3sc+ model STAPRO : 4757 Prod.Type:FTP pp.1–8(col.fig.:NIL) ED:LeenaPrakash PAGN:Anvar SCAN:

ARTICLE IN PRESS

Available online at www.sciencedirect.com

ScienceDirect

Statistics & Probability Letters ▮ (▮▮▮▮) ▮▮▮–▮▮▮

STATISTICS & PROBABILITY LETTERS

www.elsevier.com/locate/stapro

Testing parametric models in the presence of instrumental variables

Hajo Holzmann

Institut für Stochastik, Universität Karlsruhe, Germany

Received 20 February 2006; received in revised form 18 June 2007; accepted 7 September 2007

Abstract

In this note we develop tests for checking the hypothesis whether a regression function, which is identified via instrumental variables, belongs to some parametric family of functions. We show that this testing problem can essentially be reduced to testing whether a regression function in an ordinary nonparametric regression model vanishes. The asymptotic distribution theory of two test statistics is developed, and their power properties are investigated in a simulation study.

Keywords: Instrumental variables; Inverse estimation; Lack-of-fit tests; Nonparametric regression; Statistical inverse problem

1. Introduction

Instrumental variable (IV) techniques are routinely used for estimating the structural parameters in econometric models. In parametric regression models, there is a variety of $\sqrt{n}$-consistent estimation methods, including two-stage least-squares (2SLS) estimation (Basmann, 1959; Theil, 1953) and limited information Q1 maximum likelihood (Anderson and Rubin, 1949, 1950).

Recently, there has also been increasing interest for nonparametric methods of estimation in regression models with IVs, see Blundell et al. (2007), Carrasco et al. (2007), Darolles et al. (2006), Hall and Horowitz (2005) or Newey and Powell (2003). To be more specific, let us introduce the model. Suppose that (Y_k, X_k, Z_k), $Z_k = (Z_{k,1}, Z_{k,2})$, $k = 1, \ldots, N$, are i.i.d. observations from the regression model

$$Y_k = \phi(X_k, Z_{k,1}) + U_k, \quad E(U_k|Z_k) = 0, \tag{1}$$

cf. Newey and Powell (2003). Here ϕ is the unknown regression function of interest, the Y_k are real-valued, $X_k \in \mathbb{R}^q$, $Z_k = (Z_{k,1}, Z_{k,2}) \in \mathbb{R}^{p_1+p_2}$ are called the instruments, and the U_k are unobservable real-valued errors. Note that in this general model we allow for joint components between the covariates $(X_k, Z_{k,1})$ and the instruments $(Z_{k,1}, Z_{k,2})$.

Nonparametric estimation in model (1) involves solving an integral equation of the first kind, which leads to the theory of ill-posed inverse problems (Kress, 1999). In fact, let μ_1 and μ_2 denote the distributions of

E-mail addresses: holzmann@stoch.uni-karlsruhe.de, Hajo.Holzmann@stoch.uni-karlsruhe.de (H. Holzmann).

0167-7152/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.spl.2007.09.025

Please cite this article as: Holzmann, H., Testing parametric models in the presence of instrumental variables. Statist. Probab. Lett. (2007), doi:10.1016/j.spl.2007.09.025

$(X_k, Z_{k,1})$ and $(Z_{k,1}, Z_{k,2})$. Introduce the operator

$$A : L_2(\mu_1) \to L_2(\mu_2), \quad (A\psi)(z) = E(\psi(X_1, Z_{1,1})|Z_1 = z),$$

and let $E(Y_1|Z_1 = z) =: r(z)$. If the operator A is injective, which we will assume from now on, the function ϕ in model (1) is determined uniquely as the solution of

$$A\phi = r. \tag{2}$$

Thus estimation in (1) typically proceeds by estimating A and r, inserting these estimates into (2) and then solving this equation by some regularization method. However, as is well known from the general theory of statistical inverse problems (cf. Mair and Ruymgaart, 1996), such estimates typically have slow convergence rates.

