

National Centre of Competence in Research
Financial Valuation and Risk Management

Working Paper No. 379

A SPECIFICATION TEST FOR NONPARAMETRIC INSTRUMENTAL VARIABLE REGRESSION

P. Gagliardini O. Scaillet

First version: April 2007
Current version: April 2007

This research has been carried out within the NCCR FINRISK projects on
“New Methods in Theoretical and Empirical Asset Pricing” and “Financial
Econometrics for Risk Management”

**A SPECIFICATION TEST FOR NONPARAMETRIC
INSTRUMENTAL VARIABLE REGRESSION**

P. Gagliardini* and O. Scaillet†

First version: April 2007 ‡

*University of Lugano and Swiss Finance Institute.

†HEC Genève and Swiss Finance Institute.

‡Both authors received support by the Swiss National Science Foundation through the National Center of Competence in Research: Financial Valuation and Risk Management (NCCR FINRISK).

A Specification Test for Nonparametric Instrumental Variable Regression

Abstract

We consider testing for correct specification of a nonparametric instrumental variable regression. In this ill-posed inverse problem setting, the test statistic is based on the empirical minimum distance criterion corresponding to the conditional moment restriction evaluated with a Tikhonov Regularized estimator of the functional parameter. Its asymptotic distribution is normal under the null hypothesis, and a consistent bootstrap is available to get simulation based critical values. We explore the finite sample behavior with Monte Carlo experiments. Finally, we provide an empirical application for an estimated Engel curve.

Keywords and phrases: Specification Test, Nonparametric Regression, Instrumental Variables, Minimum Distance, Tikhonov Regularization, Ill-posed Inverse Problems, Generalized Method of Moments, Bootstrap, Engel Curve.

JEL classification: C13, C14, C15, D12.

AMS 2000 classification: 62G08, 62G20.

1 Introduction

Testing for correct specification of a relationship that is written as a moment restriction has a long history in econometrics. At the end of the 50's Sargan suggests a specification test for an instrumental variable (IV) linear model (Sargan (1958)), and its generalization for a nonlinear-in-parameters IV model (Sargan (1959)). Hansen (1982) extends this type of specification test to the general nonlinear framework known as the Generalized Method of Moments (GMM). These tests are known as Hansen-Sargan tests or “ J -tests”, and are part of standard software reports on IV and GMM estimation.

In this paper we consider testing for correct specification of a nonparametric instrumental variable regression defined by the conditional moment restriction

$$E_0 [Y - \varphi_0(X) | Z] = 0, \tag{1}$$

where $E_0 [\cdot | Z]$ denotes expectation with respect to the true conditional distribution F_0 of $W = (Y, X)$ given Z , and the parameter of interest φ_0 is a function defined on $[0, 1]$. There has recently been much interest in nonparametric estimation of φ_0 in (1) (see, e.g., Ai and Chen (2003), Darolles, Florens, and Renault (2003), Newey and Powell (2003), Hall and Horowitz (2005)), and testing a parametric specification in (1) (see, e.g., Donald, Imbens, and Newey (2003), Tripathi and Kitamura (2003, TK), Horowitz (2006)). Up to now there is no attempt to directly test whether (1) holds or not on the data in a functional setting. Since Equation (1) is a linear integral equation of the first kind in φ_0 , we face an ill-posed inverse problem. In a different ill-posed setting, namely parametric GMM estimation with

a continuum of moment conditions, Carrasco and Florens (2000) also study specification testing, and show the asymptotic normality of their J -test statistic.

Section 2 describes the null hypothesis of correct specification and the concept of overidentification in a nonparametric IV setting. We clarify these notions with two simple examples. Section 3 gives the test statistic and its asymptotic properties. We further explain how to implement a consistent bootstrap to get simulation based critical values. Section 4 explores the finite sample behavior with Monte Carlo experiments. Section 5 provides an empirical application for an estimated Engel curve. In Appendices 1 and 2 we gather the list of regularity conditions and the technical arguments justifying the asymptotic distribution under the null hypothesis.

2 The null hypothesis and overidentification

The conditional moment restriction (1) is a linear integral equation

$$A_{F_0}\varphi_0 = r_{F_0}, \tag{2}$$

with $A_F\varphi(z) = \int f(w|z)\varphi(x)dw$ and $r_F(z) = \int yf(w|z)dy$ for $F \in \mathcal{F}$, where \mathcal{F} denotes the set of all conditional distributions of W given Z such that $r_F \in L^2(\mathcal{Z};\mu)$ and A_F is a compact linear operator from $L^2[0,1]$ into $L^2(\mathcal{Z};\mu)$. The definition of the appropriate measure μ on the support \mathcal{Z} of Z depends on the estimation method. The model $\mathcal{M} \subset \mathcal{F}$ is the subset of constrained distributions such that equation $A_F\varphi = r_F$ admits a unique

solution φ , that is

$$\mathcal{M} = \{F \in \mathcal{F} : r_F \in \text{Range}(A_F) \text{ and } \ker(A_F) = \{0\}\},$$

(Darolles, Florens, Renault (2003)). The condition $r_F \in \text{Range}(A_F)$ ensures existence of a solution. The condition $\ker(A_F) = \{0\}$ on the null space of A_F ensures uniqueness of the solution, since it is equivalent to operator A_F being injective (Carrasco, Florens, Renault (2006)).¹

The null hypothesis of correct specification is

$$H_0 : F_0 \in \mathcal{M}, \tag{3}$$

while the alternative hypothesis is $H_1 : F_0 \in \mathcal{F} \setminus \mathcal{M}$.

It is well-known that in the standard parametric GMM setting, the test of correct specification is meaningful only in an overidentified case, that is, when the number of marginal moment restrictions is larger than the number of parameters. In our functional setting with conditional moment restrictions, the definition of overidentification is less straightforward, since the number of moment restrictions is infinite and the parameter is infinite dimensional. The model is overidentified if $\mathcal{M} \subsetneq \mathcal{F}$. Otherwise, if $\mathcal{M} = \mathcal{F}$ a unique solution φ_F of $A_F \varphi_F = r_F$ always exists for any $F \in \mathcal{F}$. In this case, the conditional moment restriction (1) has no informational content for the true conditional distribution of W given Z since (3) does not imply a constraint on F_0 . This is the analogue of the just identified case in the standard parametric GMM setting.

¹ Primitive assumptions on the distribution F that ensure the identification condition $\ker(A_F) = \{0\}$ are derived, e.g., in Newey and Powell (2003).

It turns out that the nonparametric instrumental variable regression model is overidentified by construction: \mathcal{M} cannot be equal to \mathcal{F} . First, note that: (i) operator A_F is compact, and thus $\text{Range}(A_F) \subsetneq L^2(\mathcal{Z}; \mu)$, for any $F \in \mathcal{F}$, and (ii) $\text{Range}(A_F)$ depends only on the conditional distribution $F_{X|Z}$ of X given Z , while r_F depends only on the conditional distribution $F_{Y|Z}$ of Y given Z . Thus, given the margin $F_{X|Z}$, there exist admissible $F \in \mathcal{F}$ such that $r_F \notin \text{Range}(A_F)$. Second, there exist admissible $F \in \mathcal{F}$ such that $\ker(A_F) \neq \{0\}$.

In light of (3) and the former discussion, the notion of misspecification in a nonparametric IV regression setting is intimately linked with the properties of $\text{Range}(A_F)$. To clarify this we develop two simple examples. First, consider $Y = \varphi_*(X) + U$, where $E[U|Z = z] = \rho(z)$, for a given $\varphi_* \in L^2[0, 1]$ and $\rho \in L^2(\mathcal{Z}; \mu)$. Further, assume that $\ker(A_F) = \{0\}$. Then, r_F is such that $r_F = A_F \varphi_* + \rho$, and $F \in \mathcal{M}$ if and only if

$$\rho \in \text{Range}(A_F). \quad (4)$$

Example 1: Let (X_*, U, Z) be jointly normal with zero means, unit variances, $\text{Cov}(X_*, Z) = \rho_{X_*Z} \neq 0$, and $\text{Cov}(U, Z) = \rho_{UZ}$. Let further $X = H(X_*)$ be a monotone transformation of X_* such that $H^{-1} \in L^2[0, 1]$. Then, $\rho(z) = \rho_{UZ}z$ is linear, and we have $\rho = A_F \Delta\varphi$ where $\Delta\varphi(x) = \frac{\rho_{UZ}}{\rho_{X_*Z}} H^{-1}(x)$. Thus, $\rho \in \text{Range}(A_F)$.

In Example 1, even if the innovation and the instrument are correlated, this does not prevent $F \in \mathcal{M}$ to be satisfied, and this for any ρ_{UZ} . This exemplifies a key difference between restrictions induced by a parametric conditional moment setting and their nonparametric counterpart. In the finite-dimensional setting with $E_0[Y - \varphi(X, \theta_0)|Z] = 0$, we get a correct

specification in Example 1 if $\varphi(x, \theta_0) = \varphi_*(x) + \frac{\rho_{UZ}}{\rho_{X_*Z}} H^{-1}(x)$. If this restriction holds for a given ρ_{UZ} , it will typically not for another correlation value. The condition for correct specification also differs in the standard linear IV setting:

$$Y = \theta'_* X + U, \quad X = \Gamma' Z + V,$$

where V is uncorrelated with Z . There the orthogonality condition $E_0[(Y - \theta'_* X)Z] = 0$ is correctly specified if $E_0[ZZ']^{-1}E_0[ZU] \in \text{Range}(\Gamma_0)$.

Condition (4) generally imposes a restriction on the smoothness of function ρ . This statement is made precise by Picard Theorem (e.g., Kress (1999), Theorem 15.18), which fully characterizes the range of a compact operator in terms of its spectral decomposition. Here we limit ourselves to a second example. Assume that operator A_F is such that the functions in its range are continuous, i.e., $\text{Range}(A_F) \subset C[0, 1]$. Then, for a discontinuous function ρ we have $\rho \notin \text{Range}(A_F)$ and $F \in \mathcal{F} \setminus \mathcal{M}$.

Example 2: Let (X_*, Z) be as in Example 1 and $U = V + \eta$, where V is independent of Z , and $\eta = aI\{Z \leq 0\} - aI\{Z > 0\}$, with $a \neq 0$. Using the smoothness of $f_{X|Z}(x|z)$ w.r.t. z and the Lebesgue Theorem, it follows that $\text{Range}(A_F) \subset C[0, 1]$. Thus, $\rho \notin \text{Range}(A_F)$.

A similar argument is possible when $\text{Range}(A_F) \subset C^1[0, 1]$, and function ρ is not differentiable. In the Monte Carlo section we consider discontinuous and non-differentiable functions ρ to investigate the power of our testing procedure.

3 The test statistic and its asymptotic properties

3.1 The test statistic

Estimation of functional parameter φ_0 from conditional moment restriction (1) is an ill-posed inverse problem. Different estimation procedures have been proposed in the literature (see Ai and Chen (2003), Darolles, Florens, and Renault (2003), Newey and Powell (2003), Hall and Horowitz (2005)). Here we focus on the approach of Gagliardini and Scaillet (2006, GS), and consider the Tikhonov Regularised (TiR) estimator defined by

$$\hat{\varphi} = \arg \min_{\varphi \in \Theta} Q_T(\varphi) + \lambda_T \|\varphi\|_H^2, \text{ where } Q_T(\varphi) = \frac{1}{T} \sum_{t=1}^T \hat{\Omega}(Z_t) \left[\hat{r}(Z_t) - \hat{A}\varphi(Z_t) \right]^2,$$

$\hat{r}(z) = \int y \hat{f}(w|z) dw$, $\hat{A}\varphi(z) = \int \hat{f}(w|z) \varphi(x) dw$ and \hat{f} is a kernel estimator of f . We assume that parameter φ_0 belongs to a subset Θ of the Sobolev space $H^2[0, 1]$, which is the completion of the linear space $\{\varphi \in C^1[0, 1] \mid \nabla\varphi \in L^2[0, 1]\}$ w.r.t. 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 $\|\varphi\|_H = \langle \varphi, \varphi \rangle_H^{1/2}$. The minimum distance criterion Q_T is penalized by a term that involves the squared Sobolev norm $\|\varphi\|_H^2$. Penalization is required to overcome ill-posedness and is tuned by regularization parameter $\lambda_T > 0$, which converges to 0, as $T \rightarrow \infty$. Weighting function $\hat{\Omega}(z)$ is $\hat{\Omega}(z) = \hat{V}(z)^{-1}$, where $\hat{V}(z)$ is obtained by kernel smoothing using a first-step estimator $\bar{\varphi}$ and converges to $V_0[Y - \varphi_0(X) \mid Z = z] =: \Omega_0(z)^{-1}$ for any z , P -a.s.. The TiR estimator is given in closed form by

$$\hat{\varphi} = (\lambda_T + \hat{A}^* \hat{A})^{-1} \hat{A}^* \hat{r},$$

where \hat{A}^* denotes the linear operator defined by $\langle \varphi, \hat{A}^* \psi \rangle_H = \frac{1}{T} \sum_{t=1}^T \left(\hat{A} \varphi \right) (Z_t) \hat{\Omega}(Z_t) \psi(Z_t)$, for $\varphi \in H^2[0, 1]$ and ψ measurable. The operator \hat{A}^* is an estimator of the linear operator $A_{F_0}^*$, which denotes the adjoint operator of A_{F_0} w.r.t. the scalar products $\langle \cdot, \cdot \rangle_H$ and $\langle \cdot, \cdot \rangle_{L_{\Omega_0}^2(F_Z)}$, where $\langle \psi_1, \psi_2 \rangle_{L_{\Omega_0}^2(F_Z)} := E_0 \left[\psi_1(Z)' \Omega_0(Z) \psi_2(Z) \right]$.

