A SPECIFICATION TEST FOR NONPARAMETRIC INSTRUMENTAL VARIABLE REGRESSION

P. Gagliardini* and O. Scaillet†

First version: April 2007 ‡

This version: October 2008

*University of Lugano and Swiss Finance Institute.

†HEC Université de Genève and Swiss Finance Institute.

‡We thank J.-P. Florens and the two referees for constructive criticism and careful reading, which led to improving and correcting an earlier version of the paper. We also thank the participants at ESEM 2007 for helpful comments. 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. First we study the notion of correct specification, misspecification and overidentification in this ill-posed inverse problem setting. Second we study a test statistic based on the empirical minimum distance criterion corresponding to the conditional moment restriction evaluated with a Tikhonov Regularized estimator of the functional parameter. The test statistic admits an asymptotic normal distribution under the null hypothesis, and the test is consistent under global alternatives. A bootstrap procedure is available to get simulation based critical values. Finally, we explore the finite sample behavior with Monte Carlo experiments, and provide an empirical illustration 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

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

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 $\mathcal{X} \subset \mathbb{R}$. 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. Equation (1) is a linear integral equation of the first kind in φ_0, and 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. Below we often refer to the handbook chapter by Carrasco, Florens and Renault (2006, CFR) for background on ill-posed problems in econometrics.

Section 2 outlines the specification testing problem. Section 2.1 describes the null hypothesis of correct specification, the alternative hypothesis of misspecification, and the concept of overidentification in a nonparametric IV setting. We clarify these notions with two Gaussian examples in Section 2.2. In Section 2.3 we introduce appropriate regularity spaces to derive the asymptotics of the test statistic under the null and the alternative hypotheses. Section 3 describes the testing procedure and its asymptotic properties. We give the test statistic in Section 3.1, establish its asymptotic normality under the null hypothesis in Section 3.2, and show consistency of the test under global alternatives in Section 3.3. Results are first given under a known weighting function in the construction of the test statistic before discussing the extension to an estimated weighting function in Section 3.4. We further explain how to implement a bootstrap procedure to get simulation based critical values in Section 3.5. Section 4 explores the finite sample behavior with Monte Carlo experiments. Section 5 provides an empirical illustration for an estimated Engel curve. In Appendices 1-5 we gather the list of regularity conditions and the technical arguments justifying the asymptotic distribution under the null hypothesis and consistency of the test under the alternative hypothesis. All omitted proofs of technical Lemmas are collected in a Technical Report, which is available online at our web pages.

2 The specification testing problem

2.1 The null hypothesis and overidentification

Let $L^2(\mathcal{X})$ and $L^2(\mathcal{Z})$ be the L^2-spaces w.r.t. the scalar products $\langle \varphi_1, \varphi_2 \rangle_{L^2(\mathcal{X})} = \int_{\mathcal{X}} \varphi_1(x)\varphi_2(x)$ $\Pi_{\mathcal{X}}(dx)$ and $\langle \psi_1, \psi_2 \rangle_{L^2(\mathcal{Z})} = \int_{\mathcal{Z}} \psi_1(z)\psi_2(z)\Pi_{\mathcal{Z}}(dz)$, respectively, where $\Pi_{\mathcal{X}}$ and $\Pi_{\mathcal{Z}}$ are given measures on the supports $\mathcal{X} \subset \mathbb{R}$ of X, and $\mathcal{Z} \subset \mathbb{R}$ of Z. The parameter set is the Sobolev space $H^2(\mathcal{X})$, which is the completion of the linear space $\{\varphi \in L^2(\mathcal{X}) \mid \nabla\varphi \in L^2(\mathcal{X})\}$ w.r.t. the scalar product $\langle \varphi_1, \varphi_2 \rangle_H := \langle \varphi_1, \varphi_2 \rangle_{L^2(\mathcal{X})} + \langle \nabla\varphi_1, \nabla\varphi_2 \rangle_{L^2(\mathcal{X})}$. The Sobolev space $H^2(\mathcal{X})$ is an Hilbert space w.r.t. the scalar product $\langle ., . \rangle_H$, and the corresponding Sobolev norm is denoted by $\|\varphi\|_H = \langle \varphi, \varphi \rangle_H^{1/2}$.

The conditional moment restriction (1) corresponds to the linear integral equation

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

for $\varphi_0 \in H^2(\mathcal{X})$, with $A_F\varphi(z) = \int f(w|z)\,\varphi(x)dw$, $r_F(z) = \int yf(w|z)dy$, and f the pdf of W given Z.

Assumption 1: *$F_0 \in \mathcal{F}$, where $\mathcal{F}$ denotes the set of conditional distributions F of W given Z such that $r_F \in L^2(\mathcal{Z})$ and A_F is a compact linear operator from $H^2(\mathcal{X})$ into $L^2(\mathcal{Z})$.*

The assumption of compactness of the operator A_{F_0} implies that (2) is an integral equation of the first kind which yields an ill-posed inverse problem (see CFR, Sections 3 and 5.5).

The model $\mathcal{M} \subset \mathcal{F}$ is the subset of distributions F such that equation $A_F\varphi = r_F$ admits a solution, that is

$$\mathcal{M} = \{F \in \mathcal{F} : r_F \in \text{Range}\,(A_F)\}, \tag{3}$$

where Range(A_F) denotes the range of operator A_F on $H^2(\mathcal{X})$. The null hypothesis of correct specification is

$$H_0 : F_0 \in \mathcal{M}, \tag{4}$$

while the alternative hypothesis is $H_1 : F_0 \in \bar{\mathcal{M}} := \mathcal{F} \setminus \mathcal{M}$. The definition (3) clarifies that the null hypothesis depends on the function space on which A_F operates, and thus on the postulated smoothness of the functional regression parameter.

Identification is a maintained hypothesis.

Assumption 2: $F_0 \in \{F \in \mathcal{F} : \ker(A_F) = \{0\}\}$.

Assumption 2 ensures that, under the null hypothesis H_0, the solution φ_0 of (2) is unique, since the condition $\ker(A_F) = \{0\}$ on the null space of operator A_F is equivalent to the injectivity of A_F (see CFR, Section 3.1). 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 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 unconditional 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 $\mathcal{M} \subsetneqq \mathcal{F}$. Otherwise, if $\mathcal{M} = \mathcal{F}$ a unique solution of $A_F\varphi = r_F$ 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 (4) does not imply a constraint on F_0. This is the analogue of the just identified case in the standard parametric GMM setting.

It turns out that the nonparametric instrumental variable regression model is overidentified by construction: $\mathcal{M}$ cannot be equal to $\mathcal{F}$. More precisely, $\mathcal{M}$ is a strict subset of $\mathcal{F}$ that can be characterized explicitly by Picard Theorem. Let us introduce the singular system $\left\{\phi_j, \psi_j, \omega_j; j = 1, 2, \ldots\right\}$ of operator A_F, [1] defined by $A_F\phi_j = \omega_j\psi_j$ and $A_F^*\psi_j = \omega_j\phi_j$, where $\phi_j \in H^2(\mathcal{X})$, $\psi_j \in L^2(\mathcal{Z})$, $\omega_j \geq 0$, and A_F^* is the adjoint operator of A_F w.r.t. the scalar products $\langle ., . \rangle_H$ and $\langle ., . \rangle_{L^2(\mathcal{Z})}$ (e.g., Kress (1999), Theorem 15.16, and CFR, Section 2.3). Functions ϕ_j are an orthonormal basis of eigenfunctions of the operator $A_F^*A_F$ in $H^2(\mathcal{X})$ to the eigenvalues $\nu_j = \omega_j^2$. Functions ψ_j are an orthonormal basis of $\ker\left(A_F^*\right)^\perp$, that is the linear subspace of $L^2(\mathcal{Z})$ orthogonal to $\ker\left(A_F^*\right)$ w.r.t. $\langle ., . \rangle_{L^2(\mathcal{Z})}$. Then, Picard Theorem (e.g., Kress (1999), Theorem 15.18) states that:

$$\mathcal{M} = \left\{ F \in \mathcal{F} : r_F \in \ker\left(A_F^*\right)^\perp \text{ and } \sum_{j=1}^{\infty} \frac{\left\langle r_F, \psi_j \right\rangle_{L^2(\mathcal{Z})}^2}{\nu_j} < \infty \right\}. \tag{5}$$

For the compact operator A_F we have $\ker\left(A_F^*\right)^\perp = \overline{\text{Range}\ (A_F)}$, that is the closure of Range (A_F) in $L^2(\mathcal{Z})$ (e.g., Kress (1999), Theorem 15.8). Thus, the set $\mathcal{M}$ consists of the distributions F such that function $r_F \in \overline{\text{Range}\ (A_F)}$ and such that the basis coefficients $\left\langle r_F, \psi_j \right\rangle_{L^2(\mathcal{Z})}$ weighted by the inverse eigenvalues ν_j satisfy a summability condition. This summability condition alone is not an equivalent characterization of (3) even under Assumption 2. The

[1] To simplify the notation, we omit the index F in ϕ_j, ψ_j and ω_j.

set $\bar{\mathcal{M}}$ characterizing the alternative hypothesis of misspecification can be decomposed as:

$$\begin{aligned} \bar{\mathcal{M}} &= \left\{ F \in \mathcal{F} : r_F \in \ker\left(A_F^*\right)^{\perp} \text{ and } \sum_{j=1}^{\infty} \frac{\left\langle r_F, \psi_j \right\rangle_{L^2(\mathcal{Z})}^2}{\nu_j} = \infty \right\} \\ & \quad \cup \left\{ F \in \mathcal{F} : r_F \notin \ker\left(A_F^*\right)^{\perp} \right\} \quad =: \quad \bar{\mathcal{M}}_c \cup \bar{\mathcal{M}}_s. \end{aligned}$$

For $F \in \bar{\mathcal{M}}_c$ we have $r_F \in \overline{\text{Range}\ (A_F)}$, while for $F \in \bar{\mathcal{M}}_s$ we have $r_F \notin \overline{\text{Range}\ (A_F)}$. Thus, the alternatives in $\bar{\mathcal{M}}_c$ are "arbitrarily close" to correct specification, while the alternatives in $\bar{\mathcal{M}}_s$ can be "separated" from the set of correctly specified models. We call the former *close* misspecifications, and the latter *separated* misspecifications.