Therefore it might still be desirable to use a parametric model, after checking its validity via a lack-of-fit test. More specifically, one should test the hypothesis that the regression function ϕ belongs to a parametric family of functions of interest,

$$\mathrm{H} : \phi \in (\phi_\theta)_{\theta \in \Theta}, \quad \Theta \subset \mathbb{R}^l. \tag{3}$$

Since the operator A is assumed to be injective, one could equivalently test the hypothesis

$$\mathrm{H}' : (A\phi) \in ((A\phi_\theta))_{\theta \in \Theta}, \tag{4}$$

and by further preconditioning with A^*,

$$\mathrm{H}'' : (A^*A\phi) \in ((A^*A\phi_\theta))_{\theta \in \Theta}, \tag{5}$$

since A^*A is also injective.

A natural approach is to test (3) directly. Typically this involves comparing a parametric with a nonparametric fit, and therefore requires a sophisticated nonparametric estimation of ϕ. In Holzmann et al. (2006), for the related statistical inverse problem of deconvolution density estimation, indirect testing procedures based on H are investigated in detail. However, in the IV regression model, where the operator A has to be estimated as well, this seems to be rather impractical.

Therefore, in this paper we concentrate on tests based on H'. We show that testing H' can essentially be reduced to testing whether a regression function in an ordinary nonparametric regression model vanishes. Thus, one can apply testing procedures suggested for direct regression models. A variety of methods are
Q2 available (see e.g. Härdle and Mammen, 1993; Spokoiny, 1996; Stute, 1997; Dette, 199[illegible]t-Sahalia et al.,
Q3 2001), which might be adapted to our context. We will focus on the test of Aït-Sahalia et al. (2001), but other tests might be used as well.

Tests based on H'' must be designed specifically for IV regression. They are motivated by the fact that preconditioning with A^* is often useful when estimating ϕ, since the operator A^*A is self-adjoint and therefore one can apply spectral regularization methods to approximate its inverse. For the purpose of comparison we will briefly consider one such test based on H''; the test suggested by Horowitz (2006) also appears to fall into this category. In the scenarios considered in a simulation study, it turned out that the adapted test of Aït-Sahalia et al. (2001) is at least as powerful as the suggested method based on testing H''.

The paper is organized as follows. In Section 2 we introduce the relevant test statistics and discuss their asymptotic distribution theory. Section 3 investigates the performance of competing tests in a small simulation study. Some technical assumptions and arguments are given in the Appendix.

2. The test statistics

2.1. Simple hypothesis

Let us start by investigating a simple hypothesis

$$\mathrm{H} : \phi = \phi_0, \tag{6}$$

for some given function ϕ_0. For (6), the hypothesis (4) takes the form $\mathrm{H}' : A\phi_0 = r$, or equivalently

Please cite this article as: Holzmann, H., Testing parametric models in the presence of instrumental variables. Statist. Probab. Lett. (2007), doi:10.1016/j.spl.2007.09.025

$$E(\tilde{U}_k | Z_k = z) = 0, \quad \tilde{U}_k = Y_k - \phi_0(X_k, Z_{k,1}). \tag{7}$$

Note that $\tilde{U}_k = Y_k - \phi_0(X_k, Z_{k,1})$ is completely known. Therefore, all tests for checking whether a given regression function, namely that of $\tilde{U}_k$ on Z_k, vanishes, can be applied. Let us describe the test introduced by Aït-Sahalia et al. (2001), adapted to our setting. Let K be a symmetric kernel on $\mathbb{R}$ and denote by $K_p(x) = \Pi_{k=1}^n K(x_k)$ the p-fold product kernel on $\mathbb{R}^p$, $p = p_1 + p_2$. Further let $K_{p,h}(x) = K_p(x/h)/h^p$ for the bandwidth parameter $h > 0$. We estimate r and A by the Nadaraya–Watson kernel estimators,

$$\hat{r}(z) = \frac{\sum_{k=1}^N Y_k K_{p,h}(z - Z_k)}{\sum_{k=1}^N K_{p,h}(z - Z_k)}, \quad \hat{A}\phi(z) = \frac{\sum_{k=1}^N \phi(X_k, Z_{k,1}) K_{p,h}(z - Z_k)}{\sum_{k=1}^N K_{p,h}(z - Z_k)},$$

where we suppress the dependence on N and h in the notation. This gives rise to the test statistic considered by Aït-Sahalia et al. (2001) (see also Härdle and Mammen, 1993),

$$T_N^{\mathrm{ABS}} = \frac{1}{N} \sum_{k=1}^N (\hat{r}(Z_k) - (\hat{A}\phi_0)(Z_k))^2 a(Z_k),$$

where a is a compactly supported weight function. The following theorem can be easily deduced from their results.

