$$

$$\begin{aligned} \left| \frac{1}{T\tau(1-\tau)} \sum_{t=1}^T m(\varphi, Z_t)^2 I_t - Q_\infty(\varphi) \right| &\leq \frac{1}{T\tau(1-\tau)} \sum_{t=1}^T m(\varphi, Z_t)^2 (1 - I_t) \\ &\quad + \left| \frac{1}{T\tau(1-\tau)} \sum_{t=1}^T m(\varphi, Z_t)^2 - Q_\infty(\varphi) \right| \\ &=: I_{1,T}(\varphi) + I_{2,T}(\varphi). \end{aligned}$$

Since $|m(\varphi, \cdot)| \leq 2$, the $I_{1,T}(\varphi)$ term is bounded by $\frac{4}{T\tau(1-\tau)} \sum_{t=1}^T (1 - I_t)$, uniformly in $\varphi \in \Theta$. Now, $E\left[\frac{1}{T} \sum_{t=1}^T (1 - I_t)\right] \leq P[Z \in \mathcal{Z}_T^c] + P\left[\inf_{z \in \mathcal{Z}_T} \hat{f}(z) \leq (\log T)^{-1}\right] = O\left(T^{-\bar{b}}\right)$, for any $\bar{b} > 0$, from Assumption A.7 and a large deviation bound argument as in the proof of Lemma

A.2 (i). We get $\sup_{\varphi \in \Theta} I_{1,T}(\varphi) = O_p\left(T^{-\bar{b}}\right)$, for any $\bar{b} > 0$. To bound the $I_{2,T}(\varphi)$ term, we use

$$\begin{aligned}
m(\varphi, z) &= \int_{\mathcal{X}} [F_{U|X,Z}(g^{-1}(x, \varphi(x)) | x, z) - \tau] f_{X|Z}(x|z) dx. \text{ Then} \\
&\frac{1}{T} \sum_{t=1}^T m(\varphi, Z_t)^2 - E[m(\varphi, Z)^2] \\
&= \int_{\mathcal{X}} \int_{\mathcal{X}} \frac{1}{T} \sum_{t=1}^T \{f_{X|Z}(x|Z_t) f_{X|Z}(\xi|Z_t) \\
&\quad [F_{U|X,Z}(g^{-1}(x, \varphi(x)) | x, Z_t) - \tau] [F_{U|X,Z}(g^{-1}(\xi, \varphi(\xi)) | \xi, Z_t) - \tau] \\
&\quad - E[f_{X|Z}(x|Z) f_{X|Z}(\xi|Z)] \\
&\quad [F_{U|X,Z}(g^{-1}(x, \varphi(x)) | x, Z) - \tau] [F_{U|X,Z}(g^{-1}(\xi, \varphi(\xi)) | \xi, Z) - \tau]\} dx d\xi.
\end{aligned}$$

We get $\sup_{\varphi \in \Theta} I_{2,T}(\varphi) \leq \frac{1}{\tau(1-\tau)\sqrt{T}} \sup_{\varrho \in [0,1]^4} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^T (a(Z_t, \varrho) - E[a(Z, \varrho)]) \right|$, where $a(z, \varrho) := f_{X|Z}(x|z) f_{X|Z}(\xi|z) [F_{U|X,Z}(u | x, z) - \tau] [F_{U|X,Z}(v | \xi, z) - \tau]$, $\varrho := (x, \xi, u, v) \in [0, 1]^4$. Using Assumptions A.3 (ii)-(iv), function a is bounded and Lipschitz w.r.t. ϱ : $|a(\cdot, \varrho_1) - a(\cdot, \varrho_2)| \leq C |\varrho_1 - \varrho_2|$, for a constant C . By Andrews (1994), Theorem 2, the family $\mathcal{F} := \{a(\cdot, \varrho) : \varrho \in [0, 1]^4\}$ satisfies the Pollard entropy condition. By Andrews (1994), Theorem 1, the empirical process $\nu_T(\varrho) := \frac{1}{\sqrt{T}} \sum_{t=1}^T (a(Z_t, \varrho) - E[a(Z, \varrho)])$, $\varrho \in [0, 1]^4$, is stochastically equicontinuous. Since we can apply a CLT for any $\varrho \in [0, 1]^4$, by the fundamental convergence result for empirical processes (see e.g. Andrews (1994), p. 2251), $\nu_T(\cdot)$ converges weakly. By the Continuous Mapping Theorem, $\sup_{\varrho \in [0,1]^4} |\nu_T(\varrho)| = O_p(1)$, and thus $\sup_{\varphi \in \Theta} I_{2,T}(\varphi) = O_p(1/\sqrt{T})$. Hence, the conclusion follows. ■

Proof of Proposition 2: By Lemma A.2 (ii) and the condition on λ_T , we have

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

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 $\hat{\varphi} \in K_\delta$ wp $\gtrsim 1 - \delta$, where notation wp $\gtrsim 1 - \delta$ means "with probability at least $1 - \delta$ for all sufficiently large sample sizes".

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

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

Using Lemma A.2 (iii), we get:

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

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

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

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

Appendix 4: Proof of Propositions 3

This appendix concerns the derivation of the asymptotic MISE of the Q-TiR estimator. The steps are as follows: computing the Frechet derivatives in A.4.1, getting the first-order condition in A.4.2, making the asymptotic expansion of the MISE in A.4.3, deriving the final expression in A.4.4.

A.4.1 Frechet derivatives

Lemma A.3: *Under Assumption A.6, the Frechet derivative of \mathcal{A} at φ_0 is the linear operator $A := D\mathcal{A}(\varphi_0)$ defined by $A\psi(z) = \int f_{X,Y|Z}(x, \varphi_0(x)|z)\psi(x) dx$, $z \in \mathcal{Z}$, for $\psi \in L^2[0, 1]$. Moreover we have $\mathcal{A}(\varphi) = \mathcal{A}(\varphi_0) + A\Delta\varphi + R(\varphi, \varphi_0)$, where the residual $R(\varphi, \varphi_0)$ is such that $\|R(\varphi, \varphi_0)\|_{L^2(F_Z)} \leq \frac{1}{2}c\|\Delta\varphi\|^2$, and $c := \sup_{x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}} |\nabla_y f_{X,Y|Z}(x, y|z)|$.*

Define $\langle \varphi, \psi \rangle_{L^2(\hat{F}_{Z,\tau})} := \frac{1}{T\tau(1-\tau)} \sum_{t=1}^T I_t \varphi(Z_t) \psi(Z_t)$ and $\|\psi\|_{L^2(\hat{F}_{Z,\tau})}^2 := \langle \psi, \psi \rangle_{L^2(\hat{F}_{Z,\tau})}$. Then, $\langle \varphi, \hat{A}\psi \rangle_{L^2(\hat{F}_{Z,\tau})} = \langle \hat{A}^*\varphi, \psi \rangle_H$.