The test statistic is build from the minimized criterion $Q_T(\hat{\varphi})$ after appropriate trimming, recentering and scaling. More specifically, we first replace the integrals w.r.t. kernel density estimator \hat{f} with kernel regression estimators which are easier to compute. Then, we use the asymptotic equivalence (see Appendix 2) between a trimmed version $\tilde{Q}_T(\hat{\varphi})$ of $Q_T(\hat{\varphi})$ and the statistic ξ_T :

$$\xi_T = \frac{1}{T} \sum_{t=1}^T \left(\sum_{s=1}^T \psi_{ts} \right)^2,$$

with

$$\psi_{ts} = \frac{\hat{\Omega}(Z_t)^{1/2} (Y_s - \hat{\varphi}(X_s)) K \left(\frac{Z_s - Z_t}{h_T} \right) I \{Z_t \in S_*\}}{\sum_{j=1}^T K \left(\frac{Z_j - Z_t}{h_T} \right)},$$

where K is a kernel, h_T is a bandwidth, and S_* is a compact subset of the support \mathcal{Z} of Z . Trimming based on a fixed S_* is standard in nonparametric specification testing for technical and practical reasons. As in TK, the use of a fixed support implies that the test is consistent only against alternatives for which (1) is violated on S_* .

3.2 The asymptotic distribution under the null hypothesis

Let us denote $K_{**} := \int (K * K)(x)^2 dx$, where $(K * K)(x) = \int K(y) K(x - y) dy$, and $vol(S_*) := \int_{S_*} dz$.

Proposition 1 Under Assumptions B.1-B.10 and H_0 :

$$\zeta_T := Th_T^{1/2} \frac{\xi_T - \xi_{2,T}}{\sigma} \xrightarrow{d} N(0, 1),$$

where $\xi_{2,T} = \frac{1}{T} \sum_{t=1}^T \sum_{s=1, s \neq t}^T \psi_{ts}^2$ and $\sigma^2 = 2K_{**} \text{vol}(S_*)$.

Proof. See Appendix 2. ■

In Appendix 2 we show that, under our list of regularity conditions and the null hypothesis, the test statistic ζ_T is asymptotically equivalent to $Th_T^{1/2} \xi_{5,T} / \sigma$ with $\xi_{5,T} := \frac{1}{T} \sum_{t=1}^T \sum_{s=1, s \neq t}^T \sum_{u=1, u \neq t, u \neq s}^T \psi_{ts} \psi_{tu}$. Their asymptotic distribution is unaffected by the use of estimate $\hat{\varphi}$ instead of true function φ_0 in the trimmed criterion \tilde{Q}_T . This explains why the asymptotic distribution of ζ_T under the null hypothesis is $N(0, 1)$ as for the specification test of a parametric conditional moment restriction in TK. In our ill-posed inverse problem setting, however, the control of the impact on the test statistic of estimation error is more complicated, because of the lower rate of convergence of the estimator $\hat{\varphi}$ (see GS for a discussion of such a rate). In particular, asymptotic negligibility of regularization bias and estimation variability requires an assumption on the rate of convergence of the regularization parameter λ_T in addition to an assumption on the rate of convergence of the bandwidth h_T (see Assumptions B.7 and B.10 in Appendix 1).

In the Monte Carlo section we report the finite sample performance of a test based on ζ_T . For programming purpose, this test statistic can be expressed in a matrix format:

$$\zeta_T = h_T^{1/2} [\iota' \Psi' \Psi \iota - \iota' (\Psi \odot \Psi) \iota + \text{trace}(\Psi \odot \Psi)] / \sigma,$$

where Ψ is the $T \times T$ matrix with elements ψ_{ts} , ι is a $T \times 1$ vector of ones, and \odot denotes the Hadamard (or element-by-element) product. We also consider an asymptotically equivalent

statistic based on the penalized value of the criterion, namely $\zeta_T + Th_T^{1/2} \lambda_T \|\hat{\varphi}\|_H^2 / \sigma$.²

Other possibilities include statistics such as:

$$Th_T^{1/2} \xi_{5,T} / \sigma = h_T^{1/2} [\iota' \Psi' \Psi \iota - 2 \text{diag}(\Psi)' \Psi \iota - \iota' (\Psi \odot \Psi) \iota + 2 \text{trace}(\Psi \odot \Psi)] / \sigma,$$

where $\text{diag}(\Psi)$ is the $T \times 1$ vector of the diagonal elements of Ψ , or its penalized counterpart $Th_T^{1/2} (\xi_{5,T} + \lambda_T \|\hat{\varphi}\|_H^2) / \sigma$. We have checked that Monte Carlo results for the latter two test statistics are qualitatively similar to those of the corresponding tests based on ζ_T .

3.3 Bootstrap computation of the critical values

In a GMM framework, asymptotic approximation can be bad, and bootstrapping provides one approach to improved inference (Hall and Horowitz (1996)). To establish the asymptotic distributional result of the previous section we have followed some of the developments in TK. They exploit the central limit theorem for generalized quadratic forms in de Jong (1987), which is a generalization of the central limit theorem for degenerate U -statistics in Hall (1984). The usual bootstrap of this type of statistics is known to fail. To get bootstrap consistency, an appropriate recentering is required (Arcones and Gine (1992)). Here we parallel the bootstrap construction of Horowitz (2006).³ His technique relies on sampling from a pseudo-true model which coincides with the original model if the null hypothesis is true, and satisfies a version of the conditional moment restriction if the null hypothesis is

² Statistic $\zeta_T + Th_T^{1/2} \lambda_T \|\hat{\varphi}\|_H^2 / \sigma$ is asymptotically equivalent to ζ_T if estimator $\hat{\varphi}$ is such that $\|\hat{\varphi}\|_H = O_p(1)$, and $\gamma > 1 - \bar{\eta}/2$ in Assumption B.10.

³ Other resampling techniques such as empirical likelihood bootstrap (Brown and Newey (2002)), m -out-of- n (moon) bootstrap (Bickel, Gotze and van Zwet (1997)), and subsampling (Politis, Romano and Wolf (1999)) provide other approaches to improved inference in our setting. They are however less simple to implement. The wild bootstrap (Haerdle and Mammen (1993)) and the simulation-based multiplier method (Hansen (1996)) cannot be used.

false. The idea is to get a bootstrap which imposes the conditional moment restriction on the resampled data regardless of whether the null hypothesis holds for the original model.

For a bootstrap test based on ζ_T the steps are as follows.

Bootstrap test algorithm

1. Compute $\bar{U}_t := Y_t - \hat{\varphi}(X_t) - \left(\hat{r}(Z_t) - \hat{A}\hat{\varphi}(Z_t) \right)$, $t = 1, \dots, T$.
2. Make T independent draws $(\tilde{X}_{t,b}, \tilde{Z}_{t,b}, \tilde{U}_{t,b})$ with replacement from $\{(X_t, Z_t, \bar{U}_t); 1 \leq t \leq T\}$, and take $\tilde{Y}_{t,b} := \hat{\varphi}(\tilde{X}_{t,b}) + \tilde{U}_{t,b}$ to get the bootstrap sample $(\tilde{X}_{t,b}, \tilde{Y}_{t,b}, \tilde{Z}_{t,b})$, $t = 1, \dots, T$.
3. Compute the bootstrap statistic $\tilde{\zeta}_{T,b}$ based on the bootstrap sample.
4. Repeat steps 2 and 3 B times, where B is an integer.
5. Reject the null hypothesis at significance level α if $p_B < \alpha$, where the bootstrap p -value is $p_B := \frac{1}{B} \sum_{b=1}^B I\{|\tilde{\zeta}_{T,b}| > |\zeta_T|\}$.

Step 2 implements the constraints $E[\tilde{Y} - \varphi_0(\tilde{X}) | \tilde{Z} = z] = 0$ and $V[\tilde{Y} - \varphi_0(\tilde{X}) | \tilde{Z} = z] = \Omega_0(z)^{-1}$ on the bootstrap sample whether H_0 holds or not. A test based on the decision rule in Step 5 is consistent: it satisfies $\lim P[\text{reject } H_0] = \alpha$ if H_0 is true, and $\lim P[\text{reject } H_0] = 1$ if H_0 is false. This can be justified by showing that the limit distribution of $\tilde{\zeta}_{T,b}$ is an independent copy of the limit distribution of ζ_T . The proof follows the same arguments as in the proof of Proposition 1 but applied to the bootstrap sample instead of the original sample. Therefore we omit these developments. Besides, the bootstrap provides asymptotic

refinements to critical values since we exploit an asymptotically pivotal test statistic. In our Monte Carlo results the bootstrap reduces significantly the finite sample distortions of level that occur when asymptotic critical values are used. Similar steps and comments hold for a bootstrap test based on the other asymptotic equivalent test statistics mentioned in Section 3.2.

4 A Monte-Carlo study

4.1 Data generating process under the null hypothesis

Following GS (see also Newey and Powell (2003)) we draw the errors U and V and the instrument Z as

$$\begin{pmatrix} U \\ V \\ Z \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{UV} & 0 \\ \rho_{UV} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right), \quad \rho_{UV} = .5,$$

and build $X_* = Z + V$. Then we map X_* into a variable $X = \Phi(X_*)$, which lives in $[0, 1]$.

The function Φ denotes the cdf of a standard Gaussian variable, and is assumed to be known.

We generate Y according to $Y = \sin(\pi X) + U$. Since the correlation ρ_{UV} is equal to .5 there is endogeneity, and an instrumental variable estimation is required. The moment condition is $E_0[Y - \varphi_0(X) | Z] = 0$, where the functional parameter is $\varphi_0(x) = \sin(\pi x)$, $x \in [0, 1]$. The chosen function resembles the shape of the Engel curve found in the empirical application.

4.2 Estimation procedure

The estimation procedure follows GS. To compute numerically the estimator $\hat{\varphi}$ we use a series approximation $\varphi(x) \simeq \theta'P(x)$ based on standardized shifted Chebyshev polynomials of the first kind (see Section 22 of Abramowitz and Stegun (1970) for their mathematical properties). These orthogonal polynomials are best suited for an unknown function φ_0 on $[0, 1]$. We take orders 0 to 5 which yields six coefficients to be estimated in the approximation $\varphi(x) \simeq \sum_{j=0}^5 \theta_j P_j(x)$, 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 $\|\varphi\|_H^2 = \int_0^1 \varphi^2 + \int_0^1 (\nabla\varphi)^2 \simeq \sum_{i=0}^5 \sum_{j=0}^5 \theta_i \theta_j \int_0^1 (P_i P_j + \nabla P_i \nabla P_j)$. The coefficients in the quadratic form $\theta' D \theta$ are explicitly computed with a symbolic calculus package:

$$D = \begin{pmatrix} \frac{1}{\pi} & 0 & \frac{-\sqrt{2}}{3\pi} & 0 & \frac{-\sqrt{2}}{15\pi} & 0 \\ \vdots & \frac{26}{3\pi} & 0 & \frac{38}{5\pi} & 0 & \frac{166}{21\pi} \\ & & \frac{218}{5\pi} & 0 & \frac{1182}{35\pi} & 0 \\ & & & \frac{3898}{35\pi} & 0 & \frac{5090}{63\pi} \\ \vdots & & & & \frac{67894}{315\pi} & 0 \\ \dots & & & & \dots & \frac{82802}{231\pi} \end{pmatrix}.$$

Such a simple and exact form eases implementation ⁴, and improves on speed.

The kernel estimator of the conditional moment $\hat{r}(z) - \hat{A}\varphi(z)$ is approximated through

⁴ The Gauss programs developed for this section and the empirical application are available on request from the authors.

$\hat{r}(z) = \theta' \hat{P}(z)$ where $\hat{P}(z) = \frac{\sum_{t=1}^T P(X_t) K\left(\frac{Z_t - z}{h_T}\right)}{\sum_{t=1}^T K\left(\frac{Z_t - z}{h_T}\right)}$, $\hat{r}(z) = \frac{\sum_{t=1}^T Y_t K\left(\frac{Z_t - z}{h_T}\right)}{\sum_{t=1}^T K\left(\frac{Z_t - z}{h_T}\right)}$, and K is the Gaussian kernel. The explicit form of the resulting ridge-type estimator $\hat{\theta}$ is given in GS. The bandwidth is selected via the standard rule of thumb $h = 1.06 \hat{\sigma}_Z T^{-1/5}$ (Silverman (1986)), where $\hat{\sigma}_Z$ is the empirical standard deviation of observed Z_t .⁵ Here weighting function $\Omega_0(z)$ is equal to unity, and assumed to be known.