Theorem 1. *Under Assumptions* 1–4 *given in* Appendix A, *we have under the hypothesis* (6) *that*

$$N h^{p/2} T_N^{\mathrm{ABS}} - h^{-p/2} \gamma \xrightarrow{\mathscr{L}} \mathrm{N}(0, \sigma^2),$$

where

$$\gamma = C_1 \int \sigma^2(z) a(z)\, \mathrm{d}z \quad \textit{and} \quad \sigma^2 = 2C_2 \int \sigma^4(z) a^2(z)\, \mathrm{d}z,$$

$$C_1 = \int K_p^2(z)\, \mathrm{d}z = K_p^{(2)}(0), \quad C_2 = \int (K_p^{(2)}(z))^2\, \mathrm{d}z = K_p^{(4)}(0),$$

and $K_p^{(j)}$ is the jth convolution product of K_p with itself, f_Z the marginal density of Z_1 and $\sigma^2(z) = E(U_1^2 | Z_1 = z)$ denotes the conditional variance function.

Remark 1. Note that in order to apply Theorem 1, the constants γ and σ have to be estimated from the data. Aït-Sahalia et al. (2001) give suggestions, but a variety of estimators for the conditional variance function are available, especially in a homoscedastic design (cf. e.g. Hall and Marron, 1990).

Remark 2. Aït-Sahalia et al. (2001) show that their test can detect linear local alternatives at a distance of $N^{-1/2} h^{-p/4}$ from the null, and it is easily seen that this fact also extends to our present situation, where the distance is taken from the function ϕ_0. In contrast, Horowitz' (2006) test can even detect linear alternatives at a distance of $N^{-1/2}$. We could also use a $N^{-1/2}$ consistent test for testing (7), e.g. that introduced by Stute (1998). However, such tests may have small power against non-linear local alternatives or smooth classes of alternatives, cf. Spokoiny (1996) and Bachmann and Dette (2005), and thus need not have good power properties in general.

Now let us briefly consider a statistic for testing the simple hypothesis (6) via H″, i.e. by comparing $A^* r$ to $A^* A \phi_0$. Here we will assume that the covariate is only given by X_k, and that the instrument is $Z_k = Z_{k,2}$, so that the covariate and the instrument do not have a joint component. We estimate the adjoint operator A^* by

$$\hat{A}^*(\psi)(x) = \frac{\sum_{k=1}^N \psi(Z_k) K_{q,h}(x - X_k)}{\sum_{k=1}^N K_{q,h}(x - X_k)}.$$

Denote by $\hat{f}_{h,X}(x)$, $\hat{f}_{h,Z}(Z_k)$ and $\hat{f}_{h,XZ}(x, Z_k)$ kernel density estimates of the densities f_X, f_Z and f_{XZ}, respectively, i.e. $\hat{f}_{h,X}(x) = N^{-1} \sum_{k=1}^N K_{q,h}(x - X_k)$, and similarly for the other density estimates. Then we obtain estimates of $A^* r$ and of $A^* A \phi_0$ as follows:

