Testing Homogeneity in Demand Systems Nonparametrically: Theory and Evidence

Berthold R. Haag — HypoVereinsbank
Stefan Hoderlein — Brown University*
Sonya Mihaleva — Brown University

September 24, 2009

Abstract

Homogeneity of degree zero has often been rejected in empirical studies that employ parametric models. This paper proposes a test for homogeneity that does not depend on the correct specification of the functional form of the empirical model. The test statistic we propose is based on kernel regression and extends nonparametric specification tests to systems of equations with weakly dependent data. We discuss a number of practically important issues and further extensions. In particular, we focus on a novel bootstrap version of the test statistic. Moreover, we show that the same test also allows to assess the validity of functional form assumptions. When we apply the test to British household data, we find homogeneity generally well accepted. In contrast, we reject homogeneity with a standard almost ideal parametric demand system. Using our test for functional form we obtain however that it it precisely this functional form assumption which is rejected. Our findings indicate that the rejections of homogeneity obtained thus far are due to misspecification of the functional form and not due to incorrectness of the homogeneity assumption.

Keywords: Homogeneity, Nonparametric, Bootstrap, Specification Test, System of Equations.

*Brown University, Department of Economics, Robinson Hall #302C, Providence, RI 02912, USA, email: stefan_hoderlein@yahoo.com. We have received helpful comments from Joel Horowitz, Enno Mammen and from seminar participants at the ESEM, in Stanford and UCL/IFS. Excellent research assistance by David Hohlfeldt, as well as financial support by Landesstiftung Baden-Württemberg "Eliteförderungsprogramm" is gratefully acknowledged.

1 Introduction

Homogeneity of degree zero, more colloquially also sometimes called “absence of money illusion”, is arguably the key implication of a linear budget constraint in a standard utility maximization problem and appears in many areas of applied economics. Parametric tests of this hypothesis are very common in the analysis of consumer demand. Indeed, testing this hypothesis as well as the other integrability constraints has spurned much of the research on parametric flexible functional form models throughout the 70s and 80s, e.g., the Translog, Jorgenson et al. (1982), and the Almost Ideal, Deaton and Muellbauer (1980). Results of these tests were often negative or inconcusive, and as a result there is today a certain scepticism towards the homogeneity assumption, see Lewbel (1999) for an overview.

This paper argues that scepticism is not warranted. By testing homogeneity within a given parametric model it is actually the joint hypothesis of homogeneity and functional form that is analyzed. Hence, even though many studies find homogeneity rejected, it is not clear whether it is truly homogeneity that is rejected, or whether it is not rather the functional form that is not compatible with the data.

As alternative we propose a direct test of homogeneity that does not rely on any parametric form assumption. Moreover, this test allows us to simultaneously check the functional form assumptions, and therefore allows to understand better the previously obtained results. The test we propose is of the type of a nonparametric specification test for omission of variables, and is related to Fan and Li (1996), Lavergne and Voung (2000) and Aït-Sahalia, Bickel and Stoker (2002).

From an methodological point of view, we extend this literature in two main directions: First, we extend these tests to cover hypotheses in systems of equations. This is important in our specific application. But it is also obviously of greater importance throughout economics where many applications feature systems of equations[1]. Second, we propose a “wild bootstrap” procedure and establish formally its validity. The bootstrap is extremely valuable for applied work as it helps to avoid dealing with involved limiting distributions, and complicated preestimation of elements thereof. In addition, it is well known that L_2- distance tests are an instance when the asymptotic distribution theory provides a poor approximation, see Hjellvik and Tjøstheim (1995). In this case, the bootstrap has the advantage of generating better approximations to the unknown finite sample distribution. For recent work on the bootstrap, consider Davidson and Flachaire (2008). In addition to these two main contributions, we also extend the literature by discussing additional issues that are important for applications: Specif-

[1] Alternative applications appear throughout Economics. Examples include the standard sample selection and IV models, but also in simultaneous equations (market equilibrium) models.

ically, we show how to handle local polynomials, semiparametric alternatives and dependent data.

From an economic point of view, to the best of our knowledge this is the first nonparametric test of homogeneity of degree zero. There are some nonparametric tests of related economic hypothesis: Lewbel (1995) and Haag, Hoderlein and Pendakur (2009) are both concerned with testing Slutsky symmetry. Because this hypothesis involves levels and derivatives, the specific structure of these tests are quite different. Kim and Tripathi (1999) discuss nonparametric estimation under the restricion of homogeneity of degree zero. Other parametric tests of homogeneity of degree *zero* date back at least to the flexible functional forms demand systems (Jorgenson et al. (1982), Deaton and Muellbauer (1980)). Related is also the work of Stoker (1989), but he discusses testing in an average derivative setup which is very different from ours. In Hoderlein (2009) we discuss identification and implementation of economic hypotheses, amongst them also homogeneity. The test is only applied at fixed positions of nonparametric estimators, and does hence not integrate over the function. Moreover, in Hoderlein (2009) we do not provide any large sample theory, and thus the paper is very different. Finally, related is also the work of Yatchew and Bos (1997) who provide specification testing that may be applied to homogeneity, using a nonparametric least squares method that is related to spline estimation and different from our kernel based approach.

This paper is organized as follows: We consider the test statistic in the second section. This will proceed in the following fashion. We introduce the test formally, and discuss the conventional asymptotic theory. Moreover, in the same section we propose also a bootstrap version of the test statistic. We establish the validity of this bootstrap version, along with a cooking recipe for implementation. Finally, we discuss three extensions to the basic test statistic which are particularly important for our application. These are local polynomials, (semi-)parametric alternatives and dependent (i.e., mixing) data.

In the third section, we focus on the application. We will implement the test for homogeneity of degree zero and functional form using the British FES data. Broadly speaking, our results are affirmative of the homogeneity hypotheses and trace rejections previously obtained in the literature to misspecifications of the functional form. Finally, we conclude with an outlook.