Lemma A.4: *Under Assumption A.10, the Frechet derivative of $\hat{\mathcal{A}}$ at $\bar{\varphi}$ is the linear operator $\bar{A} := D\hat{\mathcal{A}}(\bar{\varphi})$ defined by $\bar{A}\psi(z) = \int \hat{f}_{X,Y|Z}(x, \bar{\varphi}(x)|z)\psi(x) dx$, $z \in \mathcal{Z}$, for $\psi \in L^2[0, 1]$. 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., $\|\hat{R}(\varphi, \bar{\varphi})\|_{L^2(\hat{F}_{Z,\tau})} \leq \frac{1}{2\sqrt{\tau(1-\tau)}}\hat{c}\|\varphi - \bar{\varphi}\|^2$, and $\hat{c} := \sup_{x \in \mathcal{X}, y \in \mathbb{R}, z \in \mathcal{Z}_T} |\nabla_y \hat{f}_{X,Y|Z}(x, y|z)|$.*

We will denote the Frechet derivative of $\hat{\mathcal{A}}$ at φ_0 by $\hat{A}_0 := D\hat{\mathcal{A}}(\varphi_0)$, and denote $\hat{A} := D\hat{\mathcal{A}}(\hat{\varphi})$.

A.4.2 First-order condition

Proof of Equation (9): By Assumption A.9, let $r > 0$ be such that $B_r(\varphi_0) \cap H^2[0, 1]$ is contained in Θ . The estimator $\hat{\varphi}$ is such that, when $\|\Delta\hat{\varphi}\| < r$ we have:

$$\forall \psi \in H^2[0, 1], \exists \rho = \rho(\psi) > 0 : \hat{\varphi} + \varepsilon\psi \in \Theta \text{ for any } \varepsilon \text{ s.t. } |\varepsilon| < \rho.$$

Thus, when $\|\Delta\hat{\varphi}\| < r$ the estimator $\hat{\varphi}$ satisfies the first order condition

$$\left. \frac{d}{d\varepsilon} L_T(\hat{\varphi} + \varepsilon\psi) \right|_{\varepsilon=0} = 0, \quad \forall \psi \in H^2[0, 1].$$

Writing $L_T(\varphi) = \left\| \hat{\mathcal{A}}(\varphi) - \tau \right\|_{L^2(\hat{F}_Z, \tau)}^2 + \lambda_T \|\varphi\|_H^2$, we have

$$\begin{aligned} \left. \frac{d}{d\varepsilon} L_T(\hat{\varphi} + \varepsilon\psi) \right|_{\varepsilon=0} &= 2 \left\langle \hat{\mathcal{A}}(\hat{\varphi}) - \tau, D\hat{\mathcal{A}}(\hat{\varphi})\psi \right\rangle_{L^2(\hat{F}_Z, \tau)} + 2\lambda_T \langle \hat{\varphi}, \psi \rangle_H \\ &= 2 \left\langle \hat{\mathcal{A}}(\hat{\varphi}) - \tau, \hat{A}\psi \right\rangle_{L^2(\hat{F}_Z, \tau)} + 2\lambda_T \langle \hat{\varphi}, \psi \rangle_H \\ &= 2 \left\langle \hat{A}^* \left(\hat{\mathcal{A}}(\hat{\varphi}) - \tau \right) + \lambda_T \hat{\varphi}, \psi \right\rangle_H. \end{aligned}$$

By the consistency of $\hat{\varphi}$ (Proposition 2), $P[\|\Delta\hat{\varphi}\| < r] \rightarrow 1$. We show below that $P[\|\Delta\hat{\varphi}\| \geq r] = O(T^{-\bar{b}})$, for any $\bar{b} > 0$.

A.4.3 Asymptotic expansion

In this section we provide the asymptotic expansion of the MISE of estimator $\hat{\varphi}$. Our strategy consists of three steps. In Step (i) we show that the nonlinearity term $\hat{\mathcal{K}}_T(\Delta\hat{\varphi})$ in Equation (11) satisfies a quadratic bound w.p.a. 1 (Lemma A.5). In Step (ii) we exploit Equation (11) and the quadratic nature of $\hat{\mathcal{K}}_T(\Delta\hat{\varphi})$ to get a bound on the difference between $E[\|\Delta\hat{\varphi}\|^2]$ and $E\left[\left\|\Delta\hat{\psi}\right\|^2\right]$ (Lemmas A.6 and A.7). The bound involves probabilities of large deviations

for $\|\Delta\hat{\varphi}\|$ and $\|\Delta\hat{\psi}\|$. In Step (iii) we bound these probabilities by a large deviation result for penalized minimum distance estimators (Lemmas A.8 and A.9). Combining the three steps, we get the asymptotic expansion of the MISE $E[\|\Delta\hat{\varphi}\|^2]$ in terms of the expectations of powers of $\|\Delta\hat{\psi}\|$ (Lemma A.10).

(i) Quadratic bound for the nonlinearity term

Lemma A.5: *Under Assumptions A.1-A.3, A.6, A.7, A.10, for any $\bar{b} > 0$ and any constant $C > \frac{1}{2\sqrt{\tau(1-\tau)}} \sup_{x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}} |\nabla_y f_{X,Y|Z}(x, y|z)|$: $P\left[\|\hat{\mathcal{K}}_T(\Delta\hat{\varphi})\| > \frac{C}{\sqrt{\lambda_T}} \|\Delta\hat{\varphi}\|^2\right] = O\left(T^{-\bar{b}}\right)$.*

(ii) Control of the nonlinearity term

First we consider the nonstochastic analogue of Equation (11).

Lemma A.6: *Let function φ satisfy $\varphi = \psi + \varepsilon\mathcal{K}(\varphi)$, where ψ is a known function, \mathcal{K} a nonlinear operator such that $\|\mathcal{K}(\varphi)\| \leq \|\varphi\|^2$, and $\varepsilon > 0$. If $\varepsilon\|\psi\| < 1/8$, then either $|\|\varphi\|^2 - \|\psi\|^2| \leq 32\varepsilon\|\psi\|^3$, or $\|\varphi\|^2 \geq \frac{3}{8\varepsilon^2}$.*

We can use Lemma A.6 with $\varepsilon = \varepsilon_T = \frac{C}{\sqrt{\lambda_T}}$ to bound the difference $\|\Delta\hat{\varphi}\|^2 - \|\Delta\hat{\psi}\|^2$ on the set $\left\{\varepsilon_T \|\Delta\hat{\psi}\| < 1/8 \wedge \|\Delta\hat{\varphi}\|^2 < \frac{3}{8\varepsilon_T^2} \wedge \|\hat{\mathcal{K}}_T(\Delta\hat{\varphi})\| \leq \varepsilon_T \|\Delta\hat{\varphi}\|^2\right\}$, and derive the following result.