Please cite this article as: Holzmann, H., Testing parametric models in the presence of instrumental variables. Statist. Probab. Lett. (2007), doi:10.1016/j.spl.2007.09.025

$$\hat{A}^*\hat{r}(x) = \frac{1}{N}\sum_{k=1}^{N} Y_k \frac{\hat{f}_{h,XZ}(x,Z_k)}{\hat{f}_{h,X}(x)\hat{f}_{h,Z}(Z_k)}, \quad \hat{A}^*\hat{A}\phi_0(x) = \frac{1}{N}\sum_{k=1}^{N} \phi_0(X_k)\frac{\hat{f}_{h,XZ}(x,Z_k)}{\hat{f}_{h,X}(x)\hat{f}_{h,Z}(Z_k)},$$

using a straightforward computation. Assume that the model is homoscedastic, i.e. that $E(U_1^2|Z_1) = \sigma^2$ is constant. Darolles et al. (2006) show that under the regularity conditions listed in their Appendix B, under the hypothesis (6) one has that

$$\sqrt{N}(\hat{A}^*\hat{r} - \hat{A}^*\hat{A}\phi_0) \xrightarrow{\mathscr{L}} \mathrm{N}(0, \sigma^2 A^* A), \tag{8}$$

where convergence is in the sense of functional convergence in distribution in the Hilbert space $L_2(\mu_1)$ (cf. van der Vaart and Wellner, 1996), and $\mathrm{N}(0, \sigma^2 A^* A)$ is the normal distribution on $L_2(\mu_1)$ with mean vector 0 and covariance operator $\sigma^2 A^* A$. Since the squared norm is a continuous function on $L_2(\mu_1)$, (8) immediately implies that

$$\mathrm{N}\|\hat{A}^*\hat{r} - \hat{A}^*\hat{A}\phi_0\|^2_{L_2(\mu_1)} \xrightarrow{\mathscr{L}} \sigma^2 \sum_{k\geqslant 1} \lambda_k \chi^2_{k,1}, \tag{9}$$

where the λ_k are the eigenvalues of the operator A^*A and $\chi^2_{k,1}$ are independent chi-squared random variables with one degree of freedom. The statistic in (9) is not yet practically feasible, since the norm depends on the unknown density f_X. However, it can be replaced by a kernel estimate to obtain

$$S_{N,1} = \int (\hat{A}^*\hat{r}(x) - \hat{A}^*\hat{A}\phi_0(x))^2 \hat{f}_{h,X}(x)\,\mathrm{d}x,$$

or simply by an average

$$S_{N,2} = \frac{1}{N^2}\sum_{j=1}^{N}\left(\sum_{k=1}^{N}(Y_k - \phi_0(X_k))\frac{\hat{f}_{h,XZ}(X_j,Z_k)}{\hat{f}_{h,X}(X_j)\hat{f}_{h,Z}(Z_k)}\right)^2.$$

Under sufficient regularity (in particular uniform convergence of the kernel estimator $\hat{f}_{h,X}$ to f_X), the asymptotics in (9) still hold true for $S_{N,1}$ and $S_{N,2}$. However, it is not easy to estimate the critical value on this basis, and bootstrapping may become necessary. We do not pursue this further here, since we only use $S_{N,2}$ for illustration purposes in the simulations section.

2.2. Composite hypothesis

In this section we briefly discuss testing a composite hypothesis by using a modification of the statistics T_N^{ABS}.

Suppose that (3) holds for some $\phi_0 = \phi_{\theta_0}$. In this case, typically θ_0 can be estimated $\sqrt{N}$-consistently, e.g. by a 2SLS estimator $\hat{\theta}$ (cf. Basmann, 1959). In order to test the composite hypothesis (3), in the statistic T_N^{ABS} we replace the unknown ϕ_0 by the estimated $\phi_{\hat{\theta}}$, i.e. we use

$$T_{N,\hat{\theta}}^{\mathrm{ABS}} = \frac{1}{N}\sum_{k=1}^{N}(\hat{r}(Z_k) - (\hat{A}\phi_{\hat{\theta}})(Z_k))^2 a(Z_k),$$

The following theorem shows that if the parameter θ_0 can be estimated sufficiently fast, the statistic $T_{N,\hat{\theta}}^{\mathrm{ABS}}$ asymptotically behaves as in case of a simple hypothesis.

Theorem 2. *Under Assumptions* 1–6 *given in* Appendix A, *we have under the hypothesis* (3) *that*

$$Nh^{p/2}(T_{N,\hat{\theta}}^{\mathrm{ABS}} - T_N^{\mathrm{ABS}}) = o_P(1),$$

and consequently, the conclusions of Theorem 1 *remain valid for the statistic* $T_{N,\hat{\theta}}^{\mathrm{ABS}}$.