2 From Economic Hypothesis to Test Statistic

2.1 Economic Background

Homogeneity of degree zero is arguably the key implication of a linear budget constraint in a standard utility maximization problem and appears in many areas of applied economics. To define this property formally, let the demand function of an individual be defined as follows. Assume the individual has preferences $b \in \mathcal{B}$, where $\mathcal{B}$ is a preference space, e.g., the space of r times continuously differentiable utility functions. Let the corresponding utility optimal demand be denoted as $y = \phi(p, z; b)$. Here, p is a H vector of log prices, z is log income (in demand income is - under a separability assumption on preferences - usually total expenditure), and y is a $H-1$ vector of budget shares. We assume that the adding up constraint has already been imposed, and we take H to be the outside alternative.

On individual level, the linearity of the budget constraint imposes that

$$\phi(p, z; b) = \phi(p + \lambda, z + \lambda; b),$$

for $\lambda \in \mathbb{R}$, and any $(b, p, z) \in \mathcal{B}\times\ \mathbb{R}^{H+1}$. Substituting for $\lambda = -z$, Homogeneity of degree zero corresponds then to the property that $\phi(p, z; b) = \phi(p - z, 0; b)$, implying also $\phi(p - z, z; b) = \phi(p - z; b)$, i.e., using the notation $x = p - z$, this imposes the restriction that

$$\phi(x, z; b) = \phi(x; b), \tag{2.1}$$

for any $(b, x, z) \in \mathcal{B}\times\ \mathbb{R}^{H+1}$. The original homogeneity of degree zero hypotheses can thus be transformed to an omission of variables hypothesis for every individual.

This hypothesis on the $H-1$ dimensional system of equations has to hold for any preference ordering b. However we do not observe the individual's preference ordering b, and only observe a finite dimensional vector of household covariates (denoted q), as well as the economic variables y, x, z. Moreover, we assume that there exists a heterogeneous population, so that all variables B, X, Z, Q are random variables, and have a nondegenerate distribution function F_{BXZQ}. To bridge the gap between the unobservable world of preference and the observable regressions, we assume that $F_{B|XZQ} = F_{B|XQ}$, which is implied if homogeneity of degree zero in x and z also were to hold for the distribution of unobservables, but more generally holds if Z is independent of B given X. This conditional independence assumption of matching type is typical for the general nonseparable setup we consider. We establish in Hoderlein (2009) in this scenario that

$$\phi(X, Z; B) = \phi(X; B) \ \text{ a.s. } \mathbb{P}_{XZB} \Longrightarrow \ \mathbb{E}(Y|X, Z, Q) = \mathbb{E}(Y|X, Q) \ \text{ -a.s. } \mathbb{P}_{XZQ}, \tag{2.2}$$

i.e., we can use the observable mean regression to test for homogeneity of degree zero. Finally, we can extend this approach to allow for endogeneity in a control function fashion by assuming that we have instruments S such that

$$Z = \mathbb{E}(Z|S, P) + V,$$

if we assume that $F_{B|XZQV} = F_{B|XQV}$, cf. again Hoderlein (2009), by simply adding the control function residuals V to our list of control regressors Q in the mean regression.

2.2 Transforming the Hypothesis

In the following, for ease of notation we neglect possible dependencies of the regression function on Q and V, and focus on a model that captures the relationship between the random vectors Y, X and Z. Here $Y \in \mathbb{R}^{d_Y}$ is a d_Y-dimensional dependent variable, and $X \in \mathbb{R}^{d_X}, Z \in \mathbb{R}^{d_Z}$ are predictors, such that $d_Y = H - 1, d_X = H$ and $d_Z = 1$. The hypothesis to be tested is whether Z can be omitted from the regression of Y on (X, Z). For testing this hypothesis, we define the following functions

$$\mu(x,z) = \mathbb{E}(Y \mid X = x, Z = z)$$
$$m(x) = \mathbb{E}(Y \mid X = x).$$

If it is possible to exclude Z from the regression, then these functions should coincide almost surely. Hence, we will base the test statistic on the null hypothesis

$$H_0\colon \quad \mathbb{P}(\mu(X,Z) = m(X)) = 1,$$

while the alternative is that they differ on a subset of the support of Z of positive measure. The null is equivalent to the condition that the L_2 distance of the two functions is zero. Using a positive and bounded weighting function $a(x,z)$ this condition can be written as

$$\Gamma = \mathbb{E}\Big(\sum_{j=1}^{d_Y} \big(\mu^j(X,Z) - m^j(X)\big)^2 a(X,Z)\Big) = 0. \tag{2.3}$$

Using the fact that $m^j(X) = \mathbb{E}(m^j(X) \mid X, Z)$, we base the test on

$$\Gamma = \mathbb{E}\Big(\sum_{j=1}^{d_Y} \big(\mu^j(X,Z) - \mathbb{E}(m^j(X) \mid X,Z)\big)^2 a(X,Z)\Big). \tag{2.4}$$

As mentioned above, alternative test statistics for the single equation case ($d_Y = 1$) have been proposed in the literature. Aït-Sahalia, Bickel and Stoker (2002) base their test statistic directly on equation (2.3), while Fan and Li (1996) propose to base a test statistic on

$$\mathbb{E}\big((Y - m(X))\,\mathbb{E}(Y - m(X) \mid X,Z) f(X,Z) a(X,Z)\big).$$

To avoid technical problems, Fan and Li (1996) use $a(X,Z) = f^2(X,Z)a'(X,Z)$ and a leave-one-out estimator for the conditional expectation. Another possibility would be to compare residual sums of squares, i. e. basing a test statistic on

$$\mathbb{E}\big(((Y - m(X))^2 - (Y - \mu(X,Z))^2) a(X,Z)\big)$$

which would be an adaptation of the tests by Dette (1999) and Fan, Zhang and Zhang (2001) to the problem of omitting variables. To our knowledge such a test has not yet been implemented. We expect that its local power properties are worse than those of a test based on (2.3) or (2.4) (see Dette (1999), who shows these worse power properties for the case of a linear null hypothesis).