2.2 Examples

In light of definition (3), the notion of misspecification in a nonparametric IV regression setting is intimately linked with the properties of Range (A_F). To illustrate this point, we develop two simple examples based on the regression model $Y = m(X) + U$, where $E\left[U|Z = z\right] = \rho(z)$ for $m \in H^2\left(\mathcal{X}\right)$ and $\rho \in L^2(\mathcal{Z})$. Then, r_F is such that $r_F = A_F m + \rho$, and $F \in \mathcal{M}$ if and only if

$$\rho \in \text{Range}\left(A_F\right). \tag{6}$$

Example 1: *Let (X, U, Z) be jointly normal with zero means, unit variances, $Cov\,(X, Z) = \rho_{XZ} \neq 0$, and $Cov\,(U, Z) = \rho_{UZ}$ under F. Then, $\rho(z) = \rho_{UZ} z$ and we have $\rho = A_F \Delta\varphi$ where $\Delta\varphi(x) = \dfrac{\rho_{UZ}}{\rho_{XZ}} x$. Thus, $\rho \in \text{Range}\left(A_F\right)$.*

In Example 1 the moment restriction is correctly specified, even when the innovation and the instrument are correlated ($\rho_{UZ} \neq 0$). This exemplifies a 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 and only if there exists θ_0 such that $\varphi(x,\theta_0)=m(x)+\dfrac{\rho_{UZ}}{\rho_{XZ}}x$, $\forall x\in\mathcal{X}$.

For the second example, assume that operator A_F is such that the functions in its range are continuous, i.e., Range $(A_F)\subset C(\mathcal{Z})$. Then, for a discontinuous function ρ we have $\rho\notin \text{Range}\,(A_F)$ and $F\in\mathcal{F}\backslash\mathcal{M}$.

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* Range $(A_F)\subset C(\mathcal{Z})$. *Thus,* $\rho\notin\text{Range}\,(A_F)$.

A similar argument is possible when Range $(A_F)\subset C^1(\mathcal{Z})$, 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.

In light of definition (5), we can also revisit Examples 1 and 2 through the characterization of the set $\mathcal{M}$ in terms of the singular system of operator A_F. Let us derive the singular value decomposition when the distribution F is such that (X,Z) is jointly normal, with zero means, unit variances and correlation $\rho_{XZ}\neq 0$, and when $\Pi_{\mathcal{X}}$ and $\Pi_{\mathcal{Z}}$ are the standard Gaussian measure on $\mathcal{X}=\mathcal{Z}=\mathbb{R}$. The singular system of operator A_F is given by $\phi_j(x)=\dfrac{1}{\sqrt{j}}H_{j-1}(x)$, $\psi_j(z)=H_{j-1}(z)$ and $\omega_j=\dfrac{1}{\sqrt{j}}\rho_{XZ}^{j-1}$, where the H_j are the Hermite polynomials (see CFR, Section 2.3, for the case of A_F operating on $L^2(\mathbb{R})$ instead of $H^2(\mathbb{R})$). The adjoint operator is $A_F^*=\mathcal{D}^{-1}\tilde{A}$ (see Gagliardini and Scaillet (2006, GS)), where $\tilde{A}$ is the conditional expectation operator for Z given X defined by $\tilde{A}\psi(x)=\displaystyle\int_{\mathcal{Z}} f\,(z|x)\,\psi(z)dz$, and

$\mathcal{D}^{-1}$ is the inverse of the differential operator $\mathcal{D} = 1 - \nabla^2 + x\nabla$. Moreover $\ker(A_F^*) = \{0\}$, and thus $\bar{\mathcal{M}}_s$ is empty. We deduce that $F \in \mathcal{M}$ if and only if $\sum_{j=1}^{\infty} \frac{j}{\rho_{XZ}^{2(j-1)}} \langle r_F, H_{j-1}\rangle^2_{L^2(\mathcal{Z})} < \infty$ and the alternative $\bar{\mathcal{M}}$ consists of close misspecifications only: $\bar{\mathcal{M}} = \bar{\mathcal{M}}_c$.

Examples 1 and 2 (Cont.): *The condition for correct specification becomes* $\sum_{j=1}^{\infty} \frac{j}{\rho_{XZ}^{2(j-1)}} \langle \rho, H_{j-1}\rangle^2_{L^2(\mathcal{Z})} < \infty$. *In Example 1, the condition is satisfied since* $\rho(z) = \rho_{UZ} H_1(z)$, *and the series equals* $\frac{2\rho_{UZ}^2}{\rho_{XZ}^2}$. *In Example 2, the condition is not satisfied since* $\rho(z) = aI\{z \leq 0\} - aI\{z > 0\}$, $\langle \rho, H_{j-1}\rangle^2_{L^2(\mathcal{Z})} \geq C/(j-1)$ *for* $j > 1$ *and some constant* $C > 0$, *and the series diverges.*

2.3 Regularity spaces

In this section we introduce a sequence of function spaces, that characterize either the smoothness of the function φ_0 under the null hypothesis H_0 of correct specification, or the strength of a close misspecification under the alternative hypothesis H_1. These regularity spaces coincide with those introduced in Darolles, Florens, Renault (2003) and CFR, Section 3.2, under the null hypothesis H_0. We use these regularity spaces in Section 3 to derive the large sample properties of the test statistic. The postulated smoothness drives the rate of convergence to zero of the regularization bias under the null hypothesis (Section 3.2), and the rate of divergence of the noncentrality parameter under the alternative hypothesis (Section 3.3).

Let us define the sets for $\beta \geq 0$:

$$\mathcal{M}_\beta = \left\{ F \in \mathcal{F} : r_F \in \ker(A_F^*)^\perp \text{ and } \sum_{j=1}^{\infty} \frac{\langle r_F, \psi_j\rangle^2_{L^2(\mathcal{Z})}}{\nu_j^{1+\beta}} < \infty \right\}.$$

The sequence of sets $\mathcal{M}_\beta$ is decreasing w.r.t. the parameter β, that is $\mathcal{M}_{\beta_1} \subset \mathcal{M}_{\beta_2}$ for $\beta_1 \geq \beta_2$. We have $\mathcal{M}_0 = \mathcal{M}$ when $\beta = 0$. The condition $F_0 \in \mathcal{M}_\beta$ for $\beta \geq 0$ implies the null hypothesis H_0 of correct specification, and the parameter β characterizes the regularity of function $\varphi_0 \in H^2(\mathcal{X})$. More precisely, since $\left\langle r_{F_0}, \psi_j \right\rangle^2_{L^2(\mathcal{Z})} = \dfrac{1}{\nu_j} \left\langle A_{F_0}\varphi_0, A_{F_0}\phi_j \right\rangle^2_{L^2(\mathcal{Z})} = \nu_j \left\langle \varphi_0, \phi_j \right\rangle^2_{H^2(\mathcal{X})}$, the condition $F_0 \in \mathcal{M}_\beta$ is equivalent to the source condition $\varphi_0 \in \Phi_\beta := \left\{ \varphi \in H^2(\mathcal{X}) : \sum_{j=1}^{\infty} \dfrac{\left\langle \varphi, \phi_j \right\rangle^2_H}{\nu_j^\beta} < \infty \right\}$ introduced in Darolles, Florens, Renault (2003). The sets Φ_β, $\beta \geq 0$, are dense in $H^2(\mathcal{X})$ (CFR, Proposition 3.5), and the regularity of $\varphi_0 \in \Phi_\beta$ increases as β increases.

Similarly, let us define the sets for $-1 < \bar{\beta} \leq 0$:

$$\bar{\mathcal{M}}_{c,\bar{\beta}} = \left\{ F \in \mathcal{F} : r_F \in \ker\left(A_F^*\right)^\perp \text{ and } \sum_{j=1}^{\infty} \frac{\left\langle r_F, \psi_j \right\rangle^2_{L^2(\mathcal{Z})}}{\nu_j^{1+\bar{\beta}}} = \infty \right\}.$$

The sequence of sets $\bar{\mathcal{M}}_{c,\bar{\beta}}$ is increasing w.r.t. the parameter $\bar{\beta}$. We have $\bar{\mathcal{M}}_{c,0} = \bar{\mathcal{M}}_c$ when $\bar{\beta} = 0$. The condition $F_0 \in \bar{\mathcal{M}}_{c,\bar{\beta}}$ for $-1 < \bar{\beta} \leq 0$ implies the alternative hypothesis H_1 in the form of a close misspecification. The parameter $\bar{\beta}$ characterizes the strength of the misspecification, in terms of a lack of regularity of r_F, that increases as $\bar{\beta}$ decreases.

Examples 1 and 2 (Cont.): *In Example 1, $F \in \mathcal{M}_\beta$ for all $\beta \geq 0$. In Example 2, $F \in \bar{\mathcal{M}}_{c,\bar{\beta}}$ for all $-1 < \bar{\beta} \leq 0$.*

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)). They differ according to the definition of the operators, the scalar products, and the regularization scheme. Ideally we would like to develop a testing theory as versatile as possible irrespective of the chosen estimator. Unfortunately the asymptotic properties and the regularity conditions under the null and alternative hypotheses are much affected by these differences, and it is difficult to provide a unified treatment independent of how φ_0 is estimated. Here we focus on the approach of GS designed for functions in $H^2[0,1]$. The assumption of a compact support $\mathcal{X} = [0,1]$ greatly simplifies the derivation of the asymptotic properties. By the same token we set $\Pi_{\mathcal{X}}(dx) = dx$ and $\Pi_{\mathcal{Z}}(dz) = \Omega_0(z) I\{z \in S^*\} F_Z(dz)$, where Ω_0 is a given positive function on $\mathcal{Z}$ with $\Omega_0(z) = 1/V_0\left[Y - \varphi_0(X) | Z = z\right]$ under H_0, set $S^* \subset \mathcal{Z}$ is compact, and F_Z is the true cdf of Z. We consider the Tikhonov Regularized (TiR) estimator defined by

$$\hat{\varphi} = \arg \min_{\varphi \in H^2[0,1]} Q_T(\varphi) + \lambda_T \|\varphi\|_H^2 \ , \text{ where } Q_T(\varphi) = \frac{1}{T} \sum_{t=1}^{T} \Omega_0(Z_t) I\{Z_t \in S^*\} \left[\hat{r}(Z_t) - \hat{A}\varphi(Z_t)\right]^2, \tag{7}$$

$\hat{r}(z) = \int y \hat{f}(w|z)\, dw$, $\hat{A}\varphi(z) = \int \hat{f}(w|z)\, \varphi(x) dw$, and $\hat{f}$ is a kernel estimator of f. The minimum distance criterion $Q_T(\varphi)$ is penalized by a term that involves the squared Sobolev norm $\|\varphi\|_H^2$ (see Chernozhukov, Gagliardini and Scaillet (2007) for a theoretical underpinning for including a derivative term in a penalization approach). Penalization is required to overcome ill-posedness and is tuned by regularization parameter $\lambda_T > 0$, which converges to 0 as $T \to \infty$. The TiR estimator is given in closed form by $\hat{\varphi} = (\lambda_T + \hat{A}^* \hat{A})^{-1} \hat{A}^* \hat{r}$, where $\hat{A}^* = \mathcal{D}^{-1} \widehat{\tilde{A}}$ and $\widehat{\tilde{A}}$ denotes the linear operator defined by $\widehat{\tilde{A}}\psi(x) =$

$\frac{1}{T}\sum_{t=1}^{T}\hat{f}\left(x|Z_t\right)I\left\{Z_t\in S^*\right\}\Omega_0\left(Z_t\right)\psi\left(Z_t\right)$, for $\psi\in L^2(\mathcal{Z})$. The linear operator $\hat{A}^*$ is an estimator of $A^*_{F_0}$.