Please cite this article as: Holzmann, H., Testing parametric models in the presence of instrumental variables. Statist. Probab. Lett. (2007), doi:10.1016/j.spl.2007.09.025

This result cannot be deduced from Aït-Sahalia et al. (2001) since instead of the covariate Z_k the dependent variable $Y_k - \phi_{\hat{\theta}}(X_k, Z_{k,1})$ is influenced by the estimated parameter $\hat{\theta}$. We give the proof of Theorem 2 in the Appendix.

3. Simulation study

In this section we report on the findings of a small simulation study. First we investigate the quality of the normal approximations of T_N^{ABS} and $T_{N,\hat{\theta}}^{\mathrm{ABS}}$ as given in Theorems 1 and 2. Then we compare the powers of the tests based on testing H′ via the statistic T_N^{ABS} and H″ via $S_{N,2}$.

3.1. Normal approximation

The data are sampled from the model

$$Y = X^2 + U, \quad X = Z + V, \tag{10}$$

where

$$\begin{pmatrix} U \\ V \\ Z \end{pmatrix} \text{i.i.d.} \sim \mathrm{N}\left(0, \begin{pmatrix} 1 & 0.5 & 0 \\ 0.5 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right),$$

cf. Newey and Powell (2003) for a similar design. In Fig. 1 the distribution of $Nh^{1/2}T_N^{\mathrm{ABS}}$ is displayed together with the asymptotic distributions of Theorem 1 for sample sizes of $N = 100$ and 500 and bandwidths $h = 0.4$ and 0.1. The statistic was simulated $m = 10000$ times, and the weight function $a(z) = 1_{[-2,2]}(z)$ and the normal kernel were used. The approximation is reasonably close, even for the smaller sample size. When varying the bandwidth it turns out that the quality of the approximation is not very sensitive to bandwidth selection. In Fig. 2 we see corresponding simulation results for $Nh^{1/2}T_{N,\hat{\theta}}^{\mathrm{ABS}}$, where the parametric model is given by $Y = \theta_0 + \theta_1 X + \theta_2 X^2$, the instruments used in the 2SLS estimation were $(1, Z, Z^2)$. The matrix arising in the 2SLS estimation is rather ill conditioned, therefore for $N = 100$ in the second step we used a ridge parameter

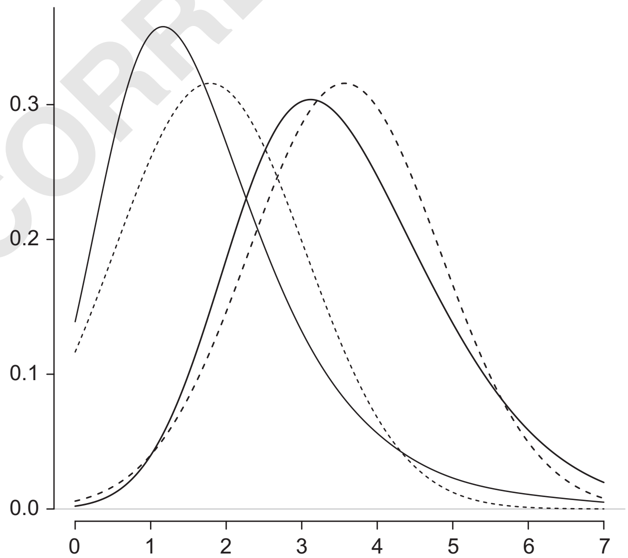

Fig. 1. Distribution of test statistic T_N^{ABS} under the simple hypothesis. Solid line for $N = 100$ and bandwidth $h = 0.4$, dashed line the asymptotic distribution. Thick lines for $N = 500$ and $h = 0.1$.