2.3 Sample Counterpart Test Statistic

As test statistic serves the sample counterpart of Γ in (2.4). Given a sample of n independent and identically distributed random vectors $(Y_1, X_1, Z_1), \ldots, (Y_n, X_n, Z_n)$, we replace the unknown functions $m(x)$ and $\mu(x,z)$ by their Nadaraya-Watson estimators $\hat{m}_{\tilde{h}}(x)$ and $\hat{\mu}_h(x,z)$. Formally, these are defined as vectors with the one-dimensional estimators, $\hat{m}^j_{\tilde{h}}(x) = \sum_{i=1}^n K_{\tilde{h}}(x - X_i) Y_i^j / \sum_{i=1}^n K_{\tilde{h}}(x - X_i)$ and $\hat{\mu}^j_h(x,z) = \sum_{i=1}^n K_h(x - X_i, z - Z_i) Y_i^j / \sum_{i=1}^n K_h(x - X_i, z - Z_i)$, where $K_h(u) = K(u/h)/h$ with a kernel K and bandwidths h and $\tilde{h}$. As an estimator for $\mathbb{E}(m^j(X) \mid X = x, Z = z)$ we propose

$$\widehat{\mathcal{K}_n m^j_{\tilde{h}}}(x,z) = \frac{\sum_{i=1}^n K_h(x - X_i, z - Z_i)\hat{m}^j_{\tilde{h}}(X_i)}{\sum_{i=1}^n K_h(x - X_i, z - Z_i)}.$$

Then, the statistic is given by

$$\hat{\Gamma}_{\mathcal{K}} = \frac{1}{n} \sum_{j=1}^{d_Y} \sum_{i=1}^{n} \left(\hat{\mu}^j_h(X_i, Z_i) - \widehat{\mathcal{K}_n m^j_{\tilde{h}}}(X_i, Z_i)\right)^2 A_i \tag{2.5}$$

with $A_i = a(X_i, Z_i)$. The additional smoothing step associated with $\widehat{\mathcal{K}_n m_{\tilde{h}}}(x,z)$ eliminates the bias coming from $\hat{\mu}_h(x,z)$, thereby reduces the number of bias components in the asymptotic expression. This reduction in turn allows to employ less restrictive requirements on the bandwidths. The superiority of $\hat{\Gamma}_{\mathcal{K}}$ over the tests of Aït-Sahalia, Bickel and Stoker (2002) and Fan and Li (1996) can be stated in terms of the local power properties of the tests and will be discussed after theorem 2.

2.4 Asymptotic Distribution of the Test Statistic

In order to treat the asymptotic distribution of the test statistic, we introduce the following assumptions. The first two assumptions are concerned with the data generating process.

Assumption 1. *The data $(Y_i, X_i, Z_i), i = 1, \ldots, n$ are independent and identically distributed with density $f(y,x,z)$.*

Assumption 2. *For the data generating process*

1. *The continuously differentiable weighting function $a(x,z)$ is nonzero and bounded with compact support $\mathcal{A} \subset \mathbb{R}^{d_X + d_Z}$.*

2. *$f(y,x,z)$ is r-times continuously differentiable ($r \geq 2$). f and its partial derivatives are bounded and square-integrable on $\mathcal{A}$.*

3. *$\mu(x,z)$ and $m(x)$ are $r+1$-times continuously differentiable.*

4. *$f(x,z) = \int f(y,x,z)\,dy$ is bounded from below on $\mathcal{A}$, i. e. $\inf_{(x,z)\in\mathcal{A}} f(x,z) = b > 0$.*

5. *The covariance matrix*

$$\Sigma(x,z) = (\sigma^{ij}(x,z))_{1\le i,j\le d_Y} = \mathbb{E}((Y-\mu(X,Z))(Y-\mu(X,Z))' \mid X=x, Z=z)$$

   is square-integrable (elementwise) on $\mathcal{A}$ and.

6. *$\mathbb{E}((Y^j-\mu^j(X,Z))^2(Y^k-\mu^k(X,Z))^2) < \infty$ for every $1 \le j,k \le d_Y$.*

The first assumption may be relaxed to allow for dependent data. We will discuss this extension in section 2.6.3. Assumption 2 contains standard differentiability and integrability assumptions that do not deserve further mentioning.

The following assumptions are concerned with the kernel and the bandwidth sequences. For simplicity, we assume product kernels in both regressions. Therefore we formulate our assumptions for one-dimensional kernel functions. To simplify things further, instead of bandwidth vectors $\mathbf{h} \in \mathbb{R}^{d_X+d_Z}$ and $\tilde{\mathbf{h}} \in \mathbb{R}^{d_X}$ we assume that we have only one single bandwidth for each regression $(h, \tilde{h})$. We shall make use of the following notation: Define kernel constants

$$\kappa_k = \int u^k K(u)\,\mathrm{d}u \qquad \text{and} \qquad \kappa_k^2 = \int u^k K(u)^2\,\mathrm{d}u$$
$$\kappa_* = \int \Big(\int K(u)K(u-v)\,\mathrm{d}u\Big)^2\,\mathrm{d}v$$

Then, our assumptions regarding kernels and bandwidths are as follows:

Assumption 3. *The one-dimensional kernel is Lipschitz continuous, bounded, has compact support, is symmetric around 0 and of order r (i. e. $\int u^k K(u)\,du = 0$ for all $k < r$ and $\int u^r K(u)\,du < \infty$).*