4.3 Simulation results

The sample size is fixed at $T = 1000$. Size and power are computed with 1000 repetitions. We use a fixed trimming at 5% in the upper and lower tails, i.e., $S_* = [-1.645, 1.645]$. We look at a grid of values for the regularization parameter $\lambda \in \{.00001, .0007, .0009, .005\}$. The values .0009 and .0007 are the optimal values of λ minimizing the asymptotic MISE of the estimator, and minimizing the finite sample MISE, respectively (see GS for details on these computations). The data-driven procedure introduced in GS selects λ close to these optimal values. The values .00001 and .005 are far away from the optimal ones, and far beyond the quartiles of the distribution of the regularization parameters that are selected by the data-driven procedure.

Unreported simulation results show that the asymptotic approximation of Proposition 1 is poor for sample size $T = 1000$: test statistic distributions are asymmetric and size distortions are large. We often end up with no rejection at all of the null hypothesis at the

⁵ This choice is motivated by ease of implementation. Moderate deviations from this simple rule do not seem to affect estimation results significantly.

1% confidence level. In light of this, we advocate to use the bootstrap procedure of Section 3.3 for small to moderate sample sizes. The number of bootstrap samples is fixed at $B = 500$. In Table I, for each value of λ we report the rejection rates of statistic ζ_T (left column) and those of statistic $\zeta_T + Th_T^{1/2} \lambda_T \|\hat{\varphi}\|_H^2 / \sigma$ (right column), at nominal size $\alpha = .01, .05, .10$. For $\lambda = .0007, .0009$, statistic ζ_T provides undersized tests. Statistic $\zeta_T + Th_T^{1/2} \lambda_T \|\hat{\varphi}\|_H^2 / \sigma$ features better finite sample properties and yields tests which are only slightly undersized. In particular, for $\lambda = .0009$, the rejection rates are very close to the nominal ones. Selecting the too small regularization parameter $\lambda = .00001$ also results in undersized tests, both for ζ_T and $\zeta_T + Th_T^{1/2} \lambda_T \|\hat{\varphi}\|_H^2 / \sigma$. For the extremely large value $\lambda = .005$, the test becomes oversized because of regularization bias.

	Rejection rates with 1000 repetitions for ζ_T and $\zeta_T + Th_T^{1/2} \lambda_T \ \hat{\varphi}\ _H^2 / \sigma$							
	$\lambda = .00001$		$\lambda = .0007$		$\lambda = .0009$		$\lambda = .005$	
$\alpha = .01$.002	.000	.005	.010	.008	.010	.087	.198
$\alpha = .05$.016	.004	.013	.025	.033	.039	.271	.533
$\alpha = .10$.048	.063	.036	.072	.049	.092	.431	.737

TABLE I: Size of bootstrap test: $T = 1000, B = 500$

In Table II, we study the power of the bootstrap testing procedure based on ζ_T (left column) and $\zeta_T + Th_T^{1/2} \lambda_T \|\hat{\varphi}\|_H^2 / \sigma$ (right column). We generate Y as $Y = \sin(\pi X) + U + \eta$. In design 1 we take $\eta = .20I\{Z \leq 0\} - .20I\{Z > 0\}$. This yields $E_0[Y - \sin(\pi X) | Z = z] = .20I\{z \leq 0\} - .20I\{z > 0\}$, and the model specification is incorrect (discontinuity at point

$z = 0$; cf. discussion in Section 2). In design 2 we take $\eta = 0.80(|Z| - \sqrt{2/\pi})$ yielding another misspecification (non-differentiability at point $z = 0$). In both designs $U + \eta$ are maintained centered. The two cases mimick possible measurement errors in data such as the ones of the empirical section. In the first one, reported Y_t are larger in average when reported Z_t are known to be small, and vice-versa. In the second one, reported Y_t are larger in average when reported Z_t are known to be large in absolute value compared to their average value.

We find a satisfactory power for $\lambda = .0007, .0009$, under both designs. Since design 1 implies a stronger departure from the null hypothesis than design 2, the test statistics have overall better power properties in the first design. For value $\lambda = .00001$, the power is minimal.

	Rejection rates with 1000 repetitions for ζ_T and $\zeta_T + Th_T^{1/2} \lambda_T \ \hat{\varphi}\ _H^2 / \sigma$							
Design 1	$\lambda = .00001$		$\lambda = .0007$		$\lambda = .0009$		$\lambda = .005$	
$\alpha = .01$.006	.002	.648	.074	.742	.084	.896	.499
$\alpha = .05$.029	.012	.829	.290	.876	.337	.966	.811
$\alpha = .10$.073	.049	.859	.473	.902	.538	.976	.916
Design 2	$\lambda = .00001$		$\lambda = .0007$		$\lambda = .0009$		$\lambda = .005$	
$\alpha = .01$.005	.002	.056	.078	.082	.142	.963	1.000
$\alpha = .05$.013	.009	.201	.310	.251	.473	.995	1.000
$\alpha = .10$.043	.044	.289	.513	.361	.736	.998	1.000

TABLE II: Power of bootstrap test: $T = 1000, B = 500$

5 An empirical example

This section presents an empirical example with the data in Horowitz (2006) and GS.⁶

We aim at testing the specification of an Engel curve based on the moment condition $E_0 [Y - \varphi_0(X) | Z] = 0$, with $X = \Phi(X_*)$. Variable Y denotes the food expenditure share, X_* denotes the standardized logarithm of total expenditures, and Z denotes the standardized logarithm of annual income from wages and salaries. We have 785 household-level observations from the 1996 US Consumer Expenditure Survey. The estimation procedure is the same as in GS (see also the previous section). It relies on a kernel estimate of the conditional variance to get the weighting function and on a spectral approach to get a data-driven regularization parameter. The selected value is $\hat{\lambda} = .01113$. In GS the plotted estimated shape corroborates the findings of Horowitz (2006), who rejects a linear curve but not a quadratic curve at the 5% significance level to explain $\log Y$. Banks, Blundell and Lewbel (1997) consider demand systems that accommodate such empirical Engel curves. A specification test based on 1000 bootstrap samples yields bootstrap p -values of .426 and .671 for the test statistic values $\zeta_T = -.9826$ and $\zeta_T + Th_T^{1/2} \lambda_T \|\hat{\varphi}\|_H^2 / \sigma = -.3017$, respectively. Hence we do not reject the null hypothesis of a correct specification of the Engel curve modeling.

⁶ We would like to thank Joel Horowitz for kindly providing the dataset.

Appendix 1: List of regularity conditions

B.1: $\{(Y_t, X_t, Z_t) : t = 1, \dots, T\}$ is an i.i.d. sample from a distribution admitting a pdf f_{YXZ} with support $\mathcal{S} = \mathcal{Y} \times \mathcal{X} \times \mathcal{Z} \subset \mathbb{R}^d$, $\mathcal{X} = [0, 1]$, $d = 3$, such that: (i) $\sup_{\mathcal{X}, \mathcal{Z}} f_{X|Z} < \infty$; (ii) f_Z is in class $C^2(\mathbb{R})$.

B.2: For $m > 4$, $E_0[|U|^m] < \infty$, where $U := Y - \varphi_0(X)$.

B.3: Set $S_* \subset \mathcal{Z}$ is compact, contained in the interior of \mathcal{Z} such that $\inf_{S_*} f_Z > 0$.

B.4: The kernel K is (i) a pdf with support in $[-1, 1]$, (ii) symmetric, (iii) continuously differentiable, and (iv) bounded away from 0 on $[-a, a]$, for $a \in (0, 1)$.

B.5: The conditional variance $V_0(z) := V_0[U|Z = z]$ is in class $C^2(\mathbb{R})$, such that $\inf_{S_*} V_0 > 0$.

B.6: Estimators $\hat{\varphi}$ and $\bar{\varphi}$ are such that: (i) $\frac{1}{T} \sum_t |\Delta \hat{\varphi}(X_t)|^2 = O_p(T^{-1/3})$, where $\Delta \hat{\varphi} := \hat{\varphi} - \varphi_0$; (ii) $\sup_{\mathcal{X}} |\nabla^2 \hat{\varphi}| = O_p(1)$; (iii) $\sup_{z \in S_*} \left| \frac{1}{Th_T} \sum_t K\left(\frac{z - Z_t}{h_T}\right) |\Delta \bar{\varphi}(X_t)|^2 \right| = o_p(T^{-\varepsilon})$, for some $\varepsilon > 2/m$, where $\Delta \bar{\varphi} := \bar{\varphi} - \varphi_0$; (iv) $\frac{1}{T} \sum_t \mathcal{R}_T(X_t)^2 = o_p(T^{-1}h_T^{1/2})$, where $\mathcal{R}_T := \Delta \hat{\varphi} - \left[(\lambda_T + A_{F_0}^* A_{F_0})^{-1} A_{F_0}^* A_{F_0} - 1 \right] \varphi_0 - (\lambda_T + A_{F_0}^* A_{F_0})^{-1} A_{F_0}^* \hat{\psi}$ and $\hat{\psi}(z) := \int (y - \varphi_0(x)) \frac{\hat{f}_{YXZ}(w, z)}{f_Z(z)} dw$; (v) $\frac{1}{T^2 h_T} \sum_t \sum_{s \neq t} K\left(\frac{Z_s - Z_t}{h_T}\right)^2 \mathcal{R}_T(X_s)^2 I(Z_t \in S_*) = o_p(T^{-1}h_T)$.

B.7: The bandwidth h_T is such that $h_T = \bar{c}T^{-\bar{\eta}}$ for $2/9 < \bar{\eta} < \min\{1 - 4/m, 1/3, \varepsilon - 2/m\}$.

B.8: (i) The eigenvalues ν_j of operator $A_{F_0}^* A_{F_0}$ are such that $C_1 e^{-\alpha j} \leq \nu_j \leq C_2 e^{-\alpha j}$, $j \in \mathbb{N}$, for some constants $\alpha > 0$, $C_1 \leq C_2$; (ii) The orthonormal eigenfunctions ϕ_j , $j \in \mathbb{N}$, of

operator $A_{F_0}^* A_{F_0}$ are such that $\sup_{j \in \mathbb{N}} \sup_{|u| \leq 1} E_0 [\nabla^2 (A_{F_0} \phi_j) (Z + h_T u)^2] = O(1)$, as $h_T \rightarrow 0$.

B.9: The function $\varphi_0 \in H^2[0, 1]$ is such that: (i) φ_0 is in class $C^2(0, 1)$ with $\sup_{\mathcal{X}} |\nabla^2 \varphi_0| < \infty$;

(ii) $\sum_{j=1}^{\infty} \frac{\nu_j \langle \varphi_0, \phi_j \rangle_H^2}{(\lambda_T + \nu_j)^2} = O(\lambda_T^{-\delta})$ as $\lambda_T \rightarrow 0$, for $\delta < 1$.

B.10: The regularisation parameter λ_T is such that $\lambda_T = cT^{-\gamma}$ for $\frac{1 - \bar{\eta}/2}{2 - \delta} < \gamma < \min\{4\bar{\eta}, 1\}$.

Assumptions B.1-B.5 yield the assumptions used in TK for testing parametric conditional moment restrictions in the special case of a linear-in-parameter moment function. In our functional setting, compactness of \mathcal{X} in Assumption B.1 eases the definition of the parameter space, which is a subset of the Sobolev space $H^2[0, 1]$. Assumption B.1 (i) on the conditional density $f_{X|Z}$ implies that operator $A_{F_0} : L^2[0, 1] \rightarrow L^2_{\Omega_0}(F_Z)$ is compact, and this yields compactness of $A_{F_0}^* A_{F_0}$. Assumption B.1 (ii) on the conditional density f_Z , together with Assumption B.4 on the kernel, allows to exploit the results on uniform convergence of kernel estimators in Newey (1994) and a result of Devroye and Wagner (1980). Assumption B.2 is a condition ensuring finite higher moments of the innovation. The compact set S_* in Assumption B.3 solves boundary problems of kernel estimators. Assumptions B.5 on the conditional variance $V_0(z)$ is used in the proof of Lemma C.1. Assumption B.6 collects properties of the functional estimators $\hat{\varphi}$ and $\bar{\varphi}$ that are used in the proofs of technical lemmas. Specifically, Assumption B.6 (i) is used in the proof of Lemmas A.1 and A.2 to prove asymptotic negligibility of two components of the test statistic. Assumption B.6 (ii) is used to prove the asymptotic equivalence of the two statistics \tilde{Q}_T and ξ_T in Section