Lemma A.7: *Under Assumptions A.1-A.3, A6, A.7 and A.10, we have for any $\bar{b} > 0$, with*

$$\begin{aligned}
C &> \frac{1}{2\sqrt{\tau(1-\tau)}} \sup_{x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}} |\nabla_y f_{X,Y|Z}(x, y|z)|, \\
E[\|\Delta\hat{\varphi}\|^2] - E[\|\Delta\hat{\psi}\|^2] &= O\left(\frac{1}{\sqrt{\lambda_T}} E[\|\Delta\hat{\psi}\|^3] + P\left[\|\Delta\hat{\psi}\|^2 \geq \frac{\lambda_T}{64C^2}\right] \right. \\
&\quad \left. + P\left[\|\Delta\hat{\varphi}\|^2 \geq \frac{3\lambda_T}{8C^2}\right]\right) + O(T^{-\bar{b}}). \tag{40}
\end{aligned}$$

Note that for large T , probability $P\left[\|\Delta\hat{\varphi}\|^2 > \frac{3\lambda_T}{8C^2}\right]$ on the RHS of (40) controls for both the event $\|\Delta\hat{\varphi}\|^2 \geq \frac{3}{8\varepsilon_T^2}$ and the event $\|\Delta\hat{\varphi}\| \geq r$, in which the first-order condition (11) does not hold.

(iii) A large deviation bound for penalized minimum distance estimators

Lemma A.8: *We have $P[\|\hat{\varphi} - \varphi_0\| \geq \varepsilon_T] \leq k_1(T, C(\varepsilon_T, \lambda_T)) + k_2(T, C(\varepsilon_T, \lambda_T))$, where $\varepsilon_T > 0$,*

$$C(\varepsilon, \lambda) := \inf_{\varphi \in \Theta: \|\varphi - \varphi_0\| \geq \varepsilon} Q_\infty(\varphi) + \lambda \|\varphi\|_H^2 - \lambda \|\varphi_0\|_H^2, \tag{41}$$

and

$$k_1(T, \eta) := P\left[\sup_{\varphi \in \Theta} \sup_{z \in \mathcal{Z}_T} \frac{|\Delta\hat{m}(\varphi, z)|}{\sqrt{\tau(1-\tau)}} \geq \frac{\sqrt{\lambda_T \|\varphi_0\|_H^2 + 2\eta} - \sqrt{\lambda_T \|\varphi_0\|_H^2}}{4}\right], \tag{42}$$

$$k_2(T, \eta) := P\left[\sup_{\varphi \in \Theta} \left|\frac{1}{T\tau(1-\tau)} \sum_{t=1}^T m(\varphi, Z_t)^2 I_t - Q_\infty(\varphi)\right| \geq \eta/2\right]. \tag{43}$$

In an ill-posed setting, the usual "identifiable uniqueness" condition (White and Wooldrige (1991)) $\inf_{\varphi \in \Theta: \|\varphi - \varphi_0\| \geq \varepsilon} Q_\infty(\varphi) > Q_\infty(\varphi_0)$ does not hold (see GS). It is replaced by the Inequality $C(\varepsilon, \lambda) > 0$ for the penalized criterion, and the behaviour of $C(\varepsilon, \lambda)$ as $\lambda, \varepsilon \rightarrow 0$ matters for the rate of convergence of $\hat{\varphi}$. A lower bound for the function $C(\varepsilon, \lambda)$ as $\lambda \rightarrow 0$

and $\varepsilon = O(\sqrt{\lambda})$ is given in the next result.

Lemma A.9: *Suppose Assumption A.6 holds. Let $d > 0$ be a constant such that*

$$d^2 > \frac{\|\varphi_0\|_H^2}{\inf_{\psi: \|\psi\|_H \leq 2\|\varphi_0\|_H} \frac{\langle \psi, A^* A \psi \rangle_H}{\|\psi\|^2}}. \quad (44)$$

Then, for any $M < \infty$ and for λ close enough to 0 :

$$\inf_{\varphi \in \Theta: \|\varphi - \varphi_0\| \geq d\sqrt{\lambda}} Q_\infty(\varphi) + \lambda \|\varphi\|_H^2 - \lambda \|\varphi_0\|_H^2 \geq M \frac{\lambda}{\log(1/\lambda)}. \quad (45)$$

In the RHS of (45), $\frac{\lambda}{\log(1/\lambda)}$ can be replaced by $\lambda g(\lambda)$, where $g(\lambda)$ is any function of λ such that $g(\lambda) \rightarrow 0$ as $\lambda \rightarrow 0$ (see the proof of Lemma A.9). From Lemmas A.8 and A.9 we deduce that for $d > 0$ as in (44), and some constant $b > 0$:

$$\begin{aligned} P[\|\hat{\varphi} - \varphi_0\|^2 \geq d^2 \lambda_T] &\leq P\left[\sup_{\varphi \in \Theta} \sup_{z \in \mathcal{Z}_T} |\Delta \hat{m}(\varphi, z)|^2 \geq b \frac{\lambda_T}{\log(1/\lambda_T)}\right] \\ &+ P\left[\sup_{\varphi \in \Theta} \left| \frac{1}{T\tau(1-\tau)} \sum_{t=1}^T m(\varphi, Z_t)^2 I_t - Q_\infty(\varphi) \right| \geq b \frac{\lambda_T}{\log(1/\lambda_T)}\right]. \end{aligned} \quad (46)$$

From the proof of Lemma A.2 (Appendix 3.2) terms $\sup_{\varphi \in \Theta} \sup_{z \in \mathcal{Z}_T} |\Delta \hat{m}(\varphi, z)|^2$ and $\sup_{\varphi \in \Theta} \left| \frac{1}{T\tau(1-\tau)} \sum_{t=1}^T m(\varphi, Z_t)^2 I_t - Q_\infty(\varphi) \right|$ can be bounded by suprema of suitable empirical processes over compact finite-dimensional sets, which are $O_p\left((\log T)^2 \left(\frac{\log T}{Th_T^{1+d_Z}} + h_T^{2m}\right)\right)$ and $O_p\left(\frac{1}{\sqrt{T}}\right)$, respectively. When $\frac{1}{\lambda_T} \left(\frac{1}{Th_T^{1+d_Z}} + h_T^{2m} + \frac{1}{\sqrt{T}}\right) = O(T^{-\bar{\varepsilon}})$, for $\bar{\varepsilon} > 0$, these orders are negligible w.r.t. $\frac{\lambda_T}{\log(1/\lambda_T)}$, and by standard large deviation results the two probabilities in the RHS of (46) are converging to zero as $T \rightarrow \infty$ at a geometric rate. Thus,

they are negligible compared to the other terms in the RHS of (40), which converge as a negative power of T . The next result is proved by combining Inequality (46) and Lemma A.7, and yields the asymptotic expansion of the MISE $E [\|\Delta\hat{\varphi}\|^2]$.