Assumption 4. *For the bandwidths*

1. *For $n \to \infty$, the bandwidth sequence $h = O(n^{-1/\delta})$ satisfies*

$$d_X + d_Z < \delta \tag{2.6}$$

2. *For $n \to \infty$, the bandwidth sequence $\tilde{h} = O(n^{-1/\tilde{\delta}})$ satisfies*

$$2\delta\frac{d_X}{d_X+d_Z} < \tilde{\delta} \tag{2.7}$$

3. *For the order r of the kernel holds*

$$\tilde{\delta}\frac{2\delta - d_X - d_Z}{4\delta} < r \tag{2.8}$$

While the assumptions on the kernel are standard, the assumptions on the bandwidths do merit some discussion. Observe first that the optimal rate for estimating the full dimensional regression function $\mu(x, z)$, given by

$$\delta_{opt} = (d_X + d_Z) + 2r.$$

is not excluded from inequality (2.6). Under the null hypothesis, $\mu(x, z)$ does not depend on z. Then, the derivatives with respect to z are zero and the corresponding bias terms disappear. It follows that under H_0 the optimal bandwidth in the z-directions is infinite. But under the alternative and in the x-directions there exists an optimal bandwidth.

If we want to make use of this fact, however, through employing rate-optimal methods of bandwidth choice in the full dimensional regression (e.g., cross validation), then the inequalities (2.7) and (2.8) impose restrictions on the bandwidth $\tilde{h}$ of the dimension-reduced regression m. More specifically, because of (2.7), it might be necessary to use a larger-than-optimal bandwidth, and because of (2.8), to employ higher order kernels. As an example, take $d_X = 1, d_Z = 1$. It is not possible to use both δ_{opt} and $\tilde{\delta}_{opt}$ for any choice of r, because inequality (2.7) yields the restriction $\delta < \tilde{\delta}$.

An alternative representation of (2.6) - (2.8) may be given in terms of n and h. We obtain $nh^{d_X+d_Z} \to \infty$ (necessary for consistency of the kernel density estimator), $h^{d_X+d_Z}\tilde{h}^{-d_X} \to 0$ and $nh^{(d_X+d_Z)/2}\tilde{h}^{2r} \to 0$.[2] The last two conditions ensure that the estimation error of the dimension-reduced regression does not dominate the test statistic.

The restrictions on the bandwidths are much weaker than those restrictions assumed by Aït-Sahalia, Bickel and Stoker (2002). In their case the optimal rate for estimation is excluded for all regressions and higher order kernels are always needed, provided $d_X + d_Z \geq 3$. In contrast, our assumptions allow to trade the use of higher order kernel and an larger-than-optimal bandwidth.

In practise we propose to calculate data-driven bandwidths (by cross-validation) for the dimension reduced regression. In case the optimal rate is excluded, we suggest to adjust the bandwidth by $n^{1/\tilde{\delta}_{opt}-1/\tilde{\delta}}$. Although we do not formally address the issue of data-driven bandwidths $\hat{h}$ we assume that our results will hold if $\hat{h}/h \xrightarrow{P} 1$.

For the first theorem, we introduce the following quantities

$$\sigma_\Gamma^{ij} = \iint \sigma^{ij}(x, z)^2 a(x, z)^2 \,\mathrm{d}x \,\mathrm{d}z \qquad b_\Gamma^i = \iint \sigma^{ii}(x, z) a(x, z) \,\mathrm{d}x \,\mathrm{d}z.$$

The asymptotic normality of the test statistic is given by the following

[2]Note that these restrictions imply $n\tilde{h}^{d_X} \to \infty$, which ensures the consistency of the dimension reduced regression.

Theorem 1. *Let assumptions 1–4 hold. Then we have that under H_0*

$$\Sigma_{\mathcal{K}}^{-1}(nh^{(d_X+d_Z)/2}\hat{\Gamma}_{\mathcal{K}} - h^{-(d_X+d_Z)/2}B_{\mathcal{K}}) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

where

$$\Sigma_{\mathcal{K}}^2 = 2(\kappa_*)^{d_X+d_Z}\Big(\sum_{i=1}^{d_Y}\sigma_\Gamma^{ii} + 2\sum_{i<j}\sigma_\Gamma^{ij}\Big) \qquad B_{\mathcal{K}} = (\kappa_0^2)^{d_X+d_Z}\sum_{i=1}^{d_Y} b_\Gamma^i.$$

Simplifying the proofs in the appendix to one line, the test statistic can be written as

$$\hat{\Gamma}_{\mathcal{K}} = \Gamma_n + I_n + U_n \tag{2.9}$$

where $\Gamma_n = 0$ under H_0, U_n depends upon the uniform rate of convergence of the restricted estimator, and I_n is a degenerated U-statistic which dominates asymptotically. This U-statistic converges at the rate $nh^{(d_X+d_Z)/2}$, which is faster than $n^{1/2}$, under the admissible bandwidth sequence.

Next, we investigate the behavior of the test statistic under the alternative. There are a number of efficiency measures (e. g. Bahadur efficiency or Hodges-Lehman efficiency) to compare two test statistics. The most common one is the asymptotic relative efficiency (Pitman efficiency) which compares the behavior of the tests under local alternatives. To this end, define a sequence of alternatives

$$H_{1n}\colon \mu(x,z) = m(x) + \varepsilon_n(x,z)$$

where $\varepsilon_n(x,z)$ is a converging sequence of functions. Note that fixed alternatives are included for $\varepsilon_n(x,z) = \varepsilon(x,z) \neq 0$.