Please cite this article as: Holzmann, H., Testing parametric models in the presence of instrumental variables. Statist. Probab. Lett. (2007), doi:10.1016/j.spl.2007.09.025

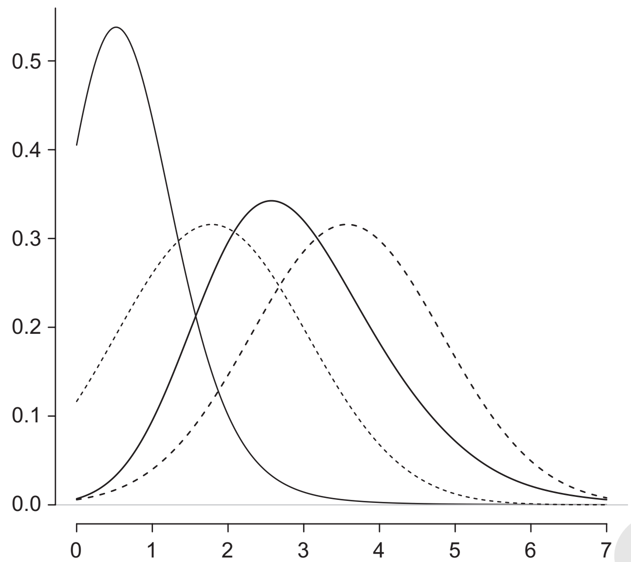

Fig. 2. Distribution of test statistic $T^{\mathrm{ABS}}_{N,\hat{\theta}}$ under the hypothesis. Solid line for $N = 100$ and bandwidth $h = 0.4$, dashed line the asymptotic distribution. Thick lines for $N = 500$ and $h = 0.1$.

Table 1
Power comparison between T^{ABS}_N and $S_{N,2}$

Alternative	N	Bandwidth	c	Power T^{ABS}_N	Power $S_{N,2}$
(A)	100	0.4	0.002	0.13	0.08
			0.004	0.30	0.14
			0.01	0.75	0.32
	200	0.3	0.002	0.17	0.07
			0.004	0.43	0.12
			0.01	0.92	0.41
	500	0.1	0.002	0.27	0.15
			0.004	0.71	0.31
			0.01	1.00	0.71
(B)	100	0.4	0.001	0.19	0.13
			0.003	0.69	0.44
	200	0.3	0.001	0.29	0.15
			0.003	0.89	0.58
	500	0.1	0.001	0.50	0.33
			0.003	0.99	0.90
(C)	100	0.4	0.02	0.09	0.10
			0.07	0.48	0.49
			0.15	0.93	0.93
	200	0.3	0.02	0.13	0.11
			0.07	0.75	0.73
			0.15	1.00	1.00
	500	0.1	0.02	0.18	0.31
			0.07	0.95	0.98

of $\alpha = 0.01$. For the composite hypothesis, the approximation is less good and the use of asymptotic critical values leads to very conservative tests. Instead a bootstrap variant of the test should be used, e.g. the wild bootstrap (cf. Härdle and Mammen, 1993).

Please cite this article as: Holzmann, H., Testing parametric models in the presence of instrumental variables. Statist. Probab. Lett. (2007), doi:10.1016/j.spl.2007.09.025

3.2. Power comparison

In this section we compare the powers of the tests based on T_N^{ABS} and on $S_{N,2}$ for a simple hypothesis via simulation. First we simulate both statistics under the simple hypothesis (10), and thus obtain precise estimates of the critical values for a test with significance level 0.95. These critical values are employed in the test decisions. As alternative models we consider

$$(A)\quad Y = X^2 + cX^5 + U, \quad c = 0.002,\ 0.004,\ 0.01,$$

$$(B)\quad Y = X^2 + cX^6 + U, \quad c = 0.001,\ 0.003,$$

$$(C)\quad Y = X^2 + c\exp(X) + U, \quad c = 0.02,\ 0.07,\ 0.15.$$

In order to estimate the power, we used $m = 10000$ repetitions in case of T_N^{ABS}, and $m = 500$ repetitions for $S_{N,2}$, which is computationally more demanding than T_N^{ABS}. The results are displayed in Table 1.The bandwidth was chosen to be equal for both statistics, since it turned out that the power of both tests is not sensitive to bandwidth selection, as long as it is not too small or too large. In the alternative model (C), both tests perform very similarly, for alternative models (A) and (B) the test based on T_N^{ABS} outperforms the test based on $S_{N,2}$.