Lemma A.10: *Suppose that Assumptions A.1-A.3, A.6, A.7, A.10 hold, and*

$$\frac{20}{\tau(1-\tau)} \sup_{x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}} |\nabla_y f_{X,Y|Z}(x, y|z)|^2 < \frac{1}{\|\varphi_0\|_H^2} \inf_{\psi: \|\psi\|_H \leq 2\|\varphi_0\|_H} \frac{\langle \psi, A^* A \psi \rangle_H}{\|\psi\|_H^2}. \quad (47)$$

Let $\lambda_T \rightarrow 0$ such that for $\bar{\varepsilon} > 0$:

$$\frac{1}{\lambda_T} \left(\frac{1}{T h_T^{1+d_Z}} + h_T^{2m} + \frac{1}{\sqrt{T}} \right) = O(T^{-\bar{\varepsilon}}). \quad (48)$$

Then $E [\|\Delta\hat{\varphi}\|^2] = E [\|\Delta\hat{\psi}\|^2] + O\left(\frac{1}{\sqrt{\lambda_T}} E [\|\Delta\hat{\psi}\|^3]\right) + O(T^{-\bar{b}})$, for any $\bar{b} > 0$.

Condition (47) is used to show that Condition (44) is satisfied, when bounding the probabilities in the RHS of (40) by means of the result in (46).

A.4.4 Proof of Proposition 3 (MISE)

From Conditions (14) and (15) we have $\frac{1}{\lambda_T^2} \frac{1}{T^2 h_T^{2+2d_Z}} = O\left(\frac{1}{\lambda_T^2} \frac{1}{T}\right) = o(\lambda_T^\varepsilon)$, $\frac{1}{\lambda_T} h_T^{2m} = O(\lambda_T^{1+\varepsilon})$, $\frac{1}{\lambda_T^2 T} = O\left(h_T^{\max\{d_Z, 2\}} \lambda_T^\varepsilon\right)$, and thus Condition (48) is satisfied as long as $h_T \asymp T^{-\eta}$ and $\lambda_T \asymp T^{-\gamma}$, $\eta, \gamma > 0$. From Assumption A.8 and Lemma A.10, the conclusion follows if:

- (i) $E \left[\|\Delta\hat{\psi}\|^2 \right] = \frac{1}{T} \sum_{j=1}^{\infty} \frac{\nu_j}{(\nu_j + \lambda_T)^2} \|\phi_j\|^2 + b(\lambda_T)^2$, up to negligible terms, and,
- (ii) $\frac{1}{\sqrt{\lambda_T}} E \left[\|\Delta\hat{\psi}\|^3 \right] = o\left(\frac{1}{T} \sum_{j=1}^{\infty} \frac{\nu_j}{(\nu_j + \lambda_T)^2} \|\phi_j\|^2 + b(\lambda_T)^2\right)$.

To show these statements, we write

$$\begin{aligned}
\hat{\mathcal{A}}(\varphi_0)(z) - \tau &= \int \int (1\{y \leq \varphi_0(x)\} - \tau) \frac{\Delta \hat{f}_{X,Y,Z}(x, y, z)}{f_Z(z)} dy dx \\
&+ \int \int (1\{y \leq \varphi_0(x)\} - \tau) \left[\hat{f}_{X,Y|Z}(x, y|z) - \frac{\hat{f}_{X,Y,Z}(x, y, z)}{f_Z(z)} \right] dy dx \\
&=: -\hat{\zeta}(z) + \hat{q}(z).
\end{aligned}$$

Then we decompose $\Delta \hat{\psi}$ as $\Delta \hat{\psi} =: \mathcal{V}_T + \mathcal{B}_T + \mathcal{R}_T$ such that

$$\Delta \hat{\psi} = (\lambda_T + A^*A)^{-1} A^* \hat{\zeta} + [(\lambda_T + A^*A)^{-1} A^*A - 1] \varphi_0 + \mathcal{R}_T,$$

where \mathcal{V}_T is the variance term, \mathcal{B}_T is the regularization bias term, and the remainder term

\mathcal{R}_T is given by

$$\begin{aligned}
\mathcal{R}_T &= \left[(\lambda_T + \hat{A}_0^* \hat{A}_0)^{-1} - (\lambda_T + A^*A)^{-1} \right] A^* \hat{\zeta} \\
&+ \left[(\lambda_T + \hat{A}_0^* \hat{A}_0)^{-1} \hat{A}_0^* \hat{A}_0 - (\lambda_T + A^*A)^{-1} A^*A \right] \varphi_0 \\
&+ (\lambda_T + \hat{A}_0^* \hat{A}_0)^{-1} \left(\hat{A}_0^* (\hat{\zeta} - \hat{q}) - A^* \hat{\zeta} \right) \\
&- (\lambda_T + \hat{A}_0^* \hat{A}_0)^{-1} \left(\hat{A}^* - \hat{A}_0^* \right) \left(\hat{\mathcal{A}}(\hat{\varphi}) - \tau \right). \tag{49}
\end{aligned}$$

We now give a series of inequalities and bounds to show that the remainder term \mathcal{R}_T can be neglected. First, from Cauchy-Schwarz inequality,

$$E \left[\left\| \Delta \hat{\psi} \right\|^2 \right] = E \left[\left\| \mathcal{V}_T + \mathcal{B}_T \right\|^2 \right] + E \left[\left\| \mathcal{R}_T \right\|^2 \right] + O \left(E \left[\left\| \mathcal{V}_T + \mathcal{B}_T \right\|^2 \right]^{1/2} E \left[\left\| \mathcal{R}_T \right\|^2 \right]^{1/2} \right), \tag{50}$$

and

$$\begin{aligned}
E \left[\left\| \Delta \hat{\psi} \right\|^3 \right] &\leq E \left[\left\| \Delta \hat{\psi} \right\|^4 \right]^{1/2} E \left[\left\| \Delta \hat{\psi} \right\|^2 \right]^{1/2} \\
&\leq C \left(E \left[\left\| \mathcal{V}_T + \mathcal{B}_T \right\|^4 \right] + E \left[\left\| \mathcal{R}_T \right\|^4 \right] \right)^{1/2} E \left[\left\| \Delta \hat{\psi} \right\|^2 \right]^{1/2}, \tag{51}
\end{aligned}$$