Theorem 2. *Let Assumptions 1–4 hold. If there exists a constant B_L such that*

$$\lambda_n \sum_{k=1}^{d_Y}\frac{1}{n}\sum_{j=1}^{n}\Big(\frac{\varepsilon_n^k(X_j,Z_j)}{f(X_j,Z_j)}\Big)^2 a(X_j,Z_j) \xrightarrow{P} B_L$$

for $\lambda_n = O(nh^{(d_X+d_Z)/2})$. Then we have that under H_{1n}

$$\Sigma_{\mathcal{K}}^{-1}(nh^{(d_X+d_Z)/2}\hat{\Gamma}_{\mathcal{K}} - h^{-(d_X+d_Z)/2}(B_{\mathcal{K}} + \kappa_0^2 B_L)) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

For a fixed alternative it holds that $nh^{(d_X+d_Z)/2}\hat{\Gamma}_{\mathcal{K}} \to \infty$.

The test cannot detect alternatives that converge to zero at a rate faster than $n^{-1/2}h^{-(d_X+d_Z)/4}$. This means that the test suffers from the curse of dimensionality because the rate decreases as the number of dimensions increase. Aït-Sahalia, Bickel and Stoker (2002) and Fan and Li (1996) have established local power properties of their tests and both obtain the same rate. Theorem 2 holds for the test of Aït-Sahalia, Bickel and Stoker (2002) in an analogous fashion. A comparison with the test of Fan and Li (1996) is only possible using $a(x,z)f(x,z)^{-2}$ as a weighting function, since Fan and Li (1996) use density weighting. The asymptotic variance differs through a kernel related constant. Because $\kappa_* < \kappa_0^2$ for a density K, our test is asymptotically relatively more efficient than the test of Fan and Li (1996).

2.5 Bootstrap-Implementation

The direct way to implement the test is to estimate the expected value $B_{\mathcal{K}}$ and the variance $\Sigma^2_{\mathcal{K}}$. This requires the estimation of integrals like

$$\int \sigma^{jj'}(x,z)^k a(x,z)^k \, \mathrm{d}x \, \mathrm{d}z \qquad k = 1,2, \; j,j' = 1,\dots,d_Y. \tag{2.10}$$

Therefore estimators of the conditional (co)variances are needed. A Nadaraya-Watson-type estimator may be defined as

$$\hat{\sigma}_h^{jj'}(x,z) = \frac{\sum_{i=1}^n K_h(x - X_i, z - Z_i)(Y_i^j - \hat{\mu}_h^j(X_i, Z_i))(Y_j^{j'} - \hat{\mu}_h^{j'}(X_i, Z_i))}{\sum_{i=1}^n K_h(x - X_i, z - Z_i)}.$$

This estimator has better properties than the difference between estimators of the second and the squared first conditional moment of Y given X and Z (see Fan and Yao, 1998). Now the integral in (2.10) can be calculated numerically. To ensure consistency of the standardized test statistic the underlying (co)variance estimators (as well as the density estimator) have to be chosen such that

$$\sup_{(x,z)\in\mathcal{A}} |\hat{\sigma}_h^{jj'}(x,z) - \sigma^{jj'}(x,z)| = o_P(h^{-(d_X+d_Z)/2})$$

Estimating the components of the asymptotic distribution of $\hat{\Gamma}_{\mathcal{K}}$ is cumbersome. Moreover, it is also problematic: In the proof of the asymptotic normality of the test statistic many terms of lower magnitude are omitted. Asymptotic approximations involving U-statistics work often very poorly in a finite sample, as was pointed out by Hjellvik and Tjøstheim (1995). To avoid this problem we propose a wild bootstrap procedure to derive critical values for the test statistic, as in Härdle and Mammen (1993). In our setting this is performed in the following way

1. Calculate (multivariate) residuals $\hat{\varepsilon}_i = Y_i - \hat{m}_{\tilde{h}}(X_i)$.
2. For each i randomly draw $\varepsilon_i^* = (\varepsilon_i^{1*}, \dots, \varepsilon_i^{d_Y,*})'$ from a distribution $\hat{F}_i$ that mimics the first three moments of $\hat{\varepsilon}_i$.
3. Generate the bootstrap sample $(Y_i^*, X_i^*, Z_i^*), i = 1,\dots,n$ by $Y_i^* = \hat{m}_{\tilde{h}}(X_i) + \varepsilon_i^*$ and $X_i^* = X_i, Z_i^* = Z_i$.
4. Calculate $\hat{\Gamma}^*_{\mathcal{K}}$ from the bootstrap sample $(Y_i^*, X_i^*, Z_i^*), i = 1,\dots,n$.
5. Repeat steps 2 to 4 often enough to obtain critical values for $\hat{\Gamma}_{\mathcal{K}}$.

Assumption 5. *For the bootstrap distribution*
The bootstrap residuals $\varepsilon_i^*, i = 1,\dots,n$ are drawn independently from distributions $\hat{F}_i$, such that $\mathbb{E}_{\hat{F}_i}\, \varepsilon_i^* = 0, \mathbb{E}_{\hat{F}_i}\, \varepsilon_i^*(\varepsilon_i^*)' = \hat{\varepsilon}_i\hat{\varepsilon}_i'$ and $\mathbb{E}_{\hat{F}_i}(\varepsilon_i^{k,*})^4 < \infty$ for all $k = 1,\dots,d_Y$.

This set of admissible distributions is very general. Apart from the simple wild bootstrap, a smooth conditional moment bootstrap as in Gozalo (1997) may also be used. In the classical wild bootstrap, residuals are drawn from a two-point distribution that takes the value $\hat{\varepsilon}_i(1-\sqrt{5})/2$ with probability $(5+\sqrt{5})/10$ and $\hat{\varepsilon}_i(1+\sqrt{5})/2$ else. Assumption 5 is fulfilled for discrete distributions, distributions with compact support and - among others - for the normal distribution. These are the most commonly used distributions in practice.