A.2.1. Assumption B.6 (iii) concerns the first-step estimator $\bar{\varphi}$ in the estimator $\hat{V}(z)$ of the conditional variance $V_0(z)$, and is used in Lemma C.1 to prove the convergence of $\hat{V}(z)$ and of its inverse. Function \mathcal{R}_T in Assumptions B.6 (iv)-(v) is the reminder term in the asymptotic expansion of $\Delta\hat{\varphi}$ w.r.t. regularisation bias $\left[(\lambda_T + A_{F_0}^* A_{F_0})^{-1} A_{F_0}^* A_{F_0} - 1\right] \varphi_0$ and sample variability $(\lambda_T + A_{F_0}^* A_{F_0})^{-1} A_{F_0}^* \hat{\psi}$. Assumptions B.6 (iv)-(v) allow us to control the reminder contribution coming from \mathcal{R}_T in Lemmas A.7 and A.8. The condition on the bandwidth h_T is given in B.7. Condition $\bar{\eta} < \min\{1 - 4/m, 1/3, \varepsilon - 2/m\}$ corresponds to the condition in Theorem 4.1 of TK for a linear-in-parameter moment condition ($\eta = \infty$ in Assumption 3.6 of TK) when we take $\varepsilon = 1$ (parametric rate of convergence of the estimator). Condition $\bar{\eta} > 2/9$ is used to prove the asymptotic equivalence of \tilde{Q}_T and ξ_T in Section A.2.1. Assumption B.8 concerns the spectral decomposition of compact operator $A_{F_0}^* A_{F_0}$ (see Kress (1999), Chapter 14). In Assumption B.8 (i) the spectrum of $A_{F_0}^* A_{F_0}$ is supposed to feature geometric decay, which corresponds to settings with severe ill-posedness. This assumption simplifies the control of series involving eigenvalues ν_j of operator $A_{F_0}^* A_{F_0}$ such as $\sum_{j=1}^{\infty} \frac{\nu_j}{\lambda_T + \nu_j}$ in the proofs of Lemmas A.5 and A.6. In GS we verify that Assumption B.8 (i) is satisfied in the Monte-Carlo setting of Section 4. Our results extend to the case of hyperbolic decay (mild ill-posedness). Regularity Assumption B.8 (ii) on the eigenfunctions ϕ_j is used in the proof of technical Lemmas A.5 and A.6. Assumption B.9 concerns the smoothness of function φ_0 . Second-order differentiability of φ_0 (Assumption B.9 (i)) is used to control the estimation bias term in $\hat{\psi}$ induced by kernel density estimator $\hat{f}(y|z)$ (see the proof of Lemma A.5). We could dispense of this assumption by adopting a different estimator of function

$A_{F_0}^* r_{F_0}$ to define estimator $\hat{\varphi}$ (see GS, footnote 8, and Hall and Horowitz (2005)). However, we would lose the interpretation of minimum distance estimator for $\hat{\varphi}$. Since estimator $\hat{\varphi}$ is not the focus of this paper, we do not detail modifications induced by alternative estimation approaches. Assumption B.9 (ii) involves Fourier coefficients $\langle \varphi_0, \phi_j \rangle_H$ w.r.t. the basis of eigenfunctions of operator $A_{F_0}^* A_{F_0}$. The slower the decay rate of $\langle \varphi_0, \phi_j \rangle_H$ with order j , the larger the coefficient δ . Assumption B.9 (ii) is tightly related to the so-called source condition (e.g., Darolles, Florens, Renault (2003)). Finally, Assumption B.10 is the condition on the rate of convergence of the regularization parameter λ_T . Condition $\gamma > \frac{1 - \bar{\eta}/2}{2 - \delta}$ is used to prove asymptotic negligibility of the terms in the test statistic coming from regularization bias (see Section A.2.3), while negligibility of sample variability from estimation of function φ_0 is implied by $\gamma < \min\{4\bar{\eta}, 1\}$ (see Section A.2.4).

Appendix 2: Asymptotic distribution under the null hypothesis

In this appendix we omit subscripts in densities, expectations and operators, hoping that context brings clarity. Furthermore, let $\mathcal{T}_* = \{1 \leq t \leq T : Z_t \in S_*\}$, $K_{st} = K\left(\frac{Z_s - Z_t}{h_T}\right)$, $\hat{\Omega}_t = \hat{\Omega}(Z_t)$, $U_t = Y_t - \varphi_0(X_t)$, $g_{\varphi_0}(w) = y - \varphi_0(x)$, $\Delta\varphi = \varphi - \varphi_0$, $I_t = I(Z_t \in S_*)$, $\mathcal{I} = \{Z_t : 1 \leq t \leq T\}$, $\hat{H}(z) = \hat{V}(z)\hat{f}(z)^2$, $H_0(z) = V_0(z)f(z)^2$, $\hat{\Sigma}(Z_t) = \frac{1}{Th_T} \sum_j K_{jt} U_j^2$, $\tilde{H}(Z_t) = E\left[\hat{\Sigma}(Z_t)|Z_t\right] E\left[\hat{f}(Z_t)|Z_t\right]$, $w_{st} = K_{st}/\sum_j K_{jt}$.

A.2.1 Asymptotically equivalent statistics

Let us consider a trimmed version of the minimized criterion

$$\tilde{Q}_T(\hat{\varphi}) = \frac{1}{T} \sum_{t=1}^T I_t \hat{\Omega}_t \left[\int (y - \hat{\varphi}(x)) \hat{f}(w|Z_t) dw \right]^2,$$

and introduce $\xi_T := \frac{1}{T} \sum_{t=1}^T \left(\sum_{s=1}^T \psi_{ts} \right)^2$, where $\psi_{ts} = \hat{\Omega}_t^{1/2} I_t (Y_s - \hat{\varphi}(X_s)) K_{st} / \sum_{j=1}^T K_{jt}$.

Statistic ξ_T corresponds to statistic \hat{T} at p. 2064 in TK, but with a functional estimator $\hat{\varphi}$ of parameter φ_0 . Using Assumption B.6 (ii) to get the asymptotic equivalence

$$\int (y - \hat{\varphi}(x)) \hat{f}(w|z) dw = \frac{\sum_{s=1}^T (Y_s - \hat{\varphi}(X_s)) K\left(\frac{Z_s - z}{h_T}\right)}{\sum_{s=1}^T K\left(\frac{Z_s - z}{h_T}\right)} + O_p(h_T^2),$$

uniformly in $z \in S_*$, and Cauchy-Schwarz inequality, we get $\tilde{Q}_T(\hat{\varphi}) = \xi_T + O_p(\xi_T^{1/2} h_T^2) + O_p(h_T^4)$. Using $h_T = \bar{c} T^{-\bar{\eta}}$ with $\bar{\eta} > 2/9$, we get $\tilde{Q}_T(\hat{\varphi}) = \xi_T + o_p((Th_T^{1/2})^{-1})$. Thus, statistics $\tilde{Q}_T(\hat{\varphi})$ and ξ_T are asymptotically equivalent to define the test.

We use the decomposition $\xi_T = \xi_{1,T} + \xi_{2,T} + \xi_{3,T} + \xi_{4,T} + \xi_{5,T}$ as in TK, where

$$\begin{aligned} \xi_{1,T} &= \frac{1}{T} \sum_{t=1}^T \psi_{tt}^2, & \xi_{2,T} &= \frac{1}{T} \sum_{t=1}^T \sum_{s=1, s \neq t}^T \psi_{ts}^2, \\ \xi_{3,T} &= \frac{1}{T} \sum_{t=1}^T \sum_{s=1, s \neq t}^T \psi_{ts} \psi_{tt} &= \xi_{4,T} \\ \xi_{5,T} &= \frac{1}{T} \sum_{t=1}^T \sum_{s=1, s \neq t}^T \sum_{u=1, u \neq t, u \neq s}^T \psi_{ts} \psi_{tu}. \end{aligned}$$

Terms $\xi_{1,T}$, $\xi_{3,T}$ and $\xi_{4,T}$ are $o_p((Th_T^{1/2})^{-1})$ (see Lemmas A.1 and A.2 in Section A.2.6), while term $\xi_{5,T}$ after appropriate rescaling is asymptotically normal (see Section A.2.2). Thus, the

test statistic is based on the difference $\xi_T - \xi_{2,T}$ satisfying

$$\xi_T - \xi_{2,T} = \xi_{5,T} + o_p((Th_T^{1/2})^{-1}). \quad (5)$$

A.2.2 Asymptotic normality

We now rewrite statistic $\xi_{5,T}$ in order to identify the contribution coming from estimation of φ_0 . We use the asymptotic expansion of the TiR estimator $\hat{\varphi}$ given in GS, Appendix 3:

$$\hat{\varphi} - \varphi_0 = (\lambda_T + A^*A)^{-1}A^*\hat{\psi} + [(\lambda_T + A^*A)^{-1}A^*A - 1]\varphi_0 + \mathcal{R}_T =: \mathcal{V}_T + \mathcal{B}_T + \mathcal{R}_T,$$

where \mathcal{V}_T is the variance term, $\hat{\psi}(z) := \int g_{\varphi_0}(w) \frac{\hat{f}(w, z)}{f(z)} dw$, \mathcal{B}_T is the (asymptotic) regularization bias, and \mathcal{R}_T is a reminder term whose expression is given in Equation (36) of GS. Thus, $\psi_{ts} = \frac{\hat{\Omega}_t^{1/2} I_t}{\sum_j K_{jt}} [U_s - \mathcal{V}_T(X_s) - \mathcal{B}_T(X_s) - \mathcal{R}_T(X_s)] K_{st}$. We get the decomposition

$\xi_{5,T} = \xi_{5,T}^* + \xi_{5,T}^B + \xi_{5,T}^V + \xi_{5,T}^R$, where the leading contribution is

$$\xi_{5,T}^* = \frac{1}{T} \sum_t \frac{\hat{\Omega}_t I_t}{\left(\sum_j K_{jt}\right)^2} \sum_{s \neq t} \sum_{u \neq t, s} K_{st} K_{ut} U_s U_u,$$

the contribution induced by regularization bias is given by

$$\begin{aligned} \xi_{5,T}^B &= \frac{1}{T} \sum_t \frac{\hat{\Omega}_t I_t}{\left(\sum_j K_{jt}\right)^2} \sum_{s \neq t} \sum_{u \neq t, s} K_{st} K_{ut} \mathcal{B}_T(X_s) \mathcal{B}_T(X_u) \\ &\quad - \frac{2}{T} \sum_t \frac{\hat{\Omega}_t I_t}{\left(\sum_j K_{jt}\right)^2} \sum_{s \neq t} \sum_{u \neq t, s} K_{st} K_{ut} U_s \mathcal{B}_T(X_u) \\ &=: J_{1,T} - 2J_{2,T} \end{aligned}$$

the contribution accounting for estimation variability is given by

$$\begin{aligned}
\xi_{5,T}^V &= \frac{1}{T} \sum_t \frac{\hat{\Omega}_t I_t}{\left(\sum_j K_{jt}\right)^2} \sum_{s \neq t} \sum_{u \neq t,s} K_{st} K_{ut} \mathcal{V}_T(X_s) \mathcal{V}_T(X_u) \\
&\quad - \frac{2}{T} \sum_t \frac{\hat{\Omega}_t I_t}{\left(\sum_j K_{jt}\right)^2} \sum_{s \neq t} \sum_{u \neq t,s} K_{st} K_{ut} [U_s - \mathcal{B}_T(X_s)] \mathcal{V}_T(X_u) \\
&=: J_{3,T} - 2J_{4,T}
\end{aligned}$$

and finally the reminder contribution is given by

$$\begin{aligned}
\xi_{5,T}^R &= \frac{1}{T} \sum_t \frac{\hat{\Omega}_t I_t}{\left(\sum_j K_{jt}\right)^2} \sum_{s \neq t} \sum_{u \neq t,s} K_{st} K_{ut} \mathcal{R}_T(X_s) \mathcal{R}_T(X_u) \\
&\quad - \frac{2}{T} \sum_t \frac{\hat{\Omega}_t I_t}{\left(\sum_j K_{jt}\right)^2} \sum_{s \neq t} \sum_{u \neq t,s} K_{st} K_{ut} [U_s - \mathcal{B}_T(X_s) - \mathcal{V}_T(X_s)] \mathcal{R}_T(X_u) \\
&=: J_{5,T} - 2J_{6,T}.
\end{aligned}$$

Statistic $\xi_{5,T}^*$ corresponds to statistic $\hat{T}_5^{(1)}$ of TK, p. 2083 (multiplied by T^{-1}). Along the lines of the proofs of Lemmas A.6 and A.7 in TK, $\xi_{5,T}^* = \frac{1}{T^3 h_T^2} \sum_t \tilde{H}(Z_t)^{-1} I_t \sum_{s \neq t} \sum_{u \neq t,s} K_{st} K_{ut} U_s U_u + O_p\left(\frac{\log T}{T h_T} \sup_{z \in S_*} \left| \hat{H}(z)^{-1} - \tilde{H}(z)^{-1} \right| \right)$. Using Lemma C.1 (iv) in Section A.2.6, Lemma A.6 in TK, and $h_T = \bar{c} T^{-\bar{\eta}}$ with $\bar{\eta} < \min\{1 - 4/m, 1/3, \varepsilon - 2/m\}$, we get $T h_T^{1/2} \xi_{5,T}^* \xrightarrow{d} N(0, 2K_{**} \text{vol}(S_*))$. Then, Proposition 1 follows using that $\xi_{5,T}^B, \xi_{5,T}^V, \xi_{5,T}^R = o_p((T h_T^{1/2})^{-1})$ (see Sections A.2.3-A.2.5).