for a constant C . Second, we can isolate the estimation bias by writing

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

Thus,

$$\begin{aligned} E \left[\|\mathcal{V}_T + \mathcal{B}_T\|^2 \right] &= E \left[\left\| (\lambda_T + A^*A)^{-1} A^* \left(\hat{\zeta} - E\hat{\zeta} \right) \right\|^2 \right] \\ &+ \left\| (\lambda_T + A^*A)^{-1} A^* E\hat{\zeta} + [(\lambda_T + A^*A)^{-1} A^*A - 1] \varphi_0 \right\|^2, \end{aligned} \quad (52)$$

and for a constant C :

$$\begin{aligned} \|\mathcal{V}_T + \mathcal{B}_T\|^4 &\leq C \left(\left\| (\lambda_T + A^*A)^{-1} A^* \left(\hat{\zeta} - E\hat{\zeta} \right) \right\|^4 \right. \\ &\quad \left. + \left\| (\lambda_T + A^*A)^{-1} A^* E\hat{\zeta} + [(\lambda_T + A^*A)^{-1} A^*A - 1] \varphi_0 \right\|^4 \right). \end{aligned} \quad (53)$$

In Lemma A.11 we give the asymptotic behavior of $E \left[\left\| (\lambda_T + A^*A)^{-1} A^* \left(\hat{\zeta} - E\hat{\zeta} \right) \right\|^2 \right]$ and $E \left[\left\| (\lambda_T + A^*A)^{-1} A^* \left(\hat{\zeta} - E\hat{\zeta} \right) \right\|^4 \right]$. In Lemma A.12 we prove that estimation bias is negligible compared to regularization bias, and in Lemma A.13 we give bounds on the remainder term. Combining Lemmas A.11 (i), A.12, A.13 (i) with Equations (50), (52), yields Statement (i) above. Then, combining Lemmas A.11 (ii), A.12, A.13 (ii) with Inequalities (51), (53) yields $E \left[\left\| \Delta\hat{\psi} \right\|^3 \right] = O \left(\left(\frac{1}{T} \sum_{j=1}^{\infty} \frac{\nu_j}{(\nu_j + \lambda_T)^2} \|\phi_j\|^2 + b(\lambda_T)^2 \right)^{3/2} \right)$. The latter in turns implies Statement (ii) by using Condition (16).

Lemma A.11: *Under Assumptions A.1, A.2, A.4, A.10, A.11 (ii), A.12, A.13 (i):*

(i) *up to negligible terms,* $E \left[\left\| (\lambda_T + A^*A)^{-1} A^* \left(\hat{\zeta} - E\hat{\zeta} \right) \right\|^2 \right] = \frac{1}{T} \sum_{j=1}^{\infty} \frac{\nu_j}{(\nu_j + \lambda_T)^2} \|\phi_j\|^2,$

(ii) $E \left[\left\| (\lambda_T + A^*A)^{-1} A^* \left(\hat{\zeta} - E\hat{\zeta} \right) \right\|^4 \right] = O \left(\left(\frac{1}{T} \sum_{j=1}^{\infty} \frac{\nu_j}{(\nu_j + \lambda_T)^2} \|\phi_j\|^2 \right)^2 \right).$

Lemma A.12: *Suppose that Assumptions A.1, A.2, A.3 (ii), A.5, A.10 hold, and $h_T^m = o(\lambda_T b(\lambda_T))$. Then, up to negligible terms:*

$$\left\| (\lambda_T + A^*A)^{-1} A^* E \hat{\zeta} + [(\lambda_T + A^*A)^{-1} A^*A - 1] \varphi_0 \right\| = b(\lambda_T).$$

Lemma A.13: *Suppose that Assumptions A.1-A.7, A.10, A.11 (ii), A.12, A.13 hold, and $\frac{(\log T)^2}{T h_T^{2(1+d_Z)}} = O(1)$, $\frac{(\log T)^2}{T h_T^{d_Z+1}} + h_T^m = o(\lambda_T b(\lambda_T))$, and $\frac{1}{T h_T^{\max\{d_Z, 2\}}} + h_T^{2m} = O(\lambda_T^{2+\varepsilon})$, $\varepsilon > 0$.*

Then: (i) $E [\|\mathcal{R}_T\|^2] = o\left(\frac{1}{T} \sum_{j=1}^{\infty} \frac{\nu_j}{(\nu_j + \lambda_T)^2} \|\phi_j\|^2 + b(\lambda_T)^2\right)$, and

(ii) $E [\|\mathcal{R}_T\|^4] = o\left(\left(\frac{1}{T} \sum_{j=1}^{\infty} \frac{\nu_j}{(\nu_j + \lambda_T)^2} \|\phi_j\|^2 + b(\lambda_T)^2\right)^2\right)$.

Appendix 5: Proof of Proposition 5

Let us show the asymptotic normality of the Q-TiR estimator. From Equation (11) and the decomposition in A.4.4, we have

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

where $\mathcal{R}_T(x)$ is defined in (49). We now show that the term (I) is asymptotically $N(0, 1)$ distributed in A.5.1, and the terms (III), (IV), and (V) are $o_p(1)$ in A.5.2, which implies Proposition 5.

A.5.1 Asymptotic normality of (I)

Since $\{\phi_j : j \in \mathbb{N}\}$ is an orthonormal basis w.r.t. $\langle \cdot, \cdot \rangle_H$, we can write:

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

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

$$\sqrt{T/\sigma_T^2(x)} (\lambda_T + A^*A)^{-1} A^* \left(\hat{\zeta} - E\hat{\zeta} \right) (x) = \sum_{j=1}^{\infty} w_{j,T}(x) Z_{j,T}, \quad (54)$$

where $Z_{j,T} := \frac{1}{\sqrt{\nu_j}} \langle \phi_j, \sqrt{T} A^* \left(\hat{\zeta} - E\hat{\zeta} \right) \rangle_H$, $j = 1, 2, \dots$,

and $w_{j,T}(x) := \frac{\sqrt{\nu_j}}{\lambda_T + \nu_j} \phi_j(x) / \left(\sum_{j=1}^{\infty} \frac{\nu_j}{(\lambda_T + \nu_j)^2} \phi_j(x)^2 \right)^{1/2}$, $j = 1, 2, \dots$.

Note that $\sum_{j=1}^{\infty} w_{j,T}(x)^2 = 1$. Equation (54) can be rewritten (see the proof of Lemma A.11)

using

$$\sum_{j=1}^{\infty} w_{j,T}(x) Z_{j,T} = -\sqrt{T} \int G_T(r) \left[\hat{f}_{X,Y,Z}(r) - E\hat{f}_{X,Y,Z}(r) \right] dr, \quad (55)$$

$r = (w, z)$, $G_T(r) := G_T(x, r) := \sum_{j=1}^{\infty} w_{j,T}(x) g_j(r)$, $g_j(r) = \frac{1}{\tau(1-\tau)} \frac{(A\phi_j)(z)}{\sqrt{\nu_j}} 1_{\varphi_0}(w)$, and

$1_{\varphi_0}(w) = 1\{y \leq \varphi_0(x)\} - \tau$.