The theoretical result concerning this bootstrap procedure is given in

Theorem 3. *Let assumptions 1–5 be true. Under H_0,*

$$\Sigma_{\mathcal{K}}^{-1}(nh^{(d_X+d_Z)/2}\hat{\Gamma}^*_{\mathcal{K}} - h^{-(d_X+d_Z)/2}B_{\mathcal{K}}) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

conditional on the data $(Y_1, X_1, Z_1), \ldots, (Y_n, X_n, Z_n)$ with probability tending to one.

To prove theorem 3 it is sufficient to assume that the bootstrap distribution $\hat{F}_i$ mimics the first two moments of $\hat{\varepsilon}_i$. Using an Edgeworth expansion in the proof, we conjecture that matching the first three moments yields a higher order approximation. Therefore we recommend to mimic three moments in applications.

2.6 Extensions

In this section we discuss extensions to the test statistic along three lines that are important modifications of our tests that we implement in our application. First, we explore the use of local polynomial estimators to replace the Nadaraya-Watson estimator, because they allow higher order bias reduction without employing higher order kernels. Second, we extend the test statistic to (semi-)parametric hypotheses. Third, we investigate the behavior of the test in the case of dependent data, because in our application we pool several cross sections, but the main variation in prices is across time. In the following discussion, we focus on the modifications of theorem 1. Changes in the proofs of the bootstrap result and the local power properties are straightforward and omitted for brevity of exposition..

2.6.1 Local Polynomials

In nonparametric regression analysis the superiority of local polynomial estimators to Nadaraya-Watson estimators is well known (see Fan and Gijbels, 1996). As a consequence, local polynomials have become the dominant approach in applied work. Therefore it is a natural extension to use local polynomial estimators for $\mu(x,z)$ and $m(x)$ in the test statistic. Recall that they are defined via minimizing

$$\sum_{i=1}^{n}\big(Y_i^j - \sum_{0\le|\mathbf{k}|\le p} b^j_{\mathbf{k}}(x,z)\big(X_i - x, Z_i - z\big)^{\mathbf{k}}\big)^2 K_h(X_i - x, Z_i - z), \tag{2.11}$$

with respect to all $b^j_{\mathbf{k}}$ for $j = 1, .., d_Y$. For vectors $\mathbf{k} = (k_1, \ldots, k_{d_X+d_Y})$ we have utilized the notation $|\mathbf{k}| = \sum_i k_i$ and $x^{\mathbf{k}} = \prod_i (x^i)^{k_i}$. Then $\hat{\mu}^{j,LP}_h(x, z)$ is defined as the solution for $b^j_{\mathbf{0}}$. Introducing the quantities

$$\hat{t}^j_{\mathbf{k}}(x, z) = \frac{1}{n}\sum_{i=1}^{n} Y^j_i \Big(\frac{(X_i - x, Z_i - z)}{h}\Big)^{\mathbf{k}} K_h(X_i - x, Z_i - z),$$
$$\hat{f}_{h,\mathbf{k}}(x, z) = \frac{1}{n}\sum_{i=1}^{n} \Big(\frac{(X_i - x, Z_i - z)}{h}\Big)^{\mathbf{k}} K_h(X_i - x, Z_i - z),$$

which are arranged in a vector $\hat{T}^j(x, z) = (\hat{t}^j_{\mathbf{k}}(x, z))_{\mathbf{k}}$ and a matrix $\hat{S}(x, z) = (\hat{f}_{h,\mathbf{k}+\mathbf{j}}(x, z))_{\mathbf{k},\mathbf{j}}$ in a lexicographical order.[3] With this notation, the estimator can be written explicitly as

$$\hat{\mu}^{j,LP}_h(x, z) = \lfloor \hat{S}^{-1}(x, z)\hat{T}^j(x, z)\rfloor_1,$$

where $\lfloor\cdot\rfloor_1$ extracts the first element of a vector. $\hat{m}^{j,LP}_{\tilde{h}}(x)$ is defined analogously. The local polynomial version of $\mathbb{E}(m^j(X) \mid X, Z)$ is defined as the solution to (2.11) where Y^j_i is replaced with $\hat{m}^{j,LP}_{\tilde{h}}(X_i)$. Explicitly it can be written as

$$\widehat{\mathcal{K}_n m^{j,LP}_{\tilde{h}}}(x, z) = \lfloor \hat{S}^{-1}(x, z)\tilde{T}^j(x, z)\rfloor_1,$$

where the elements of the vector $\tilde{T}^j(x, z)$ are given by

$$\tilde{t}^j_{\mathbf{k}}(x, z) = \frac{1}{n}\sum_{i=1}^{n} \hat{m}^{j,LP}_{\tilde{h}}(X_i)\Big(\frac{(X_i - x, Z_i - z)}{h}\Big)^{\mathbf{k}} K_h(X_i - x, Z_i - z).$$

The new test statistic is then the analog to (2.5)

$$\hat{\Gamma}^{LP}_{\mathcal{K}} = \frac{1}{n}\sum_{j=1}^{d_Y}\sum_{i=1}^{n}\big(\hat{\mu}^{j,LP}_h(X_i, Z_i) - \widehat{\mathcal{K}_n m^{j,LP}_{\tilde{h}}}(X_i, Z_i)\big)^2 A_i \tag{2.12}$$

To define the kernel constants arising in the bias and variance parts of the asymptotic distribution, we have to define the matrix $M = (\kappa_{\mathbf{j}+\mathbf{k}})_{\mathbf{j},\mathbf{k}}$ with entries $\kappa_{\mathbf{k}} = \int u^{\mathbf{k}} K(u)\,\mathrm{d}u$. In an abuse of notation we denote with $\kappa^{-1}_{\mathbf{k}}$ the elements of the first row of M^{-1}. This enables to define

$$\kappa_\Sigma = \int\Big(\int\Big(\sum_{1\le\mathbf{k}\le r}(u - v)^{\mathbf{k}}\kappa^{-1}_{\mathbf{j}}K(u - v)\Big)\Big(\sum_{1\le\mathbf{k}\le r}u^{\mathbf{k}}\kappa^{-1}_{\mathbf{j}}K(u - v)\Big)\,\mathrm{d}u\Big)^2\,\mathrm{d}v$$
$$\kappa_B = \int\Big(\sum_{1\le\mathbf{k}\le r}u^{\mathbf{k}}\kappa^{-1}_{\mathbf{j}}K(u)\Big)^2\,\mathrm{d}u$$

which we require for the derivation of the asymptotic distribution of $\hat{\Gamma}^{LP}_{\mathcal{K}}$ in the following theorem:

[3] Addition is in the Hadamard-sense, i. e. $\mathbf{j} + \mathbf{k} = (j_1 + k_1, \ldots, j_{d_X+d_Z} + k_{d_X+d_Z})$.