A.2.3 Control of the bias contribution

It follows from Lemmas A.3 for $J_{1,T}$ and A.4 for $J_{2,T}$ in Section A.2.6 that

$$\begin{aligned}
\xi_{5,T}^B &= E \left[\Omega_0(Z_t) I_t [(A\mathcal{B}_T)(Z_t)]^2 \right] (1 + o_p(1)) + O_p \left(\frac{1}{\sqrt{T}} E \left[\Omega_0(Z_t) I_t [(A\mathcal{B}_T)(Z_t)]^2 \right]^{1/2} \right) \\
&\quad + o_p((T h_T^{1/2})^{-1}).
\end{aligned}$$

Rewriting $\mathcal{B}_T = -\lambda_T (\lambda_T + A^*A)^{-1} \varphi_0$ and developing φ_0 w.r.t. the basis of eigenfunctions ϕ_j of A^*A , we have

$$E [\Omega_0(Z_t) I_t [(A\mathcal{B}_T)(Z_t)]^2] \leq E [\Omega_0(Z_t) [(A\mathcal{B}_T)(Z_t)]^2] = \langle \mathcal{B}_T, A^*A\mathcal{B}_T \rangle_H = \lambda_T^2 \sum_{j=1}^{\infty} \frac{\nu_j \langle \varphi_0, \phi_j \rangle_H^2}{(\lambda_T + \nu_j)^2}.$$

Using Assumptions B.9 (ii) and B.10, we get $\xi_{5,T}^B = o_p((Th_T^{1/2})^{-1})$.

A.2.4 Control of the variance contribution

From Lemmas A.5 for $J_{3,T}$ and A.6 for $J_{4,T}$ in Section A.2.6, we have

$$\xi_{5,T}^V = o_p \left(\frac{1}{\sqrt{Th_T^{1/2}}} E [\Omega_0(Z_t)^2 I_t [(A\mathcal{B}_T)(Z_t)]^2]^{1/2} \right) + o_p((Th_T^{1/2})^{-1}).$$

Thus, we get $\xi_{5,T}^V = o_p((Th_T^{1/2})^{-1})$ from a similar argument as in A.2.3.

A.2.5 Control of the reminder contribution

From Lemmas A.7 for $J_{5,T}$ and A.8 for $J_{6,T}$ in Section A.2.6, we deduce $\xi_{5,T}^R = o_p((Th_T^{1/2})^{-1})$.

A.2.6 Technical Lemmas

Lemmas A.1 and A.2 are akin to Lemmas A.2 and A.4 in TK. The major technical novelties are in proving Lemmas A.3-A.8 where we use conditions on the decay of the spectrum (Assumption B.8) and on the regularization parameter (Assumption B.10). To minimize the complexity of the presentation we assume $V_0(z) = \Omega_0(z) = 1$ in some steps in the proofs of Lemmas A.5-A.6. Lemma C.1 is akin to Lemmas C.2 and C.3 in TK. All proofs are given under Assumptions B.1-B.10 and H_0 .

Lemma A.1: $\xi_{1,T} = O_p((Th_T)^{-2})$.

Proof: The result follows from:

$$|\xi_{1,T}| \leq \max_{t \in \mathcal{I}_*} \left| \frac{(Th_T)^2 \hat{\Omega}_t}{(\sum_j K_{jt})^2} \right| \frac{K(0)^2}{(Th_T)^2} \frac{1}{T} \sum_t (|U_t|^2 + 2|U_t| |\Delta \hat{\varphi}(X_t)| + |\Delta \hat{\varphi}(X_t)|^2),$$

and $\max_{t \in \mathcal{I}_*} \left| \frac{(Th_T)^2 \hat{\Omega}_t}{(\sum_j K_{jt})^2} \right| = O_p(1)$, $\frac{1}{T} \sum_t |U_t|^2 = O_p(1)$, $\frac{1}{T} \sum_t |\Delta \hat{\varphi}(X_t)|^2 = o_p(1)$ (Lemma C.1 (ii) and Assumptions B.1-B.4, B.5, B.6 (i), B.7). ■

Lemma A.2: $\xi_{3,T} = o_p((Th_T^{1/2})^{-1})$.

Proof: We get:

$$\begin{aligned} \xi_{3,T} &= \frac{1}{T} \sum_t \sum_{s \neq t} \frac{\hat{\Omega}_t K(0)}{(\sum_j K_{jt})^2} U_t U_s K_{st} I_t - 2 \frac{1}{T} \sum_t \sum_{s \neq t} \frac{\hat{\Omega}_t K(0)}{(\sum_j K_{jt})^2} U_t \Delta \hat{\varphi}(X_s) K_{st} I_t \\ &\quad + \frac{1}{T} \sum_t \sum_{s \neq t} \frac{\hat{\Omega}_t K(0)}{(\sum_j K_{jt})^2} \Delta \hat{\varphi}(X_t) \Delta \hat{\varphi}(X_s) K_{st} I_t \\ &=: \xi_{31,T} + \xi_{32,T} + \xi_{33,T}. \end{aligned}$$

The first term, $\xi_{31,T}$, corresponds to statistic $\hat{T}_3^{(1)}$ of TK, p. 2082 (multiplied by T^{-1}). Along the lines of Lemma A.4 in TK, we have $\xi_{31,T} = O_p\left(\frac{1}{(Th_T)^{3/2}}\right) O_p\left(\sup_{z \in S_*} |\hat{H}(z)^{-1} - \tilde{H}(z)^{-1}|\right)$. From Lemma C.1 (iv) and $h_T = \bar{c}T^{-\bar{\eta}}$ with $\bar{\eta} < \min\{2/3, 1/2 + \varepsilon/2 - 1/m\}$, we get $\xi_{31,T} = o_p((Th_T^{1/2})^{-1})$. Let us now consider the third term, $\xi_{33,T}$. We have

$$|\xi_{33,T}| \leq \max_{t \in \mathcal{I}_*} \left| \frac{(Th_T)^2 \hat{\Omega}_t}{(\sum_j K_{jt})^2} \right| \frac{K(0)}{Th_T} \frac{1}{T^2 h_T} \sum_t \sum_{s \neq t} |\Delta \hat{\varphi}(X_t)| |\Delta \hat{\varphi}(X_s)| K_{st} I_t.$$

Applying Cauchy-Schwarz inequality twice, we deduce:

$$\frac{1}{T^2 h_T} \sum_t \sum_{s \neq t} |\Delta \hat{\varphi}(X_t) \Delta \hat{\varphi}(X_s)| K_{st} I_t \leq \frac{1}{T} \sum_t |\Delta \hat{\varphi}(X_t)|^2 \sqrt{\frac{1}{T^2 h_T^2} \sum_t \sum_{s \neq t} K_{st}^2 I_t}.$$

From $E \left[\frac{1}{T^2 h_T^2} \sum_t \sum_{s \neq t} K_{st}^2 I_t \right] = O(h_T^{-1})$ we get $\xi_{33,T} = O_p \left(\frac{1}{T h_T^{1/2} h_T} \left(\frac{1}{T} \sum_t |\Delta \hat{\varphi}(X_t)|^2 \right) \right)$.

It follows $\xi_{33,T} = o_p((T h_T^{1/2})^{-1})$ from Assumptions B.6 (i) and B.7. The argument for $\xi_{32,T}$ is similar. ■

Lemma A.3: $J_{1,T} = E \left[\Omega_0(Z_t) I_t [(A\mathcal{B}_T)(Z_t)]^2 \right] (1 + o_p(1)) + o_p((T h_T^{1/2})^{-1})$.

Proof: Define $\eta_s := \mathcal{B}_T(X_s) - E[\mathcal{B}_T(X_s) | Z_s]$ and $b_s := E[\mathcal{B}_T(X_s) | Z_s] = (A\mathcal{B}_T)(Z_s)$. Split $J_{1,T}$ into

$$\begin{aligned} J_{1,T} &= \frac{1}{T} \sum_t \frac{\hat{\Omega}_t I_t}{\left(\sum_j K_{jt} \right)^2} \sum_{s \neq t} \sum_{u \neq t, s} K_{st} K_{ut} b_s b_u + 2 \frac{1}{T} \sum_t \frac{\hat{\Omega}_t I_t}{\left(\sum_j K_{jt} \right)^2} \sum_{s \neq t} \sum_{u \neq t, s} K_{st} K_{ut} b_s \eta_u \\ &\quad + \frac{1}{T} \sum_t \frac{\hat{\Omega}_t I_t}{\left(\sum_j K_{jt} \right)^2} \sum_{s \neq t} \sum_{u \neq t, s} K_{st} K_{ut} \eta_s \eta_u \\ &=: J_{11,T} + J_{12,T} + J_{13,T}. \end{aligned}$$

Then, term $J_{11,T}$ can be written as

$$\begin{aligned} J_{11,T} &= \frac{1}{T} \sum_t \frac{(T h_T)^2 \hat{\Omega}_t I_t}{\left(\sum_j K_{jt} \right)^2} \frac{1}{T^2 h_T^2} \left(\sum_{s \neq t} K_{st} b_s \right)^2 - \frac{1}{T} \sum_t \frac{(T h_T)^2 \hat{\Omega}_t I_t}{\left(\sum_j K_{jt} \right)^2} \frac{1}{T^2 h_T^2} \sum_{s \neq t} K_{st}^2 b_s^2 \\ &=: J_{111,T} - J_{112,T}, \end{aligned}$$

where $J_{111,T}$ is the dominant term. Using

$$\begin{aligned} J_{111,T} &= \frac{1}{T} \sum_t \frac{\Omega_0(Z_t) I_t}{f(Z_t)^2} \frac{1}{T^2 h_T^2} \left(\sum_{s \neq t} K_{st} b_s \right)^2 \\ &\quad + \frac{1}{T} \sum_t \left[\frac{(T h_T)^2 \hat{\Omega}_t}{\left(\sum_j K_{jt} \right)^2} - \frac{\Omega_0(Z_t)}{f(Z_t)^2} \right] I_t \frac{1}{T^2 h_T^2} \left(\sum_{s \neq t} K_{st} b_s \right)^2, \\ E \left[\frac{\Omega_0(Z_t) I_t}{f(Z_t)^2} \frac{1}{T^2 h_T^2} \left(\sum_{s \neq t} K_{st} b_s \right)^2 \right] &= E \left[\Omega_0(Z_t) I_t [(A\mathcal{B}_T)(Z_t)]^2 \right] (1 + o(1)), \end{aligned}$$

$$\inf_{z \in \mathcal{S}_*} \frac{\Omega_0(z)}{f(z)^2} > 0, \quad \sup_{t \in \mathcal{T}^*} \left| \frac{(Th_T)^2 \hat{\Omega}_t}{\left(\sum_j K_{jt}\right)^2} - \frac{\Omega_0(Z_t)}{f(Z_t)^2} \right| = o_p(1) \quad (\text{Lemma C.1 (iii)}), \quad \text{we deduce}$$

$$J_{111,T} = E \left[\Omega_0(Z_t) I_t [(A\mathcal{B}_T)(Z_t)]^2 \right] (1 + o_p(1)).$$

Terms $J_{12,T}$ and $J_{13,T}$ can be analyzed similarly, and we consider only $J_{13,T}$ in details.

Write

$$\begin{aligned} J_{13,T} &= \frac{1}{T} \sum_t \frac{\Omega_0(Z_t) I_t}{f(Z_t)^2} \frac{1}{T^2 h_T^2} \sum_{s \neq t} \sum_{u \neq t, s} K_{st} K_{ut} \eta_s \eta_u \\ &\quad + \frac{1}{T} \sum_t \left[\frac{(Th_T)^2 \hat{\Omega}_t}{\left(\sum_j K_{jt}\right)^2} - \frac{\Omega_0(Z_t)}{f(Z_t)^2} \right] \frac{1}{T^2 h_T^2} \sum_{s \neq t} \sum_{u \neq t, s} K_{st} K_{ut} \eta_s \eta_u \\ &=: J_{131,T} + J_{132,T}. \end{aligned}$$

Note that $E[\eta_s | \mathcal{I}] = 0$ and $E[\eta_s \eta_u | \mathcal{I}] = 0$ for $s \neq u$, from the independence of the observations. Along the lines of Lemma A.7 in TK, using Assumptions B.1-B.5, B.7 and Lemma C.1 (iii) we can prove that $J_{132,T} = o_p((Th_T^{1/2})^{-1})$. Moreover, we have $J_{131,T} = \frac{1}{T} \frac{1}{T^2 h_T^2} J_{1,T}^*$, where $J_{1,T}^* = \sum_s \sum_{u>s} c_{su} \eta_s \eta_u$ and $c_{su} := 2 \sum_{t \neq s, u} \frac{\Omega_0(Z_t) I_t}{f(Z_t)^2} K_{st} K_{ut}$. Then, we get

$$E[J_{1,T}^*] = \sum_s \sum_{u>s} E[c_{su}^2 \eta_s^2 \eta_u^2] = \sum_s \sum_{u>s} E[c_{su}^2 \Gamma(Z_s) \Gamma(Z_u)],$$

where $\Gamma(Z_s) := E[\eta_s^2 | Z_s] = V[\mathcal{B}_T(X_s) | Z_s]$, and the cross-terms vanish because of the conditional independence property of the η_s variables. To compute $E[c_{su}^2 \Gamma(Z_s) \Gamma(Z_u)]$, we can use an argument similar to that in Lemma A.8 of TK, to get $E[c_{su}^2 \Gamma(Z_s) \Gamma(Z_u)] = O\left(T^2 h_T^3 E\left[\frac{\Omega_0(Z_t) I_t}{f(Z_t)} \Gamma(Z_t)\right]^2\right)$. Using Assumptions B.1, B.3, B.4, we have $E\left[\frac{\Omega_0(Z_t) I_t}{f(Z_t)} \Gamma(Z_t)\right] \leq \text{const} \cdot b(\lambda_T)^2$, where $b(\lambda_T) := \langle \mathcal{B}_T, \mathcal{B}_T \rangle^{1/2} = o(1)$. Thus, we deduce that $J_{131,T} = o_p((Th_T^{1/2})^{-1})$. The conclusion follows. ■

Lemma A.4: $J_{2,T} = O_p \left(\frac{1}{\sqrt{T}} E \left[\Omega_0(Z_t)^2 I_t [(AB_T)(Z_t)]^2 \right]^{1/2} \right) + o_p((Th_T^{1/2})^{-1})$.