Lemma A.14: *Suppose that Assumptions A.1, A.10, A.11 (i), A.13 (ii) hold, $h_T^m = o(\lambda_T)$*

and $\sqrt{h_T} \frac{\sum_{j=1}^{\infty} \frac{\nu_j}{(\lambda_T + \nu_j)^2}}{\sum_{j=1}^{\infty} \frac{\nu_j}{(\lambda_T + \nu_j)^2} \phi_j^2(x)} = o(1)$. Then: $\sqrt{T} \int G_T(r) \left[\hat{f}_{X,Y,Z}(r) - E\hat{f}_{X,Y,Z}(r) \right] dr$

$$= \frac{1}{\sqrt{T}} \sum_{t=1}^T Y_{tT} + o_p(1), \text{ where } Y_{tT} := G_T(R_t) = \sum_{j=1}^{\infty} w_{j,T}(x) g_j(R_t), R_t = (X_t, Y_t, Z_t)'$$

From Lemma A.14 it is sufficient to prove that $T^{-1/2} \sum_{t=1}^T Y_{tT}$ is asymptotically $N(0, 1)$ distributed. Note that $E[g_j(R)] = \frac{1}{\tau(1-\tau)} \frac{1}{\sqrt{\nu_j}} E[(A\phi_j)(Z) E[1_{\varphi_0}(W) | Z]] = 0$, and

$$\begin{aligned} \text{Cov}[g_j(R), g_l(R)] &= \frac{1}{\tau^2(1-\tau)^2} \frac{1}{\sqrt{\nu_j}\sqrt{\nu_l}} E[(A\phi_j)(Z) E[1_{\varphi_0}(W)^2 | Z] (A\phi_l)(Z)] \\ &= \frac{1}{\tau(1-\tau)} \frac{1}{\sqrt{\nu_j}\sqrt{\nu_l}} E[(A\phi_j)(Z) (A\phi_l)(Z)] \\ &= \frac{1}{\sqrt{\nu_j}\sqrt{\nu_l}} \langle \phi_j, A^* A \phi_l \rangle_H = \delta_{j,l}. \end{aligned}$$

Thus $E[Y_{tT}] = 0$ and $V[Y_{tT}] = \sum_{j,l=1}^{\infty} w_{j,T}(x) w_{l,T}(x) \text{Cov}[g_j(R), g_l(R)] = \sum_{j=1}^{\infty} w_{j,T}(x)^2 = 1$.

From application of a Lyapunov CLT, it is sufficient to show that

$$\frac{1}{T^{1/2}} E[|Y_{tT}|^3] \rightarrow 0, \quad T \rightarrow \infty. \quad (56)$$

To this goal, using $|Y_{tT}| \leq \sum_{j=1}^{\infty} |w_{j,T}(x)| |g_j(R_t)|$ and the triangular inequality, we get

$$\begin{aligned} \frac{1}{T^{1/2}} E[|Y_{tT}|^3] &\leq \frac{1}{T^{1/2}} E \left[\left(\sum_{j=1}^{\infty} |w_{j,T}(x)| |g_j(R)| \right)^3 \right] = \frac{1}{T^{1/2}} \left\| \sum_{j=1}^{\infty} |w_{j,T}(x)| |g_j| \right\|_3^3 \\ &\leq \frac{1}{T^{1/2}} \left(\sum_{j=1}^{\infty} |w_{j,T}(x)| \|g_j\|_3 \right)^3 = \frac{1}{T^{1/2}} \frac{\left(\sum_{j=1}^{\infty} \frac{\sqrt{\nu_j}}{\lambda_T + \nu_j} |\phi_j(x)| \|g_j\|_3 \right)^3}{\left(\sum_{j=1}^{\infty} \frac{\nu_j}{(\lambda_T + \nu_j)^2} \phi_j(x)^2 \right)^{3/2}}. \end{aligned}$$

Moreover, from the Cauchy-Schwarz inequality we have

$$\sum_{j=1}^{\infty} \frac{\sqrt{\nu_j}}{\lambda_T + \nu_j} |\phi_j(x)| \|g_j\|_3 \leq \left(\sum_{j=1}^{\infty} \frac{\nu_j}{(\lambda_T + \nu_j)^2} \phi_j(x)^2 \|g_j\|_3^2 j^{1+\varepsilon} \right)^{1/2} \left(\sum_{j=1}^{\infty} \frac{1}{j^{1+\varepsilon}} \right)^{1/2},$$

and $\sum_{j=1}^{\infty} \frac{1}{j^{1+\bar{\varepsilon}}} < \infty$, for any $\bar{\varepsilon} > 0$. Thus, we get

$$\frac{1}{T^{1/2}} E [|Y_{tT}|^3] \leq \left(\sum_{j=1}^{\infty} \frac{1}{j^{1+\bar{\varepsilon}}} \right)^{3/2} \left(\frac{1}{T^{1/3}} \frac{\sum_{j=1}^{\infty} \frac{\nu_j}{(\lambda_T + \nu_j)^2} \phi_j(x)^2 \|g_j\|_3^2 j^{1+\bar{\varepsilon}}}{\sum_{j=1}^{\infty} \frac{\nu_j}{(\lambda_T + \nu_j)^2} \phi_j(x)^2} \right)^{3/2},$$

and Condition (56) is implied by Condition (30).

A.5.2 Terms (III), (IV), and (V) are $o(1)$, $o_p(1)$, $o_p(1)$

Lemma A.15: *Suppose that Assumptions A.5, A.10 hold, and $h_T^m = O(\lambda_T^{1+\varepsilon/2} b(\lambda_T))$, $\frac{M_T(\lambda_T)}{\sigma_T^2(x)/T} = o(\lambda_T^{-\varepsilon})$, for $\varepsilon > 0$. Then: $\sqrt{T/\sigma_T^2(x)} (\lambda_T + A^*A)^{-1} A^* E \hat{\zeta}(x) = o(1)$.*

Lemma A.16: *Suppose that Assumptions A hold, and $\frac{(\log T)^2}{Th_T^{d_Z+1}} + h_T^m = O(\lambda_T^{1+\varepsilon/2} b(\lambda_T))$, $\frac{1}{Th_T^{d_Z}} + h_T^{2m} = O(\lambda_T^{2+\varepsilon})$, $\frac{(\log T)^2}{T^{2(1+d_Z)}} = O(1)$, $\frac{1}{Th_T^2} + h_T^{2m} = O(\lambda_T^{2+\varepsilon})$, $\frac{M_T(\lambda_T)}{\sigma_T^2(x)/T} = o(\lambda_T^{-\varepsilon})$, for $\varepsilon > 0$. Then: $\sqrt{T/\sigma_T^2(x)} \mathcal{R}_T(x) = o_p(1)$.*