Following Sargan (1958), (1959) and Hansen (1982), the testing procedure is based on the minimized criterion value $Q_T\left(\hat{\varphi}\right)$. The value $Q_T\left(\hat{\varphi}\right)$ is an empirical counterpart of $Q_{\lambda_T} := E_0\left[\Omega_0(Z)I\left\{Z\in S^*\right\}\left[M_{\lambda_T}r_{F_0}(Z)\right]^2\right] = \left\|M_{\lambda_T}r_{F_0}\right\|^2_{L^2(\mathcal{Z})}$, where $M_{\lambda_T}r_{F_0} := \left[1-A_{F_0}\left(\lambda_T+A^*_{F_0}A_{F_0}\right)^{-1}A^*_{F_0}\right]r_{F_0}$ is the residual in the theoretical Tikhonov regression of Y on X with instrument Z. Indeed, $M_{\lambda_T}r_{F_0}=r_{F_0}-A_{F_0}\varphi_{\lambda_T}$, where $\varphi_{\lambda_T}=\left(\lambda_T+A^*_{F_0}A_{F_0}\right)^{-1}A^*_{F_0}r_{F_0}$ is the TiR solution, namely the population counterpart of the TiR estimator $\hat{\varphi}$ for given regularization parameter λ_T and the minimizer of the penalized criterion $\left\|r_{F_0}-A_{F_0}\varphi\right\|^2_{L^2(\mathcal{Z})}+\lambda_T\left\|\varphi\right\|^2_H$. The interpretation of $Q_T\left(\hat{\varphi}\right)$ as an empirical analog of the "weighted variance of residuals" Q_{λ_T} applies under both H_0 and H_1, and remains valid no matter the function space on which A_F operates. Under H_0 the TiR solution φ_{λ_T} converges to φ_0 as T goes to infinity. Under H_1 it may converge to a pseudo-true value, but it may also not converge (see Section 3.3 and CFR, Section 3.1, for a discussion).

The test statistic is built from $Q_T\left(\hat{\varphi}\right)$ after appropriate redefinition of the smoothing, recentering and scaling. Specifically, we first replace the integrals w.r.t. kernel density estimator $\hat{f}$ with kernel regression estimators which are easier to compute. Namely, we use the asymptotic equivalence (see Appendix 2.1) between $Q_T\left(\hat{\varphi}\right)$ and the statistic

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

$$\psi_{ts} = \frac{\Omega_0(Z_t)^{1/2}\left(Y_s - \hat{\varphi}(X_s)\right) K\left(\frac{Z_s - Z_t}{h_T}\right) I\left\{Z_t \in S_*\right\}}{\sum_{j=1}^{T} K\left(\frac{Z_j - Z_t}{h_T}\right)}, \tag{8}$$

where K is a kernel and h_T is a bandwidth. Then, as in TK we recenter the statistic ξ_T by subtracting the diverging term $\xi_{2,T} = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\psi_{ts}^2$ to allow for a well-defined asymptotic distribution under the null hypothesis. After recentering we can exploit the Central Limit Theorem (CLT) for generalized quadratic forms in de Jong (1987), which is a generalization of the CLT for degenerate U-statistics in Hall (1984). The test statistic is

$$\zeta_T := T h_T^{1/2}\frac{\xi_T - \xi_{2,T}}{\sigma},$$

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

Finally, note that the trimming is based on a fixed support S_*. This 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_*. To get a coherent and simplified exposition, we have introduced the same fixed trimming in the definition of the norm $L^2(\mathcal{Z})$ and of the estimator $\hat{\varphi}$, although this is not required by fundamental reasons.

3.2 The asymptotic distribution under the null hypothesis

Let us assume that the null hypothesis H_0 of correct specification holds and F_0 is in a regularity space introduced in Section 2.3: $F_0 \in \mathcal{M}_\beta$ for a given $0 \leq \beta \leq 1$. The restriction $\beta \leq 1$ comes from a saturation effect: stronger regularity with $\beta > 1$ cannot be exploited to reduce the bias contribution in a Tikhonov regularization setting (see CFR, Section 3.3). Suppose that the bandwidth h_T and the regularization parameter λ_T converge to zero as $T \to \infty$ with rates described next.

Assumption 3: $h_T = \bar{c}T^{-\bar{\eta}}$ *with:* (i) $2/9 < \bar{\eta}$; (ii) $\bar{\eta} < \min\{1 - 4/m, 1/3\}$, *where* $m > 4$ *is defined in Assumption A.2.*

Assumption 4: $\lambda_T = cT^{-\gamma}$ *with:* (i) $\dfrac{1-\bar{\eta}/2}{1+\beta} < \gamma$; (ii) $\gamma < \min\{4\bar{\eta}, 1\}$.

Proposition 1: *Under the null hypothesis* H_0 *for* $F_0 \in \mathcal{M}_\beta$ *with* $0 \leq \beta \leq 1$, *Assumptions 1-4 and A.1-A.8, we have* $\zeta_T \xrightarrow{d} N(0,1)$.

Proof: See Appendix 2.

The proof of Proposition 1 builds on TK, and consists in first isolating the impact of the estimation of φ_0 on the test statistic, and then applying the CLT for generalized quadratic forms in de Jong (1987) to the test statistic based on $Q_T(\varphi_0)$. Under Assumptions 3 and 4 on the interplay between the bandwidth and the regularization parameter the asymptotic distribution under H_0 is unaffected by the use of estimate $\hat{\varphi}$ instead of the true function φ_0 in the criterion $Q_T(\varphi)$. This explains why the asymptotic distribution of ζ_T under H_0 is $N(0,1)$ as for the specification test of parametric conditional moment restrictions in TK.

Assumption 3(ii) on the bandwidth corresponds to the condition in Theorem 4.1 of TK for a linear-in-parameter moment condition. [2] In our ill-posed inverse problem setting, however, the control of the impact of estimation of φ_0 on the test statistic is more complicated, because of the regularization bias and the lower rate of convergence of the estimator $\hat{\varphi}$ (see GS for such a rate). More specifically, the regularization bias $\mathcal{B}_T = \left[\left(\lambda_T + A_{F_0}^* A_{F_0}\right)^{-1} A_{F_0}^* A_{F_0} - 1\right] \varphi_0 = \varphi_{\lambda_T} - \varphi_0$ of the estimator $\hat{\varphi}$ contributes the term

$$Th_T^{1/2} E_0 \left[\Omega_0(Z) I \left\{Z \in S^*\right\} \left[A_{F_0} \mathcal{B}_T(Z)\right]^2\right] = Th_T^{1/2} Q_{\lambda_T}, \tag{9}$$

to the mean of the test statistic. This term is of order $Th_T^{1/2}\lambda_T^{1+\beta}$ under $F_0 \in \mathcal{M}_\beta$, $0 \leq \beta \leq 1$, and vanishes asymptotically under Assumption 4(i) (see Appendix 2.3). Assumption 4(ii) is used to prove the asymptotic negligibility of terms induced by the estimation error $\hat{\varphi} - \varphi_{\lambda_T}$.

Finally, 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 $\xi_{5,T} := \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1, s\neq t}^{T} \sum_{u=1, u\neq t, u\neq s}^{T} \psi_{ts}\psi_{tu}$.

3.3 Consistency under global alternatives

Let us now assume that the alternative hypothesis H_1 holds.

Proposition 2: *Under the alternative hypothesis H_1, Assumptions 1-3, 4(ii) and A.1-A.8, if $\tau_T := Th_T^{1/2} Q_{\lambda_T} \to \infty$ as $T \to \infty$, we have $\sigma\zeta_T = \tau_T + z_T + Th_T^{1/2}\xi_{5,T}^{*,E} + o_p(\tau_T) + O_p(1)$, where $z_T \xrightarrow{d} N\left(0, \sigma^{*2}\right)$ with σ^{*2} defined in (18), and $\xi_{5,T}^{*,E}$ is the contribution of the estimation*

[2] To see this, set $\eta = \infty$ in Assumption 3.6 of TK. Our Assumption 3(i) is used to prove the asymptotic equivalence of $Q_T(\hat{\varphi})$ and ξ_T in Section A.2.1.

error $\hat{\varphi} - \varphi_{\lambda_T}$ defined in (17).

Proof: See Appendix 3.

Proposition 2 shows that the asymptotics of the statistic ζ_T depend on the relative orders of magnitude of τ_T and $Th_T^{1/2}\xi_{5,T}^{*,E}$. The non-centrality parameter τ_T reduces to the bias contribution in (9) under the null hypothesis. In general it is difficult to explicitly characterize the behaviour of the estimation error term $Th_T^{1/2}\xi_{5,T}^{*,E}$ under H_1. For example, as already noticed in Section 3.1, φ_{λ_T} may or may not converge under H_1. We focus on the case where estimation error is negligible compared to the non-centrality parameter, and we consider the next high-level assumption. This parallels the analysis under H_0, where we show that the estimation error term is asymptotically negligible.

Assumption 5: *Under the alternative hypothesis H_1: $Th_T^{1/2}\xi_{5,T}^{*,E} = o_p(\tau_T)$.*

Assumption 5 implies that $|\zeta_T| \geq C\tau_T$ for a constant $C > 0$.