Proof: With the notation in the proof of Lemma A.3 we have

$$\begin{aligned} J_{2,T} &= \frac{1}{T} \sum_t \frac{\hat{\Omega}_t I_t}{\left(\sum_j K_{jt} \right)^2} \sum_{s \neq t} \sum_{u \neq t, s} K_{st} K_{ut} U_s b_u \\ &\quad + \frac{1}{T} \sum_t \frac{\hat{\Omega}_t I_t}{\left(\sum_j K_{jt} \right)^2} \sum_{s \neq t} \sum_{u \neq t, s} K_{st} K_{ut} U_s \eta_u \\ &=: J_{21,T} + J_{22,T}. \end{aligned}$$

Let us first consider $J_{21,T}$. By assumptions B.1-B.5, B.7, Lemma C.1 (iii) and an argument similar to Lemma A.7 of TK, we have

$$\begin{aligned} J_{21,T} &= \frac{1}{T} \frac{1}{T^2 h_T^2} \sum_t \frac{\Omega_0(Z_t) I_t}{f(Z_t)^2} \sum_{s \neq t} \sum_{u \neq t, s} K_{st} K_{ut} U_s b_u + o_p((Th_T^{1/2})^{-1}) \\ &= \frac{1}{T^3 h_T^2} \sum_s a_s U_s + o_p((Th_T^{1/2})^{-1}) \quad =: \frac{1}{T^3 h_T^2} J_{2,T}^* + o_p((Th_T^{1/2})^{-1}), \end{aligned}$$

where $a_s = \sum_{t \neq s} \frac{\Omega_0(Z_t) I_t}{f(Z_t)^2} K_{st} \sum_{u \neq t, s} K_{ut} b_u$. From the independence of the observations and the conditional moment restriction, $E \left[(J_{2,T}^*)^2 \right] = \sum_s E \left[a_s^2 U_s^2 \right] = \sum_s E \left[a_s^2 V_0(Z_s) \right]$. To compute the expectation $E \left[a_s^2 V_0(Z_s) \right]$, we use

$$\begin{aligned} E \left[a_s^2 V_0(Z_s) \right] &= \sum_{t \neq s} E \left[\frac{\Omega_0(Z_t)^2 I_t}{f(Z_t)^4} V_0(Z_s) K_{st}^2 \left(\sum_{u \neq t, s} K_{ut} b_u \right)^2 \right] \\ &\quad + \sum_{t \neq s} \sum_{i \neq t, s} E \left[\frac{\Omega_0(Z_t) I_t}{f(Z_t)^2} \frac{\Omega_0(Z_i) I_i}{f(Z_i)^2} V_0(Z_s) K_{st} K_{si} \left(\sum_{u \neq t, s} K_{ut} b_u \right) \left(\sum_{m \neq i, s} K_{mi} b_m \right) \right], \end{aligned}$$

where the second term is the dominant one. Moreover, for $t \neq s \neq i \neq u \neq m$,

$$E \left[V_0(Z_s) K_{st} K_{si} K_{ut} K_{mi} b_u b_m \mid Z_t, Z_i \right] = O_p \left(h_T^3 V_0(Z_t) f(Z_t)^2 f(Z_i) K * K \left(\frac{Z_i - Z_t}{h_T} \right) b_t b_i \right).$$

Thus we get $E [a_s^2 V_0(Z_s)] = O (T^4 h_T^4 E [\Omega_0(Z_t) I_t b_t^2])$. We deduce

$$J_{21,T} = O_p \left(\frac{1}{\sqrt{T}} E [\Omega_0(Z_t) I_t [(A\mathcal{B}_T)(Z_t)]^2]^{1/2} \right) + o_p((Th_T^{1/2})^{-1}).$$

The second term $J_{22,T}$ can be analysed along the same lines as term $J_{13,T}$ in the proof of Lemma A.3, using $E [\eta_u | \mathcal{I}, W_s] = 0$, for $u \neq s$, and $E (\eta_u^2) = o(1)$. Hence $J_{22,T} = o_p((Th_T^{1/2})^{-1})$, and the conclusion follows. ■

Lemma A.5: $J_{3,T} = o_p((Th_T^{1/2})^{-1})$.

Proof: Write:

$$\begin{aligned} \hat{\psi}(z) &= \frac{\frac{1}{Th_T} \sum_n U_n K \left(\frac{Z_n - z}{h_T} \right)}{f(z)} + \frac{\frac{1}{Th_T} \sum_n G_{n,T} K \left(\frac{Z_n - z}{h_T} \right)}{f(z)} \\ &=: \frac{1}{T} \sum_n U_n \omega_n(z) + \frac{1}{T} \sum_n G_{n,T} \omega_n(z), \end{aligned}$$

where $G_{n,T} := \int [\varphi_0(X_n) - \varphi_0(X_n + uh_T)] K(u) du$. Then we have $\mathcal{V}_T(X_s) = \frac{1}{T} \sum_n U_n \Psi_{sn} + \frac{1}{T} \sum_n G_{n,T} \Psi_{sn}$, where $\Psi_{sn} := ((\lambda_T + A^* A)^{-1} A^* \omega_n)(X_s)$. We get

$$\begin{aligned} J_{3,T} &= \frac{1}{T^3} \sum_t \frac{\hat{\Omega}_t I_t}{\left(\sum_j K_{jt} \right)^2} \sum_n \sum_m U_n U_m \left(\sum_{s \neq t} \sum_{u \neq s,t} K_{st} K_{ut} \Psi_{sn} \Psi_{um} \right) \\ &\quad + \frac{1}{T^3} \sum_t \frac{\hat{\Omega}_t I_t}{\left(\sum_j K_{jt} \right)^2} \sum_n \sum_m G_{n,T} G_{m,T} \left(\sum_{s \neq t} \sum_{u \neq s,t} K_{st} K_{ut} \Psi_{sn} \Psi_{um} \right) \\ &\quad + 2 \frac{1}{T^3} \sum_t \frac{\hat{\Omega}_t I_t}{\left(\sum_j K_{jt} \right)^2} \sum_n \sum_m U_n G_{m,T} \left(\sum_{s \neq t} \sum_{u \neq s,t} K_{st} K_{ut} \Psi_{sn} \Psi_{um} \right) \\ &=: J_{31,T} + J_{32,T} + 2J_{33,T}. \end{aligned}$$

Let us first consider term $J_{31,T}$. Define $Q_{sn} := E[\Psi_{sn}|\mathcal{I}] = (A(\lambda_T + A^*A)^{-1}A^*\omega_n)(Z_s)$ and

$V_{sn} := \Psi_{sn} - Q_{sn}$. Then:

$$\begin{aligned}
J_{31,T} &= \frac{1}{T^3} \sum_t \frac{\hat{\Omega}_t I_t}{\left(\sum_j K_{jt}\right)^2} \sum_n \sum_m U_n U_m \left(\sum_{s \neq t} \sum_{u \neq s,t} K_{st} K_{ut} Q_{sn} Q_{um} \right) \\
&\quad + \frac{1}{T^3} \sum_t \frac{\hat{\Omega}_t I_t}{\left(\sum_j K_{jt}\right)^2} \sum_n \sum_m U_n U_m \left(\sum_{s \neq t} \sum_{u \neq s,t} K_{st} K_{ut} V_{sn} V_{um} \right) \\
&\quad + 2 \frac{1}{T^3} \sum_t \frac{\hat{\Omega}_t I_t}{\left(\sum_j K_{jt}\right)^2} \sum_n \sum_m U_n U_m \left(\sum_{s \neq t} \sum_{u \neq s,t} K_{st} K_{ut} Q_{sn} V_{um} \right) \\
&=: J_{311,T} + J_{312,T} + J_{313,T}. \tag{6}
\end{aligned}$$

We consider first term $J_{311,T}$. Using an argument similar to Lemmas A.1 and A.2 above,

Lemmas A.6 and A.7 in TK, and Lemma C.1 (iv) below, we have

$$\begin{aligned}
J_{311,T} &= \frac{1}{T^3} \sum_t \tilde{H}(Z_t)^{-1} I_t \sum_{n \neq t} \sum_{m \neq n,t} U_n U_m \left(\frac{1}{T^2 h_T^2} \sum_{s \neq t} \sum_{u \neq s,t} K_{st} K_{ut} Q_{sn} Q_{um} \right) + o_p((Th_T^{1/2})^{-1}) \\
&=: \frac{1}{T^3} J_{3,T}^* + o_p((Th_T^{1/2})^{-1}). \tag{7}
\end{aligned}$$

Term $J_{3,T}^*$ can be written as $J_{3,T}^* = \sum_n \sum_{m > n} \gamma_{nm} U_n U_m$, where

$$\gamma_{nm} := 2 \sum_{t \neq n,m} \tilde{H}(Z_t)^{-1} I_t \left(\frac{1}{T^2 h_T^2} \sum_{s \neq t} \sum_{u \neq s,t} K_{st} K_{ut} Q_{sn} Q_{um} \right).$$

By using that variables U_n and U_m are uncorrelated conditional on \mathcal{I} , we have

$$E[J_{3,T}^{*2}] = \sum_n \sum_{m > n} E[\gamma_{nm}^2 U_n^2 U_m^2] = \sum_n \sum_{m > n} E[\gamma_{nm}^2 V_0(Z_n) V_0(Z_m)].$$

To compute the expectation, we use an argument similar to Lemma A.8 in TK. To simplify

let $\Omega_0(z) = V_0(z)^{-1} = 1$. Then, $E[\gamma_{nm}^2] = O\left(\sum_{t=1, t \neq n,m}^T \sum_{i=1, i \neq n,m,t}^T R_{ti}\right)$, where

$$R_{ti} := E \left[I_t I_i \left(\frac{1}{T^2 h_T^2} \sum_{s \neq t} \sum_{u \neq s,t} K_{st} K_{ut} Q_{sn} Q_{um} \right) \left(\frac{1}{T^2 h_T^2} \sum_{p \neq i} \sum_{q \neq s,i} K_{pi} K_{qi} Q_{pn} Q_{qm} \right) \right].$$

Developing the sums, using $\frac{1}{h_T} E [K_{st} Q_{sn} | Z_t, Z_n] = O_p(f(Z_t) Q_{tn})$ for $s \neq t, n$, and the independence of observations, we get

$$R_{ti} = O(E [I_t I_i Q_{tn} Q_{tm} Q_{in} Q_{im}]) = O(E [I_t I_i E [Q_{tn} Q_{in} | Z_t, Z_i]^2]). \quad (8)$$

To compute expectations involving Q_{tn} , we use a development of $(\lambda_T + A^* A)^{-1} A^* \omega_n$ w.r.t. the basis of eigenfunctions ϕ_j of $A^* A$ to eigenvalues ν_j :

$$A(\lambda_T + A^* A)^{-1} A^* \omega_n = \sum_{j=1}^{\infty} \frac{1}{\lambda_T + \nu_j} \langle \phi_j, A^* \omega_n \rangle_H A \phi_j = \sum_{j=1}^{\infty} \frac{1}{\lambda_T + \nu_j} \langle A \phi_j, \omega_n \rangle_{L^2(F_Z)} A \phi_j.$$

Thus $Q_{tn} = \sum_{j=1}^{\infty} \frac{1}{\lambda_T + \nu_j} c_{nj} A \phi_j(Z_t)$ where

$$c_{nj} := \langle A \phi_j, \omega_n \rangle_{L^2(F_Z)} = \frac{1}{h_T} \int A \phi_j(z) K\left(\frac{Z_n - z}{h_T}\right) dz = \int A \phi_j(Z_n - h_T u) K(u) du.$$

Then

$$E [Q_{tn} Q_{in} | Z_t, Z_i] = \sum_{j=1}^{\infty} \sum_{l=1}^{\infty} \frac{1}{\lambda_T + \nu_j} \frac{1}{\lambda_T + \nu_l} E [c_{nj} c_{nl}] A \phi_j(Z_t) A \phi_l(Z_i). \quad (9)$$