Theorem 4. *Let the assumptions 1–3 hold. Let assumption 4 hold for $r = p+1$ for p odd and $r = p+2$ for p even, where p is the order of the local polynomial estimator. Then we have that under H_0*

$$\Sigma_L^{-1}(nh^{(d_X+d_Z)/2}\hat{\Gamma}_{\mathcal{K}}^{LP} - h^{-(d_X+d_Z)/2}B_L) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

where

$$\Sigma_L^2 = 2(\kappa_\Sigma)^{d_X+d_Z}\Big(\sum_{i=1}^{d_Y}\sigma_\Gamma^{ii} + 2\sum_{i<j}\sigma_\Gamma^{ij}\Big) \qquad B_L = (\kappa_B)^{d_X+d_Z}\sum_{i=1}^{d_Y} b_\Gamma^i.$$

Note, that if an even order of the local polynomial fulfills the requirements of assumption 4, then also the subsequent odd order polynomial fulfills these requirements. The use of one additional order gives therefore no gain in flexibility when choosing the bandwidth sequences. Therefore, in contrast to estimation it is natural to use an even order local polynomial for testing. If we replace the corresponding kernel constants with κ_Σ and κ_B the results of theorems 2 and 3 remain to hold. This can be seen directly from the proof of theorem 4.

2.6.2 Semiparametric Modelling

The asymptotic distribution of the test is driven by the fact that the low-dimensional estimator $\hat{m}_{\tilde{h}}(x)$ converges faster than the full-dimensional estimator $\hat{\mu}_h(x,z)$. This remains true for semiparametric hypotheses, i. e.

$$H_{0S}\colon \mathbb{P}(\mu(x,z) = m(x) + G(z,\theta)) = 1,$$

where $G(z,\theta) = \big(G^1(z,\theta) + \cdots + G^{d_Y}(z,\theta)\big)$ is a known function depending on a finite-dimensional parameter vector $\theta \in \Theta$. This includes the case when $m = 0$, i.e., we have a purely parametric null, as will be the case in our application. Denote with $\hat{\theta}$ a parametric estimator that allows us to construct estimators of the nonparametric regression part under H_{0S}, i. e.,

$$\hat{m}_{\tilde{h}}^k(x,\hat{\theta}) = \frac{\sum_{i=1}^n K_{\tilde{h}}(x-X_i)(Y_i^k - G^k(Z_i,\hat{\theta}))}{\sum_{i=1}^n K_{\tilde{h}}(x-X_i)}.$$

Then we propose to use as test statistic

$$\hat{\Gamma}_{\mathcal{K}}^S = \frac{1}{n}\sum_{j=1}^{d_Y}\sum_{i=1}^{n}\big(\hat{\mu}_h(X_i,Z_i) - \widehat{\mathcal{K}_n m_{\tilde{h}}}^{j,S}(X_i,Z_i)\big)^2 A_i,$$

with

$$\widehat{\mathcal{K}_n m_{\tilde{h}}}^{j,S}(x,z) = \frac{\sum_{i=1}^n K_h(x-X_i, z-Z_i)(\hat{m}_{\tilde{h}}^k(x,\hat{\theta}) + G^k(Z_i,\hat{\theta}))}{\sum_{i=1}^n K_h(x-X_i,z-Z_i)}.$$

To obtain an asymptotic result we require the following assumption on the speed of convergence of the semiparametric estimator:

Assumption 6. *$G^k(z,\theta) - G^k(z,\hat{\theta}) = o_P(n^{-1/2}h^{(d_X+d_Z)/4})$ for all $k = 1,\dots,d_Y$ uniformly in $\mathcal{A}_Z = \{z \mid \exists\ x\ s.t.\ (x,z) \in \mathcal{A}\}$ and $\theta \in \Theta$.*

This assumption is stated in a very general fashion. It has to be checked for a specific model and estimation problem. As an example, consider the linear model with $d_X = 0$ and $G(z,\theta) = \theta' z$. Least squares and GMM estimators are known to be root-n consistent and Assumption 6 is trivially fulfilled. Moreover, as a special case for $d_Y = 1$ we obtain the test introduced by Härdle and Mammen (1993).

The asymptotic distribution of the test is stated in the following

Theorem 5. *Let Assumptions 1–4 and 6 hold. Then we have that under H_{0S}*

$$\Sigma_{\mathcal{K}}^{-1}(nh^{(d_X+d_Z)/2}\hat{\Gamma}^S_{\mathcal{K}} - h^{-(d_X+d_Z)/2}B_{\mathcal{K}}) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

where $\Sigma_{\mathcal{K}}$ and $B_{\mathcal{K}}$ is given as in Theorem 1.