To characterize the asymptotic behaviour of τ_T under the alternative hypothesis, we can decompose $r_{F_0} = \mathcal{P}r_{F_0} + \mathcal{P}^{\perp}r_{F_0}$, where $\mathcal{P}$ and $\mathcal{P}^{\perp}$ denote the orthogonal projection operators on $\ker\left(A_{F_0}^*\right)$ and on $\ker\left(A_{F_0}^*\right)^{\perp}$, respectively. Then $Q_{\lambda_T} = \left\|M_{\lambda_T}\mathcal{P}^{\perp}r_{F_0}\right\|_{L^2(\mathcal{Z})}^2 + \left\|\mathcal{P}r_{F_0}\right\|_{L^2(\mathcal{Z})}^2$, using $M_{\lambda_T}\mathcal{P} = \mathcal{P}$ and $\mathcal{P}M_{\lambda_T}\mathcal{P}^{\perp} = 0$. Since ψ_j is an orthonormal basis of $\ker\left(A_{F_0}^*\right)^{\perp}$, we have $\mathcal{P}^{\perp}r_{F_0} = \sum_{j=1}^{\infty}\left\langle r_{F_0}, \psi_j\right\rangle_{L^2(\mathcal{Z})}\psi_j$ and we get:

$$\tau_T = Th_T^{1/2}\sum_{j=1}^{\infty}\frac{\lambda_T^2}{\left(\lambda_T + \nu_j\right)^2}\left\langle r_{F_0}, \psi_j\right\rangle_{L^2(\mathcal{Z})}^2 + Th_T^{1/2}\left\|\mathcal{P}r_{F_0}\right\|_{L^2(\mathcal{Z})}^2. \tag{10}$$

We distinguish between close and separated misspecifications. Let us first consider close alternatives, and assume that F_0 is in a regularity space introduced in Section 2.3: $F_0 \in \bar{\mathcal{M}}_{c,\bar{\beta}}$

for $-1 < \bar{\beta} \leq 0$. Then $\mathcal{P}r_{F_0} = 0$, [3] and the behaviour of the series in (10) is driven by the decay of the coefficients $\langle r_{F_0}, \psi_j \rangle_{L^2(\mathcal{Z})}$. We show in Appendix 4 that $\tau_T \geq C T h_T^{1/2} \lambda_T^{\delta+\bar{\beta}}$, for a constant $C > 0$ and any $\delta > 1$. Thus τ_T diverges under the next Assumption 4(iii) on the regularization parameter, which implies Assumption 4(ii).

Assumption 4: $\lambda_T = cT^{-\gamma}$ *with:* (iii) $\gamma < \min\left\{4\bar{\eta}, \dfrac{1-\bar{\eta}/2}{1+\bar{\beta}}, 1\right\}$.

We deduce the following bound on the divergence rate of ζ_T:

Proposition 3: *Under the alternative hypothesis H_1 for $F_0 \in \bar{\mathcal{M}}_{c,\bar{\beta}}$ with $-1 < \bar{\beta} \leq 0$, Assumptions 1-3, 4(iii), 5 and A.1-A.8, we have $T h_T^{1/2} \lambda_T^{\delta+\bar{\beta}} / \zeta_T = o_p(1)$ for any $\delta > 1$.*

Proof: See Appendix 4.

Let us now consider separated alternatives: $F_0 \in \bar{\mathcal{M}}_s$. Then $\mathcal{P}r_{F_0} \neq 0$ and it follows from (10) that $\tau_T \geq \|\mathcal{P}r_{F_0}\|^2_{L^2(\mathcal{Z})} T h_T^{1/2}$. Thus, τ_T diverges under Assumption 3. We deduce:

Proposition 4: *Under the alternative hypothesis H_1 for $F_0 \in \bar{\mathcal{M}}_s$, Assumptions 1-3, 4(ii), 5 and A.1-A.8, we have $T h_T^{1/2} / \zeta_T = O_p(1)$.*

The above results reveal that the bound on the divergence rate of the test statistic is larger under separated misspecification than under close misspecification. Under $F_0 \in \bar{\mathcal{M}}_s$, the bound corresponds to the divergence rate for the specification test of a parametric conditional moment restriction in TK, namely of the order $T h_T^{1/2}$. Under $F_0 \in \bar{\mathcal{M}}_{c,\bar{\beta}}$, the bound is close to the divergence rate in TK for the strongest departures from correct specification: when $\bar{\beta}$ is near -1, we get the order $T^{1-\epsilon} h_T^{1/2}$, $\epsilon > 0$.

We can combine Propositions 1-4 to introduce specification tests that have given as-

[3] Under H_0, we also have $\mathcal{P}r_{F_0} = 0$ since $r_{F_0} \in \ker\left(A^*_{F_0}\right)^{\perp}$.

ymptotic size under the null hypothesis, and are consistent against alternatives in suitable regularity spaces.

A. For a given $-1 < \bar{\beta}^* < 0$ let the alternative hypothesis $H_1(\bar{\beta}^*)$ be defined by $F_0 \in \bar{\mathcal{M}}_{c,\bar{\beta}^*} \cup \bar{\mathcal{M}}_s$. Let us consider the statistic ζ_T with bandwidth h_T and regularization parameter λ_T satisfying Assumptions 3, 4(i) with $\beta = 0$, and 4(iii) with $\bar{\beta} = \bar{\beta}^*$. Then, $P\left(|\zeta_T| > z_{1-\alpha/2}\right) \to \alpha$ under H_0 and $P\left(|\zeta_T| > z_{1-\alpha/2}\right) \to 1$ under $H_1(\bar{\beta}^*)$ as $T \to \infty$, where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the $N(0,1)$ distribution for $\alpha \in (0,1)$. Thus, statistic ζ_T yields a consistent test of H_0 against the alternative $H_1(\bar{\beta}^*)$. A test based on statistic ζ_T with given asymptotic size α under the null hypothesis H_0 has no asymptotic power against the alternative hypothesis $F_0 \in \bar{\mathcal{M}}_c \setminus \left(\bigcup_{-1<\bar{\beta}<0} \bar{\mathcal{M}}_{c,\bar{\beta}}\right)$. Indeed, if we set $\bar{\beta}^* = 0$, we face an incompatibility between the conditions on λ_T to have a vanishing regularization bias under the null hypothesis, and a diverging noncentrality parameter under the alternative hypothesis. Intuitively, it is difficult to distinguish the close misspecification $\bar{\mathcal{M}}_c \setminus \left(\bigcup_{-1<\bar{\beta}<0} \bar{\mathcal{M}}_{c,\bar{\beta}}\right)$ from a correctly specified model with minimal smoothness, i.e., $\varphi_0 \in \Phi_0$ under the null hypothesis.

B. For a given $0 < \beta^* \leq 1$ let the null hypothesis $H_0(\beta^*)$ be defined by $F_0 \in \mathcal{M}_{\beta^*}$. Let us consider the statistic ζ_T with bandwidth h_T and regularization parameter λ_T satisfying Assumptions 3, 4(i) with $\beta = \beta^*$, and 4(ii) with $\bar{\beta} = 0$. Then, $P\left(|\zeta_T| > z_{1-\alpha/2}\right) \to \alpha$ under $H_0(\beta^*)$ and $P\left(|\zeta_T| > z_{1-\alpha/2}\right) \to 1$ under H_1 as $T \to \infty$. Thus, statistic ζ_T yields a consistent test of $H_0(\beta^*)$ against the full set H_1 of alternatives.

3.4 Extension to an estimated weighting function

In the previous sections, results have been presented for a known weighting matrix to ease reading and derivation. Let us now replace $\Omega_0(z)$ by an estimate $\hat{\Omega}(z) = \hat{V}(z)^{-1}$ in (7) based on a kernel regression estimator of the conditional variance $V_0(z)$ and a pilot estimator $\bar{\varphi}$ of φ_0. Under H_0 the analysis remains virtually unchanged, and Proposition 1 holds under the supplementary assumptions A.9-A.10 on $V_0(z)$ and $\bar{\varphi}$ in Appendix A.5.1. The analysis complicates under global alternatives. The estimation of the weighting function plays a nontrivial role as opposed to the standard GMM setting. The difficulty comes from the population counterpart $\bar{\varphi}_{\lambda_T}$ of the first-step estimator $\bar{\varphi}$ that may not converge under H_1. Then, the estimated weighting function $\hat{\Omega}(z)$ used to compute the test statistic may not converge. This affects the definition of the population analogue φ_{λ_T} of the estimator $\hat{\varphi}$, since the limit of $\hat{A}^*$ might differ from $A^*_{F_0}$ whose definition is based on $\Omega_0(z)$. We overcome this difficulty by introducing a norm based on the population analogue $\Omega_{\lambda_T}(z)$ of $\hat{\Omega}(z)$ (see Appendix A.5.2). Then, the results in Propositions 3 and 4 can be derived under the supplementary assumptions A.11-A.13, that control for the behavior of the norm induced by $\Omega_{\lambda_T}(z)$ and for the uniform convergence of $\hat{\Omega}(z) - \Omega_{\lambda_T}(z)$ to 0. We summarize the results as follows.

Proposition 5: *Under Assumptions 1-5 and A.1-A.13, the results in Proposition 1 for* $F_0 \in \mathcal{M}_\beta$, $0 \leq \beta \leq 1$, *Proposition 3 for* $F_0 \in \bar{\mathcal{M}}_{c,\bar{\beta}}$, $-1 < \bar{\beta} \leq 0$, *and Proposition 4 for* $F_0 \in \bar{\mathcal{M}}_s$, *hold.*

Proof: See Appendix 5.

3.5 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)). However, the usual bootstrap of testing procedures based on degenerate U-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). [4] 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 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 $\bar{U}_t := Y_t - \hat{\varphi}(X_t) - \Big(\hat{r}(Z_t) - \hat{A}\hat{\varphi}(Z_t)\Big)$, $t = 1, ..., T$.

2. Make T independent draws $(\tilde{X}_{t,b}, \tilde{Z}_{t,b}, \tilde{U}_{t,b})$ with replacement from $\big\{(X_t, Z_t, \bar{U}_t);\ 1 \leq t \leq T\big\}$, and take $\tilde{Y}_{t,b} := \hat{\varphi}(\tilde{X}_{t,b}) + \tilde{U}_{t,b}$ to get the bootstrap sample $(\tilde{X}_{t,b}, \tilde{Y}_{t,b}, \tilde{Z}_{t,b})$, $t = 1, ..., T$.

3. Compute the bootstrap statistic $\tilde{\zeta}_{T,b}$ based on the bootstrap sample.

[4] 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.

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 $E[\tilde{Y} - \varphi_{\lambda_T}(\tilde{X})|\tilde{Z} = z] = 0$ and $E\left[\left(\tilde{Y} - \varphi_{\lambda_T}(\tilde{X})\right)^2 |\tilde{Z} = z\right] = \Omega_{\lambda_T}(z)^{-1}$ on the bootstrap sample whether H_0 holds or not. A test based on the decision rule in Step 5 is consistent: it satisfies $\lim_{T\to\infty} P[reject\, H_0] = \alpha$ if H_0 is true, and $\lim_{T\to\infty} P[reject\, H_0] = 1$ 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 proofs of Propositions 1-5 but applied to the bootstrap sample instead of the original sample. Therefore we omit these developments. In our Monte Carlo results the bootstrap reduces significantly the finite sample size distortions 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 Sections 3.2 and 4.2.

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

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

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 $\rho_{UV} \neq 0$ there is endogeneity, and an instrumental variable estimation is required. The moment condition is $E_0 [Y - \varphi_0(X) \mid Z] = 0$, 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 illustration.

4.2 Computation of the test statistic

The estimation of φ_0 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:

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

Such a simple and exact form eases implementation [5] , and improves on speed.