Lemma A.17: *Suppose that Assumptions A hold, and $\frac{1}{Th_T^{\max\{d_Z, 2\}}} + h_T^{2m} = O(\lambda_T^{2+\varepsilon})$, $\frac{(\log T)^2}{Th_T^{d_Z+1}} + h_T^m = O(\lambda_T b(\lambda_T))$, $\frac{(\log T)^2}{T^{2(1+d_Z)}} = O(1)$, $\frac{M_T(\lambda_T)}{\sigma_T^2(x)/T} = o(\lambda_T^{-\varepsilon})$, $T\lambda_T^3 = O(1)$, $M_T(\lambda_T) = O(\lambda_T^{1+\varepsilon})$ for $\varepsilon > 0$. Then: $\sqrt{T/\sigma_T^2(x)} \hat{\mathcal{K}}_T(\Delta \hat{\varphi})(x) = o_p(1)$.*

TABLE 1: Case 1: Separable model: $\varphi_0(x) = \sin(\pi x) + \Phi^{-1}(\tau)$.

Asymptotic and finite sample optimal regularization parameter and MISE.

Average ISE, average and quartiles of selected regularization parameters

with data driven procedure.

τ	.10	.25	.50	.75	.90
Asymptotic values					
λ	.0006	.0008	.0009	.0008	.0006
MISE	.0302	.0180	.0147	.0180	.0302
Finite sample values					
λ	.0004	.0005	.0006	.0005	.0003
MISE	.0353	.0189	.0133	.0173	.0401
Data driven values					
Ave λ	.0009	.0006	.0006	.0006	.0009
1st Qu λ	.0003	.0003	.0003	.0003	.0003
Med λ	.0004	.0004	.0004	.0004	.0004
2nd Qu λ	.0007	.0005	.0006	.0006	.0012
Ave ISE	.0423	.0225	.0178	.0237	.0466

TABLE 2: Case 2: Nonseparable model: $\varphi_0(x) = \Phi(3(x + \tau - 1))$.

Asymptotic and finite sample optimal regularization parameter and MISE.

Average ISE, average and quartiles of selected regularization parameters

with data driven procedure.

τ	.10	.25	.50	.75	.90
Asymptotic values					
λ	.013	.013	.027	.016	.016
MISE	.0010	.0013	.0010	.0012	.0009
Finite sample values					
λ	.001	.004	.012	.003	.002
MISE	.0024	.0019	.0023	.0021	.0024
Data driven values					
Ave λ	.007	.007	.011	.009	.008
1st Qu λ	.003	.002	.002	.002	.002
Med λ	.006	.005	.006	.005	.005
2nd Qu λ	.008	.010	.015	.013	.008
Ave ISE	.0043	.0029	.0038	.0033	.0036

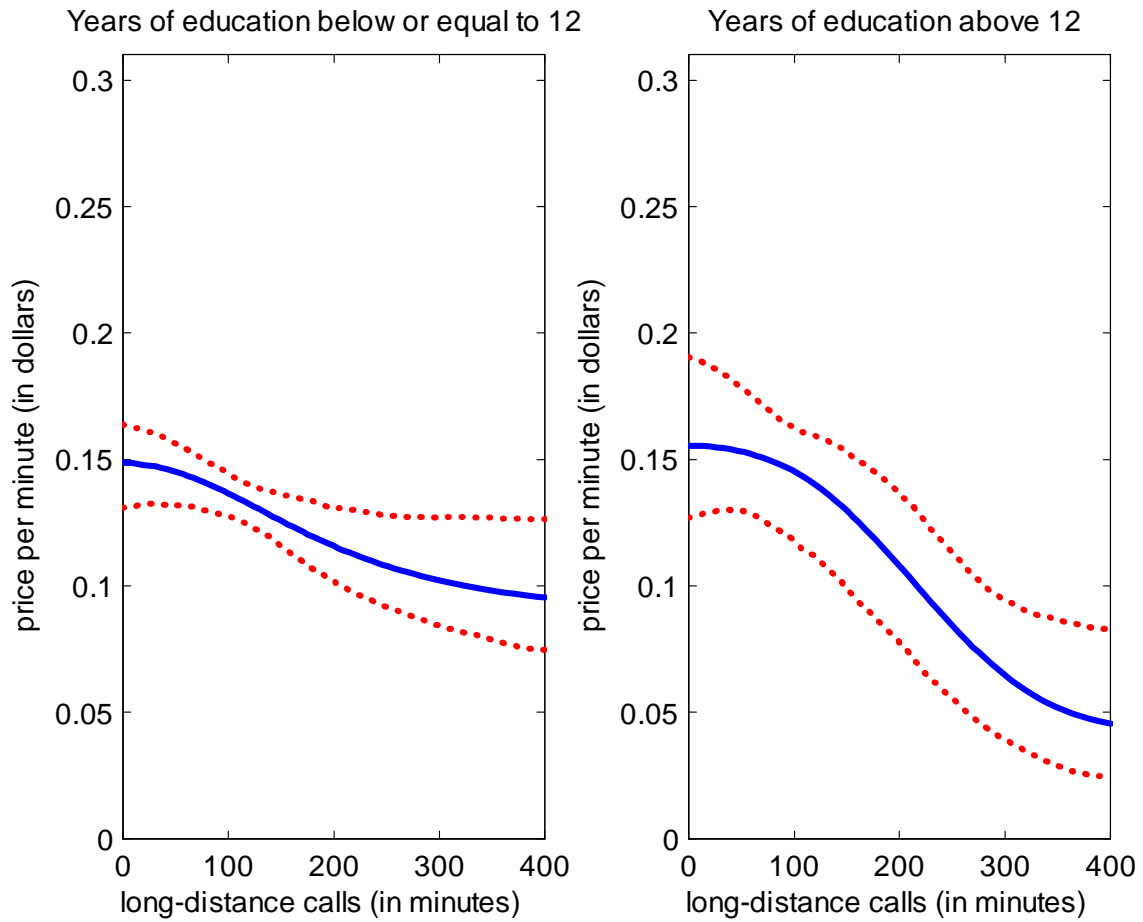


Figure 1: Estimated median structural effect (solid line) and bootstrap pointwise confidence bands at 95% (dotted lines) for the two education categories.