From the orthogonality of the eigenfunctions, and the independence of the observations, we get $E [E [Q_{tn} Q_{in} | Z_t, Z_i]^2] = \sum_{j=1}^{\infty} \sum_{l=1}^{\infty} \frac{\nu_j}{(\lambda_T + \nu_j)^2} \frac{\nu_l}{(\lambda_T + \nu_l)^2} E [c_{nj} c_{nl}]^2$, for $t \neq i$. Moreover, from Assumptions B.4 (i)-(ii) and B.8 (ii) we have

$$\begin{aligned} E [c_{nj} c_{nl}] &= E [A \phi_j(Z_n) A \phi_l(Z_n)] + O(h_T^2) \left(E [A \phi_j(Z_n)^2]^{1/2} + E [A \phi_l(Z_n)^2]^{1/2} \right) + O(h_T^4) \\ &= \nu_j \delta_{jl} + O(h_T^2) (\sqrt{\nu_j} + \sqrt{\nu_l}) + O(h_T^4), \end{aligned} \quad (10)$$

uniformly in j, l , where δ_{jl} is the Kronecker delta. Thus we get

$$\begin{aligned} R_{ti} &= O \left(\sum_{j=1}^{\infty} \frac{\nu_j^4}{(\lambda_T + \nu_j)^4} + h_T^4 \sum_{j=1}^{\infty} \frac{\nu_j^2}{(\lambda_T + \nu_j)^2} \sum_{l=1}^{\infty} \frac{\nu_l}{(\lambda_T + \nu_l)^2} + h_T^8 \left(\sum_{j=1}^{\infty} \frac{\nu_j}{(\lambda_T + \nu_j)^2} \right)^2 \right) \\ &=: O(S(\lambda_T)). \end{aligned}$$

Thus, $E [J_{3,T}^{*2}] = O(T(T-1)(T-2)(T-3)S(\lambda_T))$, which implies $Th_T^{1/2}J_{311,T} = O_p(\sqrt{h_T S(\lambda_T)}) + o_p(1)$. Using that $\sum_{j=1}^{\infty} \frac{\nu_j}{(\lambda_T + \nu_j)^2} = O\left(\frac{1}{\lambda_T}\right)$ and $\sum_{j=1}^{\infty} \frac{\nu_j^4}{(\lambda_T + \nu_j)^4} \leq \sum_{j=1}^{\infty} \frac{\nu_j^2}{(\lambda_T + \nu_j)^2} \leq \sum_{j=1}^{\infty} \frac{\nu_j}{\lambda_T + \nu_j} = O(\log(1/\lambda_T))$ under Assumption B.8 (see GS, proof of Lemma A.6), we get $S(\lambda_T) = O(\log(1/\lambda_T)) + O\left(h_T^4 \frac{1}{\lambda_T} \log(1/\lambda_T)\right) + O\left(h_T^8 \frac{1}{\lambda_T^2}\right)$. Then, $S(\lambda_T) = O(\log(1/\lambda_T))$ follows from $\lambda_T = cT^{-\gamma}$ with $\gamma < 4\bar{\eta}$ (Assumption B.10), and we get $Th_T^{1/2}J_{311,T} = o_p(1)$.

Let us now consider $J_{312,T}$ in (6). Using an argument similar to Lemmas A.1 and A.2 above, Lemmas A.6 and A.7 in TK, and Lemma C.1 (iv) below, we have

$$\begin{aligned} J_{312,T} &= \frac{1}{T^5 h_T^2} \sum_t \tilde{H}(Z_t)^{-1} I_t \sum_{n \neq t} \sum_{m \neq n, t} \sum_{s \neq t} \sum_{u \neq s, t} K_{st} K_{ut} U_n U_m V_{sn} V_{um} + o_p((Th_T^{1/2})^{-1}) \\ &= \frac{1}{T^4} \sum_n \sum_{m \neq n} \sum_s \sum_{u \neq s} \chi_{nmsu} U_n U_m V_{sn} V_{um} + o_p((Th_T^{1/2})^{-1}) =: J_{312,T}^* + o_p((Th_T^{1/2})^{-1}), \end{aligned}$$

where $\chi_{nmsu} := \frac{1}{Th_T^2} \sum_{t \neq n, m, s, u} \tilde{H}(Z_t)^{-1} I_t K_{st} K_{ut}$. Using that $E[U_n | \mathcal{I}, W_m] = 0$ for $m \neq n$, $E[V_{sn} | \mathcal{I}, W_u] = 0$ for $u \neq s$, and developing the expressions of the conditional variances, we deduce that $E[(J_{312,T}^*)^2] = O(1/(T^4 h_T \lambda_T^2))$. From $\lambda_T = cT^{-\gamma}$, $\gamma < 1$ (Assumption B.10), it follows $J_{312,T} = o_p((Th_T^{1/2})^{-1})$. Similar arguments apply for $J_{313,T}$, and from (6) we get $J_{31,T} = o_p((Th_T^{1/2})^{-1})$.

Let us now consider $J_{32,T}$. Similarly as in (6) and (7), we have $J_{32,T} = \frac{1}{T^3} J_{3,T}^{**} + o_p((Th_T^{1/2})^{-1})$, where $J_{3,T}^{**} = \sum_n \sum_{m \neq n} \gamma_{nm} G_{n,T} G_{m,T}$. From the above arguments we have $\gamma_{nm} = O_p\left(T\sqrt{S(\lambda_T)}\right)$ uniformly in n, m . Moreover, from Assumption B.9 (i), $G_{n,T} = O_p(h_T^2)$ uniformly in n . Thus, $J_{32,T} = O_p(h_T^4 \sqrt{S(\lambda_T)}) + o_p((Th_T^{1/2})^{-1})$. Since $S(\lambda_T) = O(\log(1/\lambda_T))$ (see above), $J_{32,T} = o_p((Th_T^{1/2})^{-1})$ follows from Assumptions B.7 and B.10.

Similar arguments apply to $J_{33,T}$, and the proof is concluded. ■

Lemma A.6: $J_{4,T} = o_p \left(\frac{1}{\sqrt{Th_T^{1/2}}} E \left[\Omega_0(Z_t)^2 I_t [(A\mathcal{B}_T)(Z_t)]^2 \right]^{1/2} \right) + o_p((Th_T^{1/2})^{-1})$.

Proof: Using the notation in the proof of Lemma A.3, we have

$$\begin{aligned} J_{4,T} &= -\frac{1}{T} \sum_t \frac{\hat{\Omega}_t I_t}{\left(\sum_j K_{jt} \right)^2} \sum_{s \neq t} \sum_{u \neq t, s} K_{st} K_{ut} b_s \mathcal{V}_T(X_u) \\ &\quad + \frac{1}{T} \sum_t \frac{\hat{\Omega}_t I_t}{\left(\sum_j K_{jt} \right)^2} \sum_{s \neq t} \sum_{u \neq t, s} K_{st} K_{ut} (U_s - \eta_s) \mathcal{V}_T(X_u) \\ &=: -J_{41,T} + J_{42,T}. \end{aligned}$$

Let us first consider $J_{41,T}$. Similar arguments as in the proof of Lemma A.5 show that

$$\begin{aligned} J_{41,T} &= \frac{1}{T^2} \sum_t \tilde{H}(Z_t)^{-1} I_t \sum_{n \neq t} U_n \left(\frac{1}{T^2 h_T^2} \sum_{s \neq t} \sum_{u \neq t, s} K_{st} K_{ut} b_s Q_{un} \right) \\ &\quad + \frac{1}{T^2} \sum_t \tilde{H}(Z_t)^{-1} I_t \sum_{n \neq t} G_{n,T} \left(\frac{1}{T^2 h_T^2} \sum_{s \neq t} \sum_{u \neq t, s} K_{st} K_{ut} b_s Q_{un} \right) + o_p((Th_T^{1/2})^{-1}) \\ &=: \frac{1}{T^2} J_{41,T}^* + \frac{1}{T^2} J_{41,T}^{**} + o_p((Th_T^{1/2})^{-1}). \end{aligned}$$

Furthermore, $J_{41,T}^* = \sum_n a_n U_n$ where

$$a_n = \sum_{t \neq n} \tilde{H}(Z_t)^{-1} I_t \left(\frac{1}{T^2 h_T^2} \sum_{s \neq t} \sum_{u \neq t, s} K_{st} K_{ut} b_s Q_{un} \right).$$

We have $E \left[(J_{41,T}^*)^2 \right] = \sum_n E [a_n^2 U_n^2] = \sum_n E [a_n^2 V_0(Z_n)]$. To simplify, let $\Omega_0(z) = V_0(z)^{-1} =$

1. Using an argument similar as for the derivation of (8), $E [a_n^2]$ is asymptotically equivalent to $\sum_{t \neq n} \sum_{i \neq t, n} E [I_t I_i b_t b_i E [Q_{nt} Q_{ni} | Z_t, Z_i]]$. Using (9), (10) and Cauchy-Schwarz inequality, for

$t \neq i$ we get

$$\begin{aligned} & E [I_t I_i b_t b_i E [Q_{nt} Q_{ni} | Z_t, Z_i]] \\ & \leq \left\{ \sum_{j=1}^{\infty} \frac{\nu_j^2}{(\lambda_T + \nu_j)^2} + O(h_T^2) \sum_{j=1}^{\infty} \frac{\nu_j}{\lambda_T + \nu_j} \sum_{l=1}^{\infty} \frac{\sqrt{\nu_l}}{\lambda_T + \nu_l} + O(h_T^4) \left(\sum_{j=1}^{\infty} \frac{\sqrt{\nu_j}}{\lambda_T + \nu_j} \right)^2 \right\} E [I_t b_t^2] \\ =: & S_1(\lambda_T) E [I_t b_t^2]. \end{aligned}$$

$$\text{Thus, } E \left[(J_{41,T}^*)^2 \right] = O(T^3 S_1(\lambda_T) E [I_t b_t^2]) \text{ and } \frac{1}{T^2} J_{41,T}^* = O_p \left(\frac{\sqrt{h_T^{1/2} S_1(\lambda_T)}}{\sqrt{T h_T^{1/2}}} E [I_t b_t^2]^{1/2} \right).$$

Similarly, writing $J_{41,T}^{**} = \sum_n a_n G_{n,T}$ and using $a_n = O_p \left(T \sqrt{S_1(\lambda_T) E [I_t b_t^2]} \right)$, $G_{n,T} = O_p(h_T^2)$, uniformly in n , we get $\frac{1}{T^2} J_{41,T}^{**} = O_p \left(h_T^2 \sqrt{S_1(\lambda_T)} E [I_t b_t^2]^{1/2} \right)$. Now, using that $\sum_{l=1}^{\infty} \frac{\sqrt{\nu_l}}{\lambda_T + \nu_l} \leq \left(\sum_{l=1}^{\infty} \frac{\nu_l l^2}{(\lambda_T + \nu_l)^2} \right)^{1/2} \left(\sum_{l=1}^{\infty} \frac{1}{l^2} \right)^{1/2} = O \left(\frac{1}{\lambda_T^{1/2}} \log(1/\lambda_T) \right)$ under Assumption B.8 (see Lemma A.6 is GS) we get $S_1(\lambda_T) = O \left((\log(1/\lambda_T))^2 \right)$ from Assumption B.10. Thus, $h_T^2 \sqrt{S_1(\lambda_T)} = o \left(\frac{1}{\sqrt{T h_T^{1/2}}} \right)$ from Assumption B.7, and $J_{41,T} = o_p \left(\frac{1}{\sqrt{T h_T^{1/2}}} E [I_t b_t^2]^{1/2} \right) + o_p \left((T h_T^{1/2})^{-1} \right)$.

Let us now consider $J_{42,T}$. By similar arguments as above we have

$$\begin{aligned} J_{42,T} &= \frac{1}{T^3 h_T} \sum_t \tilde{H}(Z_t)^{-1} I_t \sum_{s \neq t} \sum_{n \neq s, t} (U_s - \eta_s) U_n \left(\frac{1}{T h_T} \sum_{u \neq t, s} K_{st} K_{ut} Q_{un} \right) + o_p \left((T h_T^{1/2})^{-1} \right) \\ =: & \frac{1}{T^3 h_T} J_{42,T}^* + o_p \left((T h_T^{1/2})^{-1} \right), \end{aligned}$$

$$\text{where } J_{42,T}^* = \sum_s \sum_{n \neq s} d_{ns} (U_s - \eta_s) U_n \text{ and } d_{ns} := \sum_{t \neq s, n} \tilde{H}(Z_t)^{-1} I_t \left(\frac{1}{T h_T} \sum_{u \neq t, s} K_{st} K_{ut} Q_{un} \right).$$

Using that $E [U_s | \mathcal{I}, W_u] = E [\eta_s | \mathcal{I}, W_u] = 0$ for $s \neq u$, we get

$$E \left[(J_{42,T}^*)^2 \right] = \sum_s \sum_{n \neq s} E [d_{ns}^2 \Psi_1(Z_s)] + \sum_s \sum_{n \neq s} E [d_{ns} d_{sn} \Psi_2(Z_s) \Psi_2(Z_n)],$$

where $\Psi_1(Z_s) := E [(U_s - \eta_s)^2 | Z_s]$, $\Psi_2(Z_s) := E [(U_s - \eta_s) U_s | Z_s]$. Then, $E [d_{ns}^2 \Psi_1(Z_s)]$

is asymptotically equivalent to $\sum_{t \neq s, n} \sum_{i \neq t, s, n} E [I_t I_i \Psi_1(Z_s) K_{st} K_{si} Q_{nt} Q_{ni}]$. Using (9), (10),

$E[\Psi_1(Z_s)K_{st}K_{si}|Z_i, Z_t] = O_p\left(h_T K * K\left(\frac{Z_i - Z_t}{h_T}\right) f(Z_t)\Psi_1(Z_t)\right)$ and Cauchy-Schwarz inequality, we get $E[d_{ns}^2\Psi_1(Z_s)] = O(T^2h_T^2S_1(\lambda_T))$. A similar bound holds for $E[d_{ns}d_{sn}\Psi_2(Z_s)\Psi_2(Z_n)]$. Then, $J_{42,T} = o_p((Th_T^{1/2})^{-1})$ using the same arguments as for $J_{41,T}$. ■

Lemma A.7: $J_{5,T} = o_p((Th_T^{1/2})^{-1})$.