The kernel estimator of the conditional moment $\hat{r}(z) - \hat{A}\varphi(z)$ is approximated through $\hat{r}(z) - \theta'\hat{P}(z)$ where $\hat{P}(z) = \sum_{t=1}^{T} P(X_t) K\left(\frac{Z_t - z}{h_T}\right) / \sum_{t=1}^{T} K\left(\frac{Z_t - z}{h_T}\right)$, $\hat{r}(z) = \sum_{t=1}^{T} Y_t K\left(\frac{Z_t - z}{h_T}\right) / \sum_{t=1}^{T} K\left(\frac{Z_t - z}{h_T}\right)$, 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. [6] Here weighting function $\Omega_0(z)$ is equal to unity, and assumed to be known.

For programming purpose, this test statistic can be expressed in a matrix format:

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

where Ψ is the $T \times T$ matrix with elements ψ_{ts} in (8), ι is a $T \times 1$ vector of ones, and $\odot$ denotes

[5] The Gauss programs developed for this section and the empirical illustration are available on request from the authors.

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

the Hadamard (or element-by-element) product. We also consider an asymptotically equivalent statistic based on the penalized value of the criterion, namely $\zeta_T + Th_T^{1/2}\lambda_T \|\hat{\varphi}\|_H^2 / \sigma$. [7]

Other possibilities include statistics such as:

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

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

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 $\lambda \in \{.00001, .0007, .0009, .0012, .0015, .005\}$. The values .0009 and .0007 are the 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 with slight overpenalization. Therefore we also consider the values .0012 and .0015. 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.

[7] Statistic $\zeta_T + Th_T^{1/2}\lambda_T \|\hat{\varphi}\|_H^2 / \sigma$ is asymptotically equivalent to ζ_T under H_0 if estimator $\hat{\varphi}$ is such that $\|\hat{\varphi}\|_H = O_p(1)$, and $\gamma > 1 - \bar{\eta}/2$ in Assumption 4(i).

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 1% confidence level. In light of this, we advocate to use the bootstrap procedure of Section 3.5 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, while for $\lambda = .0012, .0015$, the rejection rates are close to the nominal ones. For $\lambda = .0007, .0009$, statistic $\zeta_T + Th_T^{1/2}\lambda_T \|\hat{\varphi}\|_H^2 /\sigma$ features good finite sample properties and yields tests which are only slightly undersized. For $\lambda = .0012, .0015$, statistic $\zeta_T + Th_T^{1/2}\lambda_T \|\hat{\varphi}\|_H^2 /\sigma$ provides oversized tests at $\alpha = .10$. Selecting the very small regularization parameter $\lambda = .00001$ results in undersized tests, both for ζ_T and $\zeta_T + Th_T^{1/2}\lambda_T \|\hat{\varphi}\|_H^2 /\sigma$. For the very large value $\lambda = .005$, the test becomes oversized because of regularization bias.

	Rejection rates with 1000 repetitions for ζ_T and $\zeta_T + Th_T^{1/2}\lambda_T \|\hat{\varphi}\|_H^2 /\sigma$											
	$\lambda = .00001$		$\lambda = .0007$		$\lambda = .0009$		$\lambda = .0012$		$\lambda = .0015$		$\lambda = .005$	
$\alpha = .01$	.002	.000	.005	.010	.008	.010	.013	.002	.013	.005	.087	.198
$\alpha = .05$	.016	.004	.013	.025	.033	.039	.046	.046	.073	.073	.271	.533
$\alpha = .10$	.048	.063	.036	.072	.049	.092	.102	.141	.107	.195	.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 $\zeta_T + T h_T^{1/2} \lambda_T \left\| \hat{\varphi} \right\|_H^2 / \sigma$ (right column). We generate Y as $Y = \sin(\pi X) + U + \eta$. In design 1 we take $\eta = .20 I\{Z \leq 0\} - .20 I\{Z > 0\}$. This yields $E_0[Y - \sin(\pi X) | Z = z] = .20 I\{z \leq 0\} - .20 I\{z > 0\}$, 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 $U + \eta$ is 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 $\lambda = .0007, .0009, .0012, .0015$, under both designs. The statistic ζ_T gives better power properties in design 1, while the statistic $\zeta_T + T h_T^{1/2} \lambda_T \left\| \hat{\varphi} \right\|_H^2 / \sigma$ gives better power properties in design 2. For value $\lambda = .00001$, i.e., for a very light penalization, the power is minimal under both designs. In our simulation experiments, choosing a regularization parameter value around the values minimizing the asymptotic or finite sample MISE delivers good performance. Of course there is no reason why an optimal choice for estimation should be optimal for testing. The design of an adaptive rate-optimal test is a challenging task even in the parametric case (Horowitz and Spokoiny (2001)), and we leave this interesting research topic to future work.

	Rejection rates with 1000 repetitions for ζ_T and $\zeta_T + Th_T^{1/2}\lambda_T \left\|\hat{\varphi}\right\|_H^2 / \sigma$											
Design 1	$\lambda = .00001$		$\lambda = .0007$		$\lambda = .0009$		$\lambda = .0012$		$\lambda = .0015$		$\lambda = .005$	
$\alpha = .01$	.006	.002	.648	.074	.742	.084	.812	.141	.842	.155	.896	.499
$\alpha = .05$	.029	.012	.829	.290	.876	.337	.932	.419	.942	.444	.966	.811
$\alpha = .10$	.073	.049	.859	.473	.902	.538	.948	.576	.952	.647	.976	.916
Design 2	$\lambda = .00001$		$\lambda = .0007$		$\lambda = .0009$		$\lambda = .0012$		$\lambda = .0015$		$\lambda = .005$	
$\alpha = .01$	.005	.002	.056	.078	.082	.142	.120	.338	.189	.566	.963	1.000
$\alpha = .05$	.013	.009	.201	.310	.251	.473	.326	.759	.438	.920	.995	1.000
$\alpha = .10$	.043	.044	.289	.513	.361	.736	.471	.930	.579	.991	.998	1.000

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

5 An empirical illustration

This section presents an empirical example with the data in Horowitz (2006) and GS. [8] We aim at testing the specification of an Engel curve based on the moment condition $E_0 [Y - \varphi_0 (X) \mid Z] = 0$, 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 reg-

[8] We would like to thank Joel Horowitz for kindly providing the dataset.

ularization 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 $\zeta_T + T h_T^{1/2} \lambda_T \left\| \hat{\varphi} \right\|_H^2 / \sigma = -.3017$, respectively. Hence we do not reject the null hypothesis of a correct specification of the Engel curve modeling.

Appendices

In Appendix 1 we list the regularity conditions and provide their detailed discussion. In Appendix 2 we show Proposition 1 on asymptotic normality of our test statistic under the null hypothesis. In Appendix 3 and 4 we show Propositions 2 and 3 on the behavior of the test statistic under global alternatives. In Appendix 5 we provide the details for extending the proofs to the setting with an estimated weighting function.

Appendix 1: List of regularity conditions

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

A.2: *For* $m > 4$: (i) $E_0\left[|Y - \varphi_0(X)|^m\right] < \infty$ *under* H_0, *and* (ii) $E_0\left[\left|Y - \varphi_{\lambda_T}(X)\right|^m\right] = O(1)$ *and* $E_0\left[|Y - E_0[Y|Z]|^m\right] < \infty$ *under* H_1.

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

A.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* $a \in (0,1)$.

A.5: *Estimator* $\hat{\varphi}$ *is such that:* (i) $\frac{1}{T}\sum_t \left|\hat{\varphi}(X_t) - \varphi_{\lambda_T}(X_t)\right|^2 = O_p(T^{-1/3})$; (ii) $\sup_{\mathcal{X}} \left|\nabla^2\hat{\varphi}\right| = O_p(1)$.

A.6: *Under* H_0: (i) $\sup_{z\in S^*}\left(\frac{1}{Th_T}\sum_s K\left(\frac{Z_s - z}{h_T}\right)\mathcal{R}_T(X_s)\right)^2 = o_p\left(\frac{1}{Th_T^{1/2}}\right)$, *where* $\mathcal{R}_T :=$

$\hat{\varphi}-\varphi_{\lambda_T}-\left(\lambda_T+A_{F_0}^*A_{F_0}\right)^{-1}A_{F_0}^*\hat{\psi}-\left(\lambda_T+A_{F_0}^*A_{F_0}\right)^{-1}\left(\hat{A}^*\hat{A}-A_{F_0}^*A_{F_0}\right)\mathcal{B}_T$ *and* $\hat{\psi}(z):=\int(y-\varphi_0(x))\frac{\hat{f}(w,z)}{f(z)}dw$; (ii) $\sup_{z\in S^*}\left(\frac{1}{Th_T}\sum_s K\left(\frac{Z_s-z}{h_T}\right)\mathcal{R}_T(X_s)^2\right)=o_p\left(\frac{1}{T}\right)$.

A.7: (i) *The eigenvalues* ν_j *of operator* $A_{F_0}^*A_{F_0}$ *are such that* $C_1e^{-\alpha j}\leq\nu_j\leq C_2e^{-\alpha j}$, $j\in\mathbb{N}$, *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* $\sup_{j\in\mathbb{N}}\sup_{|u|\leq 1}E_0\left[\nabla^2\left(A_{F_0}\phi_j\right)(Z+h_Tu)^2\right]=O(1)$, *as* $h_T\to 0$, *and* (iii) $\sup_{j\in\mathbb{N}}\sup_{z\in S^*}E_0\left[\phi_j(X)^2|Z=z\right]<\infty$.

A.8: *The function* $\varphi_0\in H^2[0,1]$ *is in class* $C^2(0,1)$ *with* $\sup_{\mathcal{X}}\left|\nabla^2\varphi_0\right|<\infty$ *under* H_0.