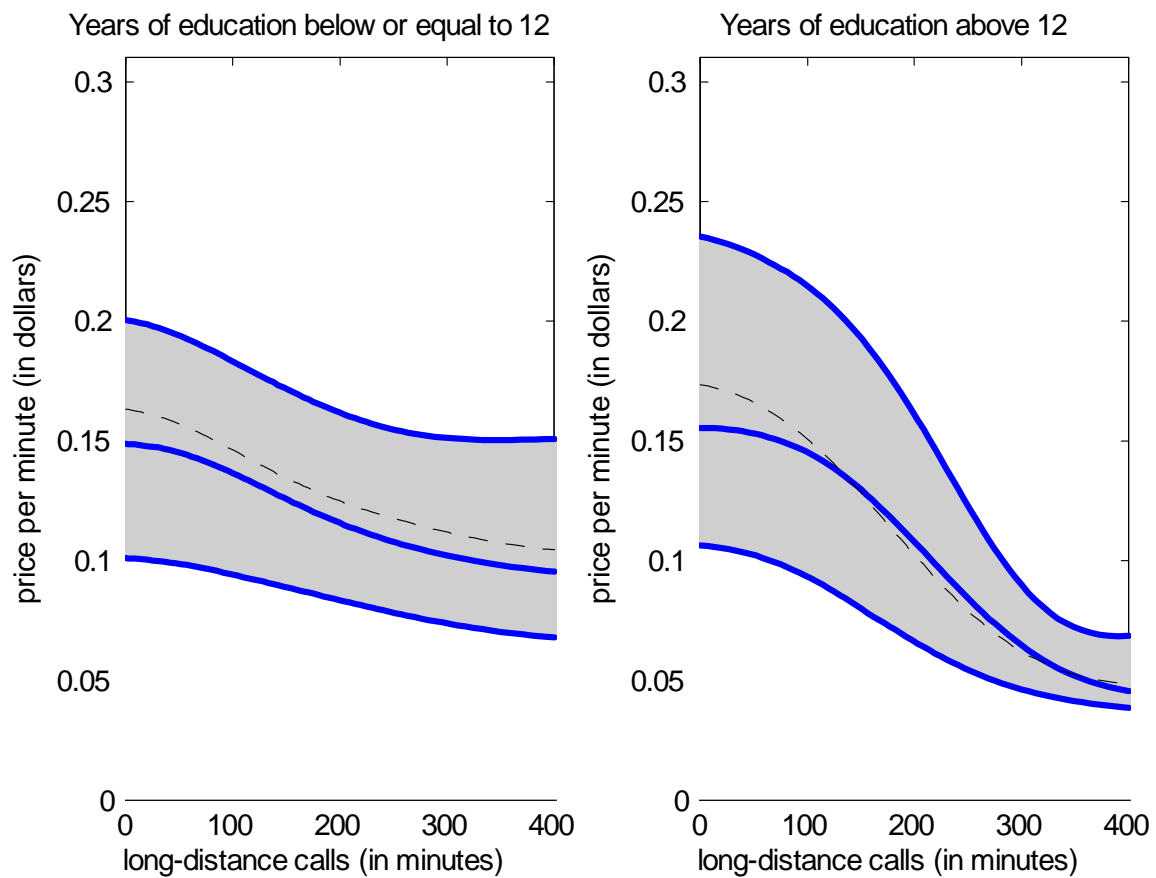


Figure 2: Estimated quartile structural effects (solid lines) and mean structural effect (dashed line) for the two education categories.

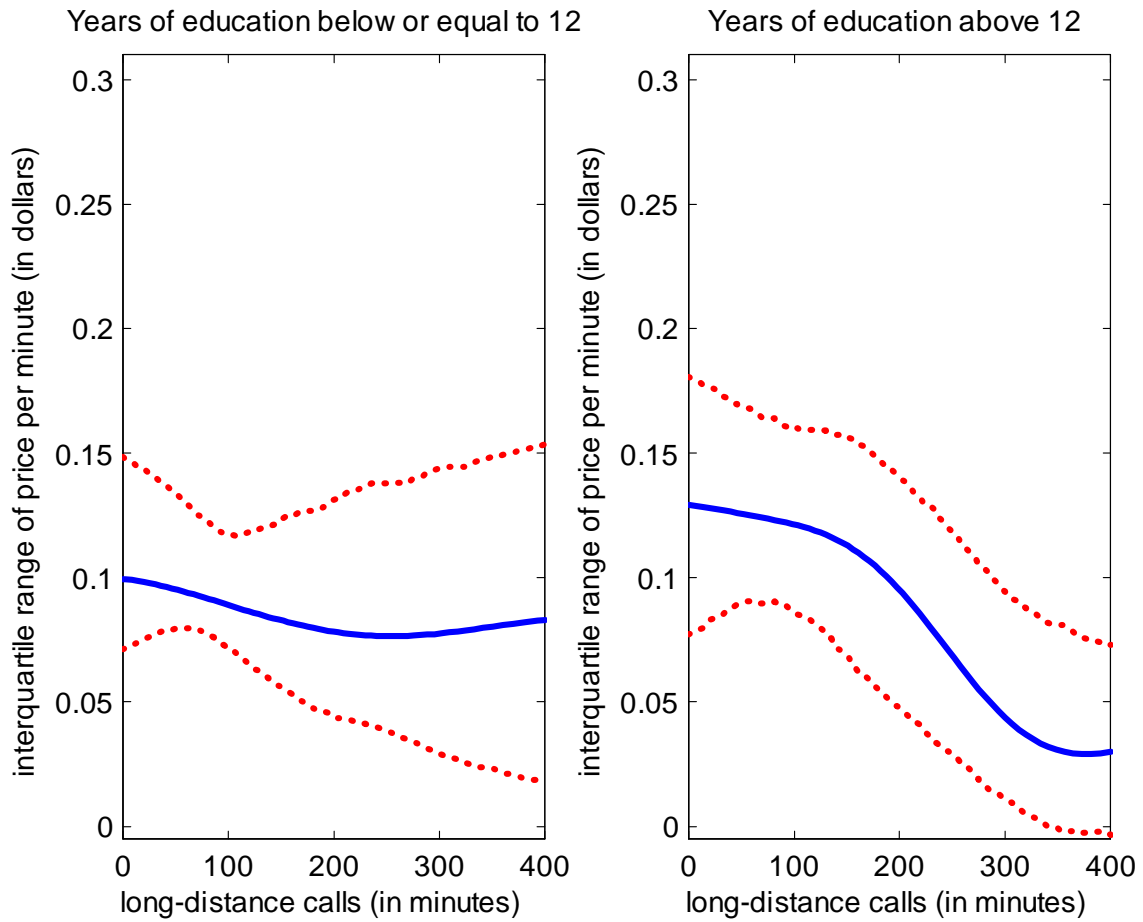


Figure 3: Estimated interquartile range of structural effect (solid line) and bootstrap pointwise confidence bands at 95% (dotted lines) for the two education categories.