Financial support of the DFG under grant MU 1230/8-2 is acknowledged.

Appendix A. Technical details

Assumption 1. The data (Y_k, X_k, Z_k), $k = 1, \ldots, N$, are i.i.d. with density $f(y, x, z)$.

Assumption 2. We have

1. The density $f(y, x, z)$ is $r+1$-times continuously differentiable, $r \geqslant 2$. f and its derivatives are bounded and square-integrable.
2. The weight function a has compact support.
3. The marginal density $f_Z(z)$ is bounded away from 0 on the compact support of a, and $f_{Z_1}(z_1)$ is bounded away from 0 for all z_1 with (z_1, z_2) in the support of a for some z_2.
4. $EU^4 < \infty$, and
$$\sigma^2(z) = E(U_1^2 | Z_1 = z)$$
is continuous and square-integrable.

Assumption 3. The kernel K is a bounded function on $\mathbb{R}^p$, symmetric around 0, with $\int |K| < \infty$, $\int K(z)\,\mathrm{d}z = 1$ and $\int z^j K(z) = 0$ for $1 \leqslant j < d$, $d > 3p/4$.

Assumption 4. The bandwidth h satisfies $h \sim N^{-1/\delta}$, where $2p < \delta < 2r + p/2$.

Assumption 5. For fixed (x, z_1), $\theta \to \phi_\theta(x, z_1)$ is continuously differentiable. In the Taylor expansion

$$\phi_\theta(x, z_1) = \phi_{\theta_0}(x, z_1) + \phi'_{\theta_0}(x, z_1)(\theta - \theta_0) + R(\theta_0, \theta, x, z_1),$$

where $\phi'_{\theta_0}(x, z_1)$ is the gradient (w.r.t. θ), we have that $\phi'_{\theta_0}(x, z_1)$ is uniformly bounded in both arguments and that $R(\theta_0, \theta, x, z_1) = \mathrm{o}(\|\theta - \theta_0\|)$, uniformly in (x, z_1).

Assumption 6. Under the hypothesis (3) the parametric estimate $\hat{\theta}$ satisfies

$$\|\hat{\theta} - \theta_0\| = \mathrm{o_P}(N^{-1/2} h^{-p/2}).$$

Proof of Theorem 2. We have

Please cite this article as: Holzmann, H., Testing parametric models in the presence of instrumental variables. Statist. Probab. Lett. (2007), doi:10.1016/j.spl.2007.09.025

$$
\begin{aligned}
T^{\mathrm{ABS}}_{N,\hat\theta} - T^{\mathrm{ABS}}_{N} &= 2N^{-1}\sum_{j=1}^{N}(\hat r(Z_j) - (\hat A\phi_0)(Z_j))(\hat A(\phi_0 - \phi_{\hat\theta})(Z_j))a(Z_j) \\
&\quad + N^{-1}\sum_{j=1}^{N}(\hat A(\phi_0 - \phi_{\hat\theta})(Z_j))^2 a(Z_j).
\end{aligned}
$$

From Assumption 5 and uniform convergence of the Nadaraya–Watson estimator on compact sets (cf. Härdle and Mammen, 1993),

$$
\begin{aligned}
\hat A(\phi_0 - \phi_{\hat\theta})(z) &= \frac{\sum_{k=1}^{N}(\phi_0'(X_k, Z_{k,1}) + o_P(1))K_{p,h}(z - Z_k)}{\sum_{k=1}^{N}K_{p,h}(z - Z_k)}(\theta_0 - \hat\theta) \\
&= (\psi(z) + o_P(1))(\theta_0 - \hat\theta),
\end{aligned}
$$

where $\psi(z) = E(\phi_0'(X_1, Z_{1,1})|Z_1 = z)$ is bounded by Assumptions 2.3 and 5. Therefore,

$$
N^{-1}\sum_{j=1}^{N}(\hat A(\phi_0 - \phi_{\hat\theta})(Z_j))^2 a(Z_j) = O(\|\theta_0 - \hat\theta\|^2)O_P(1) = o_P(N^{-1}h^{-p/2}).
$$