Assumptions A.1, A.2 (i), A.3, A.4 yield the assumptions used in TK for testing parametric conditional moment restrictions in the special case of a linear-in-parameter moment function and known weighting function. In our functional setting, compacity of $\mathcal{X}$ in Assumption A.1 eases the definition of the parameter space, which is a subset of the Sobolev space $H^2[0,1]$. We take univariate variables to avoid matrix notation and facilitate the writing and reading of the results and proofs. All our results can be extended to $\dim X>1$ and $\dim Z>1$. Assumption A.1 (i) on the conditional density $f_{X|Z}$ implies that operator $A_{F_0}:L^2[0,1]\to L^2(\mathcal{Z})$ is compact, and this yields compacity of A_{F_0} defined on $H^2[0,1]$. Assumption A.1 (ii) on the conditional density f_Z, together with Assumption A.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 A.2 is a condition ensuring finite higher moments of the innovation under H_0, and similar conditions under H_1. In particular, Assumption A.2 (ii) is used in the proofs of Lemmas C.1-C.4. The compact set S_* in

Assumption A.3 solves boundary problems of kernel estimators. Assumption A.5 concerns properties of the functional estimator $\hat{\varphi}$ that are used in the proofs of technical lemmas. Specifically, Assumption A.5 (i) is used in the proof of Lemmas B.1-B.2 and C.1-C.2 to prove asymptotic negligibility of two components of the test statistic. Assumption A.5 (ii) is used to prove the asymptotic equivalence of Q_T and ξ_T in Section A.2.1. Function $\mathcal{R}_T$ in Assumption A.6 is the reminder term in the linearization of the estimation error $\hat{\varphi}-\varphi_{\lambda_T}$ performed in Section A.2.4. The linearization includes a term induced by estimation of $A_{F_0}^* r_{F_0}$ and a term induced by estimation of $A_{F_0}^* A_{F_0}$. The reminder term $\mathcal{R}_T$ is of second-order w.r.t $\hat{\psi}$ and $\hat{A}^*\hat{A} - A_{F_0}^* A_{F_0}$. Assumption A.6 is satisfied e.g. when $\mathcal{R}_T$ is $O_p(1/\sqrt{T})$, and allows us to control the reminder contribution coming from the estimation error in Lemmas B.7-B.8. Assumption A.7 concerns the spectral decomposition of compact operator $A_{F_0}^* A_{F_0}$ (see Kress (1999), Chapter 14). In Assumption A.7 (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 B.5 and B.6. In GS we verify that Assumption A.7 (i) is satisfied in the Monte-Carlo setting of Section 4. Our results extend to the case of hyperbolic decay (mild ill-posedness). Regularity Assumptions A.7 (ii)-(iii) on the eigenfunctions ϕ_j are used in the proof of technical Lemmas B.5 and B.6. Assumption A.8 concerns the smoothness of function φ_0 under H_0. Second-order differentiability of φ_0 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 B.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)), at the cost however of an increase in the technical complexity. Since estimator $\hat{\varphi}$ is not the focus of this paper, we do not detail modifications induced by alternative estimation approaches.

Appendix 2: Proof of Proposition 1

In A.2.1 we show the equivalence between $Q_T(\hat{\varphi})$ and ξ_T. In A.2.2 we establish the asymptotic normality of the leading term before showing in A.2.3 and A.2.4 that the other terms are negligible. We gather in A.2.5 the technical lemmas and discuss their main differences w.r.t. TK. In this appendix and hereafter we omit subscripts in densities, expectations and operators. Furthermore, let $\mathcal{T}_* = \{1 \le t \le T : Z_t \in S_*\}$, $K_{st} = K\left(\dfrac{Z_s - Z_t}{h_T}\right)$, $\Omega_t = \Omega_0(Z_t)$, $U_t = Y_t - \varphi_0(X_t)$, $g_{\varphi_0}(w) = y - \varphi_0(x)$, $\Delta\varphi = \varphi - \varphi_0$, $I_t = I\{Z_t \in S_*\}$, $\mathcal{I} = \{Z_t : 1 \le t \le T\}$, $H_0(z) = V_0(z) f(z)^2$, $w_{st} = K_{st} / \sum_j K_{jt}$, $\mathcal{K}_T(V, W) = \dfrac{1}{T} \sum_t \dfrac{\Omega_t I_t}{\left(\sum_j K_{jt}\right)^2} \sum_{s \ne t} \sum_{u \ne t,s} K_{st} K_{ut} V_s W_u$.

A.2.1 Asymptotically equivalent statistics

Let us consider $\xi_T := \dfrac{1}{T} \sum_{t=1}^T \left(\sum_{s=1}^T \psi_{ts} \right)^2$, where $\psi_{ts} = \Omega_t^{1/2} I_t \left(Y_s - \hat{\varphi}(X_s)\right) K_{st} \Big/ \sum_{j=1}^T K_{jt}$. 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 A.5 (ii) to get the asymptotic equivalence

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

uniformly in $z \in S_*$, and Cauchy-Schwarz inequality, we get $Q_T(\hat{\varphi}) = \xi_T + O_p(\xi_T^{1/2} h_T^2) + O_p(h_T^4)$. Using $h_T = \bar{c}T^{-\bar{\eta}}$ with $\bar{\eta} > 2/9$, we get $Q_T(\hat{\varphi}) = \xi_T + o_p((Th_T^{1/2})^{-1})$. Thus, statistics $Q_T(\hat{\varphi})$ 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

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

Terms $\xi_{1,T}, \xi_{3,T}$ and $\xi_{4,T}$ are $o_p((Th_T^{1/2})^{-1})$ (see Lemmas B.1 and B.2 in Section A.2.5), 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

$$\xi_T - \xi_{2,T} = \xi_{5,T} + o_p((Th_T^{1/2})^{-1}). \tag{11}$$

Let us rewrite statistic $\xi_{5,T}$ in order to identify the different contributions. We use the following decomposition:

$$Y - \hat{\varphi}(X) = U - \mathcal{B}_T(X) - \mathcal{E}_T(X). \tag{12}$$

The innovation $U := Y - \varphi_0(X)$ is such that $E[U|Z] = 0$. The bias $\mathcal{B}_T(X)$ is such that $E[\mathcal{B}_T(X)|Z] = -M_{\lambda_T} r_{F_0}(Z)$, that is, minus the Tikhonov residual. Finally, $\mathcal{E}_T(X) := \hat{\varphi}(X) - \varphi_{\lambda_T}(X)$ is the estimation error. We get the decomposition $\xi_{5,T} = \bar{\xi}_{5,T} + \xi_{5,T}^B + \xi_{5,T}^E$, where the leading contribution is $\bar{\xi}_{5,T} = \mathcal{K}_T(U,U)$, the contribution induced by regularization bias is given by $\xi_{5,T}^{B} = \mathcal{K}_T\left(\mathcal{B}_T(X), \mathcal{B}_T(X)\right) - 2\mathcal{K}_T\left(U, \mathcal{B}_T(X)\right)$, and the contribution accounting for estimation error is $\xi_{5,T}^{E} = \mathcal{K}_T\left(\mathcal{E}_T(X), \mathcal{E}_T(X)\right) - 2\mathcal{K}_T\left(U - \mathcal{B}_T(X), \mathcal{E}_T(X)\right)$.

A.2.2 Asymptotic normality of the test statistic

Statistic $\bar{\xi}_{5,T}$ 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, $\bar{\xi}_{5,T} = \dfrac{1}{T^3 h_T^2} \sum_t H_0(Z_t)^{-1} I_t \sum_{s \neq t} \sum_{u \neq t,s} K_{st} K_{ut} U_s U_u + O_p\left(\dfrac{\log T}{T h_T} \sup_{z \in S_*} \left| \hat{f}(z)^{-1} - f(z)^{-1} \right|\right)$. Using $\sup_{z \in S_*} \left| \hat{f}(z)^{-1} - f(z)^{-1} \right| = O_p\left(\sqrt{\dfrac{\log T}{T h_T}} + h_T^2\right)$ from Assumptions A.1, A.3 and A.4, the CLT for generalized quadratic forms of de Jong (1987) along the lines of Lemma A.6 in TK, and $h_T = \bar{c} T^{-\bar{\eta}}$ with $2/9 < \bar{\eta} < \min\{1 - 4/m, 1/3\}$, we get $T h_T^{1/2} \bar{\xi}_{5,T} \xrightarrow{d} N(0, 2K_{**} vol(S_*))$. Then, Proposition 1 follows using that $\xi_{5,T}^{B}, \xi_{5,T}^{E} = o_p((T h_T^{1/2})^{-1})$ as shown below.

A.2.3 Control of the regularization bias contribution

It follows from Lemmas B.3 and B.4 in Section A.2.5 that

$$\xi_{5,T}^{B} = Q_{\lambda_T}\left(1 + o_p(1)\right) + O_p\left(\frac{1}{\sqrt{T}} Q_{\lambda_T}^{1/2}\right) + o_p((T h_T^{1/2})^{-1}).$$

Rewriting $\mathcal{B}_T = -\lambda_T \left(\lambda_T + A^* A\right)^{-1} \varphi_0$ and developing φ_0 w.r.t. the basis of eigenfunctions ϕ_j of $A^* A$, we have for $0 \leq \beta \leq 1$

$$\begin{aligned} Q_{\lambda_T} &= \|A \mathcal{B}_T\|_{L^2(\mathcal{Z})}^2 = \lambda_T^2 \sum_{j=1}^{\infty} \frac{\nu_j \left\langle \varphi_0, \phi_j \right\rangle_H^2}{\left(\lambda_T + \nu_j\right)^2} = \lambda_T^{1+\beta} \sum_{j=1}^{\infty} \frac{\lambda_T^{1-\beta} \nu_j^{1+\beta}}{\left(\lambda_T + \nu_j\right)^2} \frac{\left\langle \varphi_0, \phi_j \right\rangle_H^2}{\nu_j^{\beta}} \\ &\leq \lambda_T^{1+\beta} \sum_{j=1}^{\infty} \frac{\left\langle \varphi_0, \phi_j \right\rangle_H^2}{\nu_j^{\beta}} \leq C \lambda_T^{1+\beta}, \end{aligned} \tag{13}$$

from the source condition $\varphi_0 \in \Phi_\beta$ (see also CFR, proof of Proposition 3.11). Using Assumption 4(i), we get $\xi^B_{5,T} = o_p((Th_T^{1/2})^{-1})$.

A.2.4 Control of the estimation error contribution

We use $r = A\varphi_0$ under H_0 and $\hat{r} = \hat{\psi} + \widehat{A}\varphi_0 + \hat{q}$, where $\hat{\psi}(z) := \int (y - \varphi_0(x)) \frac{\hat{f}(w,z)}{f(z)} dw$ and $\hat{q}(z) = \int (y - \varphi_0(x)) \left[\hat{f}(w|z) - \frac{\hat{f}(w,z)}{f(z)}\right] dw = -\hat{\psi}(z) \frac{\hat{f}(z) - f(z)}{\hat{f}(z)}$. Then, the estimation error $\hat{\varphi} - \varphi_{\lambda_T}$ is decomposed as:

$$\begin{aligned} \hat{\varphi} - \varphi_{\lambda_T} &= \left(\lambda_T + \hat{A}^*\hat{A}\right)^{-1} A^*\hat{\psi} - \left(\lambda_T + \hat{A}^*\hat{A}\right)^{-1} \left(\hat{A}^*\hat{A} - A^*A\right) \mathcal{B}_T \\ &\quad + \left(\lambda_T + \hat{A}^*\hat{A}\right)^{-1} A^*\hat{q} + \left(\lambda_T + \hat{A}^*\hat{A}\right)^{-1} (\hat{A}^* - A^*) \left(\hat{q} + \hat{\psi}\right). \end{aligned} \tag{14}$$