2.6.3 Dependent data

The assumption of independent and identically distributed data is very restrictive, particularly in demand analysis where in many data sets time series effects are present. To deal with this complication, we extend the results of the previous sections to the case of mixing random variables. For a time series $W_i = (Y_i, X_i, Z_i), i = 1,\dots,n$ we define the sigma algebras $\mathcal{F}_s^t = \sigma(W_s, W_{s+1},\dots,W_t$ with $-\infty \le s < t \le \infty$ and the β-mixing coefficients

$$\beta(n) = \sup_{t\in\mathbb{Z}}\ \mathbb{E}\big(\sup_{A\in\mathcal{F}_{t+n}^{\infty}} |\mathbb{P}(A \mid \mathcal{F}_{-\infty}^t) - \mathbb{P}(A)|\big)$$

A process is called absolutely regular if $\beta(n) \to 0$ for $n \to \infty$. To derive the asymptotic normality of the test statistic, we invoke the following additional assumptions

Assumption 7. *For dependent data*

1. *The data $W_i = (Y_i, X_i, Z_i), i = 1,\dots,n$ are strictly stationary and absolutely regular with mixing coefficients $\beta(n)$. The stationary density is denoted by $f(w)$.*

2. *The density of the joint distribution of (W_q, W_r, W_s, W_t) is bounded and continuously differentiable for all q, r, s, t.*

3. *For some $\nu > 1$ it holds that $\mathbb{E}\,|Y^j|^{4\nu} < \infty$ for all $j = 1,\dots,d_Y$.*

4. *For the mixing coefficients we have the summability conditions*

$$\sum_{i=1}^{\infty} \beta(i)^{1-2/\nu} < \infty \qquad \sum_{i=1}^{\infty} i^{a'}\beta(i)^{1-2/a}$$

with $2 < a < 4\nu$ *and* $a' > 1 - 2/a$.

It holds that $\sum_{n=1}^{\infty} \psi(n) < \infty$ *where*

$$\psi(n) = \frac{nL(n)}{r(n)} \Big(\frac{nT(n)^2}{\tilde{h}^{d_X} \log n} \Big)^{1/4} \beta(r(n))$$

with $L(n) = (nT(n)^2/(\tilde{h}^{d_X+2} \log n))^{d_X/2}$, $r(n) = (n\tilde{h}^{d_X} / \log n)^{1/2}/T(n)$ *and* $T(n) = \big(n \log n (\log\log n)^{1+\epsilon}\big)^{1/4\nu}$.

For $m = n^{1/\delta_m}$ *with* $\delta_m > 4\delta$ *and* $1/\delta + 1/\delta_m < 3/2$ *it holds that*

$$n^6 h^2 (m^2 \beta(m)^{1-1/\nu} + n^2 \beta(m)^{2-2/\nu}) \to 0$$

as $n \to \infty$.

These assumptions are not restrictive: Many well known time series models have shown to be absolutely regular, most of them with exponentially decaying mixing coefficients. For mixing coefficients with geometric decay, the requirements of assumption 7 are directly fulfilled (for some $\nu > 1$). In our application, we consider the log of relative prices (i.e., the log of prices divided by income). This quantity is approximately stationary.

The dependence structure of Y, X and Z is only modelled in terms of differentiability assumptions on their joint density. This is general enough to cover the cases where X and Z are lagged values of Y. Beside time series regression, the test can be used to determine the order of a nonparametric AR-process as well as to test for parametric AR-structure.

This assumption enables us to state the following extension to the previous theorems.

Theorem 6. *Theorems 1–4 remain valid, if we replace assumption 1 by assumption 7.*

Asymptotic results under mixing assumptions are obtained by a trade-off between the number of existing moments and the decaying rate of the mixing coefficients. This is given in terms of the parameter ν. The use of a larger bandwidth may also reduce the requirements on the rate of decay (and the moment conditions). Here, this is given in terms of the sequence $\psi(n)$.

3 Application

In this section we put the test to work. We take our application from standard consumer demand using a workhorse data set, the FES. We start by giving an overview of the data, the methods of data clearance and of the definitions of variables involved.

Data description and data clearance: Our data source is the British Family Expenditure Survey (FES). Every year, the FES reports the income, expenditures, demographic composition and other characteristics of about 7 000 households. The sample surveyed represents about 0.05% of all households in the United Kingdom. The information is collected partly by interview and partly by records. Records are kept by each household member, and include an itemized list of expenditures during 14 consecutive days. The periods of data collection are evenly spread out over the year. The information is then compiled and provides a repeated series of yearly cross-sections. We use the years 1974-1999, but exclude the respective Christmas periods as they contain too much irregular behavior.

All the goods are grouped into four categories, Group 1 to 4. The first category is related to food consumption and consists of the subcategories food bought, food out (catering) and tobacco, which are self explanatory. The second category contains expenditures which are related to the house, namely housing (a more heterogeneous category; it consists of rent or mortgage payments as well as nondurable household goods and services). Finally, the last group consists of motoring and fuel expenditures, categories that are often related to energy prices. All other goods are included in the fourth category. For brevity, we call these categories food, housing, energy and others. These broader categories are formed since more detailed accounts suffer from infrequent purchases (recall that the recording period is 14 days) and are thus underreported. Together the first three categories account for 50-60% of expenditures. We removed outliers by excluding the upper and lower 2.5% of the population in all groups. The corresponding price indices which are components of the Retail Price Index are published at the National Statistics Online web site.

"Income" in demand analysis under an additive separability assumption of preferences over time and decisions equals total expenditure. It is obtained by adding up all expenditures, with a few exceptions which are known to have measurement error like tobacco. This is done to define nominal income; real income is then obtained by dividing through the retail price indices.

In this paper, we stratify the population to obtain more homogeneous subpopulations. More specifically, like much of the demand literature we focus on one subpopulation, namely two person households, sampled in a certain time interval, both adults, at least one of which is working and the head of household is a white collar worker. This focus is also justified because other subpopulations are much more prone to measurement problems. It is likely that there is remaining preference heterogeneity. We abstract from this problem here, but see Hoderlein (2009