Further,

$$
\begin{aligned}
&N^{-1}\sum_{j=1}^{N}(\hat r(Z_j) - (\hat A\phi_0)(Z_j))(\hat A(\phi_0 - \phi_{\hat\theta})(Z_j))a(z)\,\mathrm{d}z \\
&= N^{-1}\sum_{j=1}^{N}\frac{\sum_{k=1}^{N}U_k K_{p,h}(Z_j - Z_k)}{\sum_{k=1}^{N}K_{p,h}(Z_j - Z_k)}(\psi(Z_j) + o_P(1))a(Z_j)(\theta_0 - \hat\theta) \\
&= O_P(N^{-1/2})O(\|\hat\theta - \theta_0\|) = o_P(N^{-1}h^{-p/2}),
\end{aligned}
$$

where the last equality is obtained by straightforward computations. □

References

Aït-Sahalia, Y., Bickel, P.J., Stoker, T.M., 2001. Goodness-of-fit tests for kernel regression with application to option implied volatilities. J. Econometrics 105, 363–412.

Anderson, T.W., Rubin, H., 1949. Estimation of the parameters of a single equation in a complete system of stochastic equations. Ann. Math. Statist. 20, 46–63.

Anderson, T.W., Rubin, H., 1950. The asymptotic properties of estimates of the parameters of a single equation in a complete system of stochastic equations. Ann. Math. Statist. 21, 570–582.

Bachmann, D., Dette, H., 2005. A note on the Bickel–Rosenblatt test in autoregressive time series. Statist. Probab. Lett. 74, 221–234.

Basmann, R.L., 1959. A generalized classical method of linear estimation of coefficients in a structural equation. Econometrica 25, 77–83.

Blundell, R., Chen, X., Kristensen, D., 2007. Semi-nonparametric IV estimation of shape invariant engel curves. Econometrica, to appear.

Carrasco, M., Florens, J.-P., Renault, E., 2007. Linear inverse problems in structural econometrics: Estimation based on spectral decomposition and regularization. In: Handbook of Econometrics, Vol. 6B.

Darolles, S., Florence, J.-P., Renault, E., 2006. Nonparametric instrumental regression. Working Paper, University of Toulouse.

Hall, P., Horowitz, J.L., 2005. Nonparametric methods for inference in the presence of instrumental variables. Ann. Statist. 33, 2904–2929.

Hall, P., Marron, J.S., 1990. On variance estimation in nonparametric regression. Biometrika 77, 415–419.

Härdle, W., Mammen, E., 1993. Comparing nonparametric versus parametric regression fits. Ann. Statist. 21, 1926–1947.

Holzmann, H., Bissantz, N., Munk, A., 2006. Density testing in a contaminated sample. J. Multivariate Anal. 98, 57–75.

Horowitz, J.L., 2006. Testing a parametric model against a nonparametric alternative with identification through instrumental variables. Econometrica 74, 521–538.

Kress, R., 1999. Linear Integral Equations, second ed. Springer, New York.

Mair, B.-A., Ruymgaart, F., 1996. Statistical inverse estimation in Hilbert scales. SIAM J. Appl. Math. 56, 1424–1444.

Newey, W.K., Powell, J.L., 2003. Instrumental variable estimation of nonparametric models. Econometrica 71, 1565–1578.

Spokoiny, V.G., 1996. Adaptive hypothesis testing using wavelets. Ann. Statist. 24, 2477–2498.

Stute, W., 1997. Nonparametric model checks for regression. Ann. Statist. 25, 613–641.

Theil, H., 1953. Repeated least squares applied to a complete equations system. The Hague: Central Planning Bureau, mimeo.

van der Vaart, A.-W., Wellner, J.A., 1996. Weak Convergence and Empirical Processes. With Applications to Statistics. Springer, New York.

Please cite this article as: Holzmann, H., Testing parametric models in the presence of instrumental variables. Statist. Probab. Lett. (2007), doi:10.1016/j.spl.2007.09.025