The first two terms in the RHS are of first-order, the corresponding quantities converging to zero are $\hat{\psi}$ and $\left(\hat{A}^*\hat{A} - A^*A\right) \mathcal{B}_T$, respectively. The third and fourth terms are at least of second-order. To eliminate the estimate $\hat{A}^*\hat{A}$ in the inverse $\left(\lambda_T + \hat{A}^*\hat{A}\right)^{-1}$, we can use iteratively:

$$\left(\lambda_T + \hat{A}^*\hat{A}\right)^{-1} = (\lambda_T + A^*A)^{-1} - \left(\lambda_T + \hat{A}^*\hat{A}\right)^{-1} \left(\hat{A}^*\hat{A} - A^*A\right) (\lambda_T + A^*A)^{-1}.$$

Then Equation (14) is transformed into a development of $\hat{\varphi} - \varphi_{\lambda_T}$ in a series of terms of different orders:

$$\hat{\varphi} - \varphi_{\lambda_T} = \mathcal{E}_{T,1} + \mathcal{E}_{T,2} + \mathcal{R}_T,$$

where $\mathcal{E}_{T,1} = (\lambda_T + A^*A)^{-1} A^*\hat{\psi}$ and $\mathcal{E}_{T,2} = (\lambda_T + A^*A)^{-1} \left(\hat{A}^*\hat{A} - A^*A\right) \mathcal{B}_T$ are the first-order terms, and $\mathcal{R}_T$ contains second-, third-, etc-order terms. Thus, the estimation error contribution can be decomposed as:

$$\begin{aligned}\xi_{5,T}^{E} &= \mathcal{K}_T\left(\mathcal{E}_{T,1}(X),\mathcal{E}_{T,1}(X)\right)+\mathcal{K}_T\left(\mathcal{E}_{T,2}(X),\mathcal{E}_{T,2}(X)\right)+2\mathcal{K}_T\left(\mathcal{E}_{T,2}(X),\mathcal{E}_{T,1}(X)\right)\\ &\quad -2\mathcal{K}_T\left(U-\mathcal{B}_T(X),\mathcal{E}_{T,1}(X)\right)-2\mathcal{K}_T\left(U-\mathcal{B}_T(X),\mathcal{E}_{T,2}(X)\right)\\ &\quad +\mathcal{K}_T\left(\mathcal{R}_T(X),\mathcal{R}_T(X)\right)-2\mathcal{K}_T\left(\mathcal{R}_T(X),U-\mathcal{B}_T(X)-\mathcal{E}_{T,1}(X)-\mathcal{E}_{T,2}(X)\right).\end{aligned}$$

It follows from Lemmas B.5-B.8 and (13) that $\xi_{5,T}^{E}=o_p((Th_T^{1/2})^{-1})$.

A.2.5 Technical Lemmas

Lemmas B.1 and B.2 are akin to Lemmas A.2 and A.4 in TK. The major technical novelties are in proving Lemmas B.3-B.8 where we use conditions on the decay of the spectrum (Assumption A.7) and on the regularization parameter (Assumption 4). To minimize the complexity of the presentation we assume $V_0(z)=\Omega_0(z)=1$ in some steps in the proofs of Lemmas B.5-B.6. All proofs are given under Assumptions A.1-A.8 and H_0.

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

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

Lemma B.3: $\mathcal{K}_T\left(\mathcal{B}_T(X),\mathcal{B}_T(X)\right)=Q_{\lambda_T}\left(1+o_p(1)\right)+o_p((Th_T^{1/2})^{-1})$.

Lemma B.4: $\mathcal{K}_T\left(U,\mathcal{B}_T(X)\right)=O_p\left(\frac{1}{\sqrt{T}}Q_{\lambda_T}^{1/2}\right)+o_p((Th_T^{1/2})^{-1})$.

Lemma B.5: $\mathcal{K}_T\left(\mathcal{E}_{T,k}(X),\mathcal{E}_{T,l}(X)\right)=o_p((Th_T^{1/2})^{-1})$, $k,l=1,2$.

Lemma B.6: $\mathcal{K}_T\left(U-\mathcal{B}_T(X),\mathcal{E}_{T,k}(X)\right)=o_p\left(\frac{1}{\sqrt{Th_T^{1/2}}}Q_{\lambda_T}^{1/2}\right)+o_p((Th_T^{1/2})^{-1})$, $k=1,2$.

Lemma B.7: $\mathcal{K}_T\left(\mathcal{R}_T(X), \mathcal{R}_T(X)\right) = o_p((Th_T^{1/2})^{-1})$.

Lemma B.8: $\mathcal{K}_T\left(\mathcal{R}_T(X), U - \mathcal{B}_T(X) - \mathcal{E}_{T,1}(X) - \mathcal{E}_{T,2}(X)\right) = o_p((Th_T^{1/2})^{-1})$.

Appendix 3: Proof of Proposition 2

In A.3.1 we show that a decomposition similar to (11) holds under H_1: some differences appear in the order of the negligible terms. In A.3.2 we separate the leading term into an asymptotically distributed term and a diverging term under H_1. We gather the technical lemmas in A.3.3 and discuss differences w.r.t. those used in the proof of Proposition 1.

A.3.1 Asymptotically equivalent statistics

From the arguments in A.2.1, statistics $Q_T(\hat{\varphi})$ and ξ_T are asymptotically equivalent. Let us study the behaviour of $\xi_{i,T}$, $i = 1, ..., 5$, under H_1, and split as in (12)

$$Y - \hat{\varphi}(X) = U^* - \upsilon - \mathcal{E}_T(X). \tag{15}$$

The innovation $U^* := Y - r(Z)$ satisfies $E[U^*|Z] = 0$. The error $\upsilon := \varphi_{\lambda_T}(X) - r(Z)$ satisfies $E[\upsilon|Z] = -M_{\lambda_T} r(Z)$, that is minus the Tikhonov residual. We get $\xi_{1,T}$ is $O_p((Th_T)^{-2})$ from Lemma C.1, while $\xi_{3,T}$ and $\xi_{4,T}$ are $O_p(T^{-2}h_T^{-3/2}\tau_T) + O_p((Th_T^{1/2})^{-1})$ from Lemma C.2, which yields that under H_1 and $\tau_T \to \infty$

$$\xi_T - \xi_{2,T} = \xi_{5,T} + o_p((Th_T^{1/2})^{-1}\tau_T). \tag{16}$$

Let us investigate the behavior of $\xi_{5,T}$ by introducing a decomposition based on (15): $\xi_{5,T} = \bar{\xi}^{*}_{5,T} + \bar{\xi}^{*,\upsilon}_{5,T} + \xi^{*,E}_{5,T}$, where $\bar{\xi}^{*}_{5,T} = \mathcal{K}_T\left(U^*, U^*\right)$, $\bar{\xi}^{*,\upsilon}_{5,T} = \mathcal{K}_T\left(\upsilon, \upsilon\right) - 2\mathcal{K}_T\left(U^*, \upsilon\right)$, and

$$\xi^{*,E}_{5,T} = \mathcal{K}_T\left(\mathcal{E}_T(X), \mathcal{E}_T(X)\right) - 2\mathcal{K}_T\left(U^* - \upsilon, \mathcal{E}_T(X)\right). \tag{17}$$

A.3.2 Divergence of the test statistic

Following the same arguments as in Section A.2.2, we get $Th_T^{1/2}\bar{\xi}^{*}_{5,T} \xrightarrow{d} N(0, \sigma^{*2})$, where

$$\sigma^{*2} = 2K_{**}\int_{S^*}\left\{\Omega_0(z)V_0(z)\right\}^2 dz. \tag{18}$$

Then Proposition 2 follows using that $Th_T^{1/2}\bar{\xi}^{*,\upsilon}_{5,T} = \tau_T + o_p(\tau_T) + O_p(1)$ from Lemmas C.3 and C.4.

A.3.3 Technical Lemmas

Lemma C.1 is the analogue of Lemma B.1. Lemma C.2 differs from Lemma B.2 by the orders in the bound. Lemmas C.3 and C.4 concern the diverging contribution due to the noncentrality parameter $Th_T^{1/2}Q_{\lambda_T}$, and are the analogues of Lemmas B.3 and B.4. The remainder terms are $O_p((Th_T^{1/2})^{-1})$, since $E\left[\upsilon^2\right] = O(1)$ under H_1 in contrast to $E\left[\mathcal{B}_T(X)^2\right] = o(1)$ under H_0. All proofs are given under Assumptions A.1-A.8 and H_1.

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

Lemma C.2: $\xi_{3,T} = O_p\left((Th_T)^{-1}Q_{\lambda_T}\right) + O_p((Th_T^{1/2})^{-1})$.

Lemma C.3: $\mathcal{K}_T\left(\upsilon, \upsilon\right) = Q_{\lambda_T}(1 + o_p(1)) + O_p((Th_T^{1/2})^{-1})$.

Lemma C.4: $\mathcal{K}_T\left(U^*, \upsilon\right) = O_p\left(\frac{1}{\sqrt{T}}Q_{\lambda_T}^{1/2}\right) + O_p((Th_T^{1/2})^{-1})$.

Appendix 4: Proof of Proposition 3

From (10) and $\mathcal{P}r_{F_0} = 0$, we have:

$$\tau_T = Th_T^{1/2} \sum_{j=1}^{\infty} \frac{\lambda_T^2}{(\lambda_T + \nu_j)^2} \chi_j^2, \tag{19}$$

where $\chi_j = \langle r_{F_0}, \psi_j \rangle_{L^2(\mathcal{Z})}$. Let $\delta > 1$. Then we have $\chi_j^2 / \nu_j^{\delta + \bar{\beta}} \to \infty$ as $j \to \infty$. Indeed, by contradiction, if sequence $\chi_j^2 / \nu_j^{\delta+\bar{\beta}}$ were bounded, we would have $\sum_{j=1}^{\infty} \frac{\chi_j^2}{\nu_j^{1+\bar{\beta}}} = \sum_{j=1}^{\infty} \nu_j^{\delta-1} \frac{\chi_j^2}{\nu_j^{\delta+\bar{\beta}}} \leq C \sum_{j=1}^{\infty} \nu_j^{\delta-1} < \infty$, which is impossible because of $F_0 \in \bar{\mathcal{M}}_{c,\bar{\beta}}$. Now,

$$\begin{aligned}
\sum_{j=1}^{\infty} \frac{\lambda_T^2}{(\lambda_T + \nu_j)^2} \chi_j^2 &= \lambda_T^{\delta+\bar{\beta}} \sum_{j=1}^{\infty} \frac{\lambda_T^{2-\delta-\bar{\beta}} \nu_j^{\delta+\bar{\beta}}}{(\lambda_T + \nu_j)^2} \frac{\chi_j^2}{\nu_j^{\delta+\bar{\beta}}} \geq \lambda_T^{\delta+\bar{\beta}} \frac{\lambda_T^{2-\delta-\bar{\beta}} \nu_{N(\lambda_T)}^{\delta+\bar{\beta}}}{\left(\lambda_T + \nu_{N(\lambda_T)}\right)^2} \frac{\chi_{N(\lambda_T)}^2}{\nu_{N(\lambda_T)}^{\delta+\bar{\beta}}} \\
&\gtrsim \lambda_T^{\delta+\bar{\beta}} \frac{\chi_{N(\lambda_T)}^2}{\nu_{N(\lambda_T)}^{\delta+\bar{\beta}}} \geq C \lambda_T^{\delta+\bar{\beta}},
\end{aligned}$$

for any constant C and large T, where $N(\lambda) \asymp \frac{1}{\alpha} \log\left(\frac{\tilde{C}}{\lambda}\right)$ is such that $\nu_{N(\lambda)} \asymp \lambda$. From (19) we conclude that $\tau_T \geq CTh_T^{1/2} \lambda_T^{\delta+\bar{\beta}}$, and this yields the statement of Proposition 3 because of Assumption 4(iii).