Proof: By applying twice the Cauchy-Schwarz inequality, we have

$$|J_{5,T}| \leq \max_{t \in T^*} \left| \frac{(Th_T)^2 \hat{\Omega}_t}{\left(\sum_j K_{jt}\right)^2} \right| \frac{1}{T} \sum_t |\mathcal{R}_T(X_t)|^2 I_t \frac{1}{T^2 h_T^2} \sum_t \left(\sum_{s \neq t} \sum_{u \neq s,t} K_{st}^2 K_{ut}^2 I_t \right)^{1/2}.$$

Using $\max_{t \in T^*} \left| \frac{(Th_T)^2 \hat{\Omega}_t}{\left(\sum_j K_{jt}\right)^2} \right| = O_p(1)$, $\sum_{s \neq t} \sum_{u \neq s,t} E[K_{st}^2 K_{ut}^2 I_t] = O(T^2 h_T^2)$, uniformly in t , and Assumption B.6 (iv), the conclusion follows. ■

Lemma A.8: $J_{6,T} = o_p((Th_T^{1/2})^{-1})$.

Proof: Write $J_{6,T} = \frac{1}{T} \sum_t \frac{\hat{\Omega}_t I_t}{\left(\sum_j K_{jt}\right)^2} \sum_{s \neq t} K_{st} \mathcal{R}_T(X_s) \Phi_{t,s}$, where we set $\Phi_{t,s} := \sum_{u \neq t,s} K_{ut} (U_u - \mathcal{B}_T(X_u) - \mathcal{V}_T(X_u))$. By applying twice the Cauchy-Schwarz inequality, we get

$$|J_{6,T}| \leq \max_{t \in T^*} \left| \frac{(Th_T)^2 \hat{\Omega}_t}{\left(\sum_j K_{jt}\right)^2} \right| \frac{1}{T^3 h_T^2} \left(\sum_t \sum_{s \neq t} K_{st}^2 \mathcal{R}_T(X_s)^2 I_t \right)^{1/2} \left(\sum_t \sum_{s \neq t} \Phi_{t,s}^2 I_t \right)^{1/2}.$$

Using $\max_{t \in T^*} \left| \frac{(Th_T)^2 \hat{\Omega}_t}{\left(\sum_j K_{jt}\right)^2} \right| = O_p(1)$ and Assumption B.6 (v), the conclusion follows if we can show that $\Phi_{t,s} I_t = O_p\left(\sqrt{Th_T}\right)$, uniformly in $s \neq t$. By using the notation in the proof of

Lemma A.3 we have

$$\Phi_{t,s} := \sum_{u \neq t,s} K_{ut} (U_u - \eta_u) - \sum_{u \neq t,s} K_{ut} b_u - \sum_{u \neq t,s} K_{ut} \mathcal{V}_T(X_u) =: \Phi_{1,ts} - \Phi_{2,ts} - \Phi_{3,ts}.$$

Since variables $U_u - \eta_u$ are uncorrelated conditionally on \mathcal{I} , $E[\Phi_{1,ts}^2] = O(Th_T)$. Furthermore, $E[\Phi_{2,ts}^2 I_t] = O(T^2 h_T^2 E[I_t b_t^2]) = o(Th_T)$ by Assumptions B.5, B.9 (ii) and B.10.

Finally, as in the proof of Lemma A.5 we have

$$\begin{aligned}\Phi_{3,ts} &= h_T \sum_n U_n \left(\frac{1}{Th_T} \sum_{u \neq t,s} K_{ut} Q_{un} \right) + \frac{1}{T} \sum_n \sum_{u \neq t,s} K_{ut} U_n V_{un} \\ &\quad + h_T \sum_n G_{n,T} \left(\frac{1}{Th_T} \sum_{u \neq t,s} K_{ut} \Psi_{un} \right) =: \Phi_{31,ts} + \Phi_{32,ts} + \Phi_{33,ts}.\end{aligned}$$

From (9) and (10), $E \left[\left(\frac{1}{Th_T} \sum_{u \neq t,s} K_{ut} Q_{un} \right)^2 \right]$ and $E \left[\left(\frac{1}{Th_T} \sum_{u \neq t,s} K_{ut} \Psi_{un} \right)^2 \right]$ are asymptotically equivalent to

$$E[Q_{tn}^2] = \sum_{j=1}^{\infty} \frac{\nu_j^2}{(\lambda_T + \nu_j)^2} + O(h_T^2) \sum_{j=1}^{\infty} \frac{\nu_j^{3/2}}{(\lambda_T + \nu_j)^2} + O(h_T^4) \sum_{j=1}^{\infty} \frac{\nu_j}{(\lambda_T + \nu_j)^2} =: S_2(\lambda_T).$$

Using Cauchy-Schwarz inequality, Assumptions B.8, B.10 and similar arguments as in the proof of Lemma A.5 we get $S_2(\lambda_T) = O(\log(1/\lambda_T))$. Thus, $E[\Phi_{31,ts}^2] = O(Th_T^2 \log(1/\lambda_T))$ and $\Phi_{33,ts} = O_p\left(Th_T^3 \sqrt{\log(1/\lambda_T)}\right)$. Moreover, $E[\Phi_{32,ts}^2] = O(h_T/\lambda_T)$. From Assumptions B.7 and B.10, the conclusion follows. ■

Lemma C.1:

- (i) $\sup_{z \in S_*} \left| \hat{V}(z) - V_0(z) \right| = O_p \left(\sqrt{\frac{\log T}{Th_T}} + h_T^2 \right) + o_p(T^{-\varepsilon/2+1/m});$
- (ii) $\sup_{z \in S_*} \left| \hat{V}(z)^{-1} - V_0(z)^{-1} \right| = O_p \left(\sqrt{\frac{\log T}{Th_T}} + h_T^2 \right) + o_p(T^{-\varepsilon/2+1/m});$
- (iii) $\sup_{z \in S_*} \left| \hat{H}(z)^{-1} - H_0(z)^{-1} \right| = O_p \left(\sqrt{\frac{\log T}{Th_T}} + h_T^2 \right) + o_p(T^{-\varepsilon/2+1/m});$
- (iv) $\sup_{z \in S_*} \left| \hat{H}(z)^{-1} - \tilde{H}(z)^{-1} \right| = O_p \left(\sqrt{\frac{\log T}{Th_T}} \right) + o_p(T^{-\varepsilon/2+1/m}).$

Proof: From Cauchy-Schwarz inequality,

$$\left| \hat{V}(Z_t) - V_0(Z_t) \right| \leq \left| \sum_j w_{tj} U_j^2 - V_0(Z_t) \right| + 2A(Z_t) + B(Z_t),$$

where $A(Z_t) = \left(\sum_j w_{tj} U_j^2 \right)^{1/2} \left(\sum_j w_{tj} |\Delta \bar{\varphi}(X_j)|^2 \right)^{1/2}$ and $B(Z_t) = \sum_j w_{tj} |\Delta \bar{\varphi}(X_j)|^2$. As

in the proof of Lemma C.2 in TK and using $h_T = \bar{c}T^{-\bar{\eta}}$ with $\bar{\eta} < 1 - 4/m$,

$\sup_{Z_t \in S_*} \left| \sum_j w_{tj} U_j^2 - V_0(Z_t) \right| = O_p \left(\sqrt{\frac{\log T}{Th_T}} + h_T^2 \right)$. Further, from Lemma C.6 of TK and

Assumption B.2, $\sup_{Z_t \in S_*} \sum_j w_{tj} U_j^2 = o_p(T^{2/m})$. Then, (i) follows from Assumption B.6 (iii)

and uniform convergence of $\hat{f}(z)$ over S_* . Points (ii) and (iii) follow from (i), Assumption

B.5 and uniform convergence of $\hat{f}(z)$ over S_* . Finally, (iv) follows from Lemma C.3 in TK

and similar arguments as in (i). ■

References

- Abramowitz, M. and I. Stegun (1970): *Handbook of Mathematical Functions*, Dover Publications, New York.
- Ai, C. and X. Chen (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions", *Econometrica*, 71, 1795-1843.
- Arcones, M. and E. Gine (1992): "On the Bootstrap of U and V Statistics", *Annals of Statistics*, 20, 655-674.
- Banks, J., Blundell, R. and A. Lewbel (1997): "Quadratic Engel Curves and Consumer Demand", *Review of Economics and Statistics*, 79, 527-539.
- Bickel, P., Gotze, F. and W. van Zwet (1997): "Resampling Fewer than n Observations: Gains, Losses, and Remedies for Losses", *Statistica Sinica*, 7, 1-31.
- Brown, B. and W. Newey (2002): "Generalized Method of Moments, Efficient Bootstrapping, and Improved Inference", *Journal of Business and Economic Statistics*, 20, 507-517.
- Carrasco, M. and J.-P. Florens (2000): "Generalization of GMM to a Continuum of Moment Conditions", *Econometric Theory*, 16, 797-834.
- Carrasco, M., Florens, J.-P. and E. Renault (2006): "Linear Inverse Problems in Structural Econometrics: Estimation Based on Spectral Decomposition and Regularization", forthcoming in the *Handbook of Econometrics*.
- Darolles, S., Florens, J.-P. and E. Renault (2003): "Nonparametric Instrumental Regression", Working Paper.
- De Jong, P. (1987): "A Central Limit Theorem for Generalized Quadratic Forms", *Probab. Theory Related Fields*, 75, 261-277.
- Devroye, L. and T. Wagner (1980): "Distribution-Free Consistency Results in Nonparametric Discrimination and Regression Function Estimation", *Annals of Statistics*, 8, 231-239.
- Donald, S., Imbens, G. and W. Newey (2003): "Empirical Likelihood Estimation and Consistent Tests with Conditional Moment Restrictions", *Journal of Econometrics*, 117, 55-93.
- Gagliardini, P. and O. Scaillet (2006): "Tikhonov Regularisation for Functional Minimum Distance Estimators", Working Paper.

- Haerdle, W. and E. Mammen (1993): "Comparing Nonparametric versus Parametric Regression Fits", *Annals of Statistics*, 21, 1926-1947.
- Hall, P. (1984): "Central Limit Theorem for Integrated Square Error of Multivariate Nonparametric Density Estimators", *Journal of Multivariate Analysis*, 14, 1-16.
- Hall, P. and J. Horowitz (1996): "Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators", *Econometrica*, 64, 891-916.
- Hall, P. and J. Horowitz (2005): "Nonparametric Methods for Inference in the Presence of Instrumental Variables", *Annals of Statistics*, 33, 2904-2929.
- Hansen, B. (1996): "Inference when a Nuisance Parameter is not Identified under the Null Hypothesis", *Econometrica*, 64, 413-430.
- Hansen, L. (1982): "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica*, 50, 1029-1054.
- Horowitz, J. (2006): "Testing a Parametric Model Against a Nonparametric Alternative with Identification Through Instrumental Variables", *Econometrica*, 74, 521-538.
- Kress, R. (1999): *Linear Integral Equations*, Springer, New York.
- Newey, W. (1994): "Kernel Estimation of Partial Means and a General Variance Estimator", *Econometric Theory*, 10, 233-253.
- Newey, W. and J. Powell (2003): "Instrumental Variable Estimation of Nonparametric Models", *Econometrica*, 71, 1565-1578.
- Politis, D., Romano, J. and M. Wolf (1999): *Subsampling*, Springer-Verlag, New-York.
- Sargan, J. (1958): "The Estimation of Economic Relationships Using Instrumental Variables", *Econometrica*, 26, 393-415.
- Sargan, J. (1959): "The Estimation of Relationships with Autocorrelated Residuals By the Use of Instrumental Variables", *Journal of the Royal Statistical Society, Ser. B*, 21, 91-105.
- Silverman, B. (1986): *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.
- Tripathi, G. and Y. Kitamura (2003): "Testing Conditional Moment Restrictions", *Annals of Statistics*, 31, 2059-2095.