Appendix 5: Proof of Proposition 5

In A.5.1 we prove the asymptotic normality of the test statistic under H_0. In A.5.2 we derive the asymptotic behavior under global alternatives. We gather the technical lemmas in A.5.3, and compare them with the ones of TK and of the previous sections.

A.5.1 Asymptotic distribution under the null hypothesis

A.9: *The conditional variance function* $V_0(z) := V_0\left[Y - \varphi_0(X)|Z = z\right]$ *is in class* $C^2(\mathbb{R})$, *such that* $\inf_{S_*} V_0 > 0$ *under* H_0.

A.10: *Estimator* $\bar{\varphi}$ *is such that* $\sup_{z \in S_*} \frac{1}{Th_T} \sum_t K\left(\frac{z - Z_t}{h_T}\right) \left[\bar{\varphi}(X_t) - \varphi_0(X_t)\right]^2 = o_p\left(T^{-\varepsilon}\right)$ *under* H_0, *for some* $\varepsilon > \frac{1}{3} + \frac{2}{m}$.

Assumptions A.9 and A.10 concern the conditional variance $V_0(z)$, and the first-step estimator $\bar{\varphi}$ in the estimator $\hat{V}(z)$ of $V_0(z)$, respectively. These assumptions are used in Lemma D.1 to prove the convergence of $\hat{V}(z)$, its inverse $\hat{V}(z)^{-1}$, and $\hat{H}(z)^{-1}$ under H_0, where $\hat{H}(z) := \hat{V}(z)\hat{f}(z)^2$. Then, using $\frac{\log T}{Th_T} \sup_{z \in S_*} \left|\hat{H}(z)^{-1} - H_0(z)^{-1}\right| = o_p\left(1/\left(Th_T^{1/2}\right)\right)$ from Lemma D.1 (iii) with $\varepsilon > \frac{1}{3} + \frac{2}{m}$ (Assumption A.10) and $h_T = \bar{c}T^{-\bar{\eta}}$ with $2/9 < \bar{\eta} < \min\{1 - 4/m, 1/3\}$, the proof of $\zeta_T \xrightarrow{d} N(0,1)$ is unchanged compared to Appendix 2.

A.5.2 Asymptotic behaviour under global alternatives

Let $\bar{\varphi}_{\lambda_T}$ denote the population counterpart of the pilot estimator $\bar{\varphi}$. Let $V_{\lambda_T}(z) := E\left[\left(Y - \bar{\varphi}_{\lambda_T}(X)\right)^2 |Z = z\right]$ and $\Omega_{\lambda_T}(z) := V_{\lambda_T}(z)^{-1}$. Let $L^2_{\lambda_T}(\mathcal{Z})$ denote the L^2-space associated with measure $\Pi_{\mathcal{Z},\lambda_T}(dz) = \Omega_{\lambda_T}(z) I\left\{z \in S^*\right\} F_Z(dz)$, and $A^*_{\lambda_T}$ the corresponding adjoint of operator A. Denote by $\left\{\phi_{\lambda_T,j}, \psi_{\lambda_T,j}, \omega_{\lambda_T,j}; j = 1, 2, ...\right\}$ the singular value decomposition of operator A w.r.t. the $H^2[0,1]$ and $L^2_{\lambda_T}(\mathcal{Z})$ norms.

A.11: *There exist constants* $c_1 \leq c_2$ *such that* $c_1\Omega_0(z) \leq \Omega_{\lambda_T}(z) \leq c_2\Omega_0(z)$, *for any* $z \in S^*$ *and large* T.

A.12: *Under* H_1: $\left|\langle r, \psi_{\lambda_T,j}\rangle_{L^2_{\lambda_T}(\mathcal{Z})}\right| \geq C\left|\langle r, \psi_j\rangle_{L^2(\mathcal{Z})}\right|$ *for large* j *and* T, *and a constant* C.

A.13: *Under* H_1*:* $\sup_{z \in S^*} \left| \hat{V}(z) - V_{\lambda_T}(z) \right| = O_p\left(\sqrt{\frac{\log T}{T h_T}} + h_T^2 \right) + o_p\left(T^{-1/6}\right).$

Assumptions A.11 and A.12 concern the L^2-norm in $L^2_{\lambda_T}(\mathcal{Z})$. Specifically, Assumption A.11 implies that the L^2-norms in $L^2_{\lambda_T}(\mathcal{Z})$ and $L^2(\mathcal{Z})$ are equivalent. Assumption A.12 requires that the coefficients of r w.r.t. $\{\psi_{\lambda_T,j}; j = 1, 2, ...\}$ are uniformly bounded from below by the coefficients w.r.t. $\{\psi_j; j = 1, 2, ...\}$. Finally Assumption A.13 yields the uniform convergence of $\hat{V}(z)$ to $V_{\lambda_T}(z)$ and is the analogue of the property proved in Lemma D.1 (i) under H_0.

The asymptotic behavior of ζ_T is derived along the lines of Section 3.3 using Lemmas D.2 and D.3. The population counterpart of the TiR estimator $\hat{\varphi}$ in (7) is $\varphi_{\lambda_T} = \left(\lambda_T + A^*_{\lambda_T} A\right)^{-1} A^*_{\lambda_T} r$. The population counterpart of the minimized criterion value $Q_T(\hat{\varphi})$ is $Q_{\lambda_T} := E_0\left[\Omega_{\lambda_T}(Z) I\{Z \in S^*\} [M_{\lambda_T} r(Z)]^2\right] = \|M_{\lambda_T} r\|^2_{L^2_{\lambda_T}(\mathcal{Z})}$, where $M_{\lambda_T} r := \left[1 - A\left(\lambda_T + A^*_{\lambda_T} A\right)^{-1} A^*_{\lambda_T}\right] r$, and $\tau_T := T h_T^{1/2} Q_{\lambda_T}$. From Lemma D.2 and Assumption 5, we deduce that $|\zeta_T| \geq C \tau_T$ for some constant C. As in Equation (10):

$$\tau_T = T h_T^{1/2} \sum_{j=1}^{\infty} \frac{\lambda_T^2}{\left(\lambda_T + \nu_{\lambda_T,j}\right)^2} \left\langle r, \psi_{\lambda_T,j} \right\rangle^2_{L^2_{\lambda_T}(\mathcal{Z})} + T h_T^{1/2} \left\| \mathcal{P}_{\lambda_T} r \right\|^2_{L^2_{\lambda_T}(\mathcal{Z})}, \tag{20}$$

where $\mathcal{P}_{\lambda_T}$ denotes the orthogonal projection operator on the linear space $\ker\left(A^*_{\lambda_T}\right)$ w.r.t. the scalar product in $L^2_{\lambda_T}(\mathcal{Z})$. Let us now characterize the divergence rate of τ_T.

Consider first close misspecifications. Then, $r \in \ker(A^*)^\perp$ and $r \in \ker\left(A^*_{\lambda_T}\right)^\perp$ from Lemma D.3. Thus $\mathcal{P}_{\lambda_T} r = 0$. Morever, $c_1 A^* A \leq A^*_{\lambda_T} A \leq c_2 A^* A$ from Assumption A.11, and thus $c_1 \nu_j \leq \nu_{\lambda_T,j} \leq c_2 \nu_j$. Using Assumption A.12, $\tau_T \geq C T h_T^{1/2} \sum_{j=1}^{\infty} \frac{\lambda_T^2}{\left(\lambda_T + \nu_j\right)^2} \langle r, \psi_j \rangle^2_{L^2(\mathcal{Z})}$ from (20). Then, from the arguments in Appendix 4, we get $\tau_T \geq C T h_T^{1/2} \lambda_T^{\delta + \bar{\beta}}$, for any

$\delta > 1$.

Consider now separated misspecifications. Then, $r \notin \ker(A^*)^{\perp} = \ker\left(A^*_{\lambda_T}\right)^{\perp}$, and $\|\mathcal{P}_{\lambda_T} r\|_{L^2_{\lambda_T}(\mathcal{Z})} \geq C$ from Assumption A.11. Thus we have $\tau_T \geq T h_T^{1/2} C$ from (20), and the conclusion follows.

A.5.3 Technical Lemmas

Lemma D.1 is akin to Lemmas C.2 and C.3 (i) in TK. Lemma D.2 extends Proposition 2 for an estimated weighting function. Since $V_{\lambda_T}(z)$ might not converge, we cannot expect an asymptotically normal distribution for term $\bar{\xi}^*_{5,T}$ as in A.3.2. However, its contribution is still $O_p(1)$. Lemma D.3 shows that the orthogonal complement of the null space of the ajoint of A is the same under the two norms $L^2_{\lambda_T}(\mathcal{Z})$ and $L^2(\mathcal{Z})$.

Lemma D.1: *Under H_0 and A.1-A.10*,

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

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

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

Lemma D.2: *Under H_1, Assumptions 1-3, 4(ii), A.1-A.8 and A.11-A.13, if $\tau_T \to \infty$ as $T \to \infty$ we have $\sigma \zeta_T = T h_T^{1/2} \xi^{*,E}_{5,T} + o_p(\tau_T) + O_p(1)$, where $\xi^{*,E}_{5,T}$ is defined as in* (17) *replacing Ω_t by $\hat{\Omega}(Z_t)$.*

Lemma D.3: *We have $\ker\left(A^*_{\lambda_T}\right)^{\perp} = \ker(A^*)^{\perp}$, where the orthogonal complements are w.r.t. the scalar product in $L^2_{\lambda_T}(\mathcal{Z})$ in the LHS, and $L^2(\mathcal{Z})$ in the RHS, respectively.*

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.*

Chernozhukov, V., Gagliardini, P. and O. Scaillet (2007): "Nonparametric Instrumental Variable Estimation of Quantile Structural Effects", Working Paper.

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 Nonparametric Instrumental variable 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.

Horowitz, J., and V., Spokoiny (2001): "An Adaptive, Rate-Optimal Test of a Parametric Mean-Regression Model Against a Nonparametric Alternative", Econometrica, 69, 599-631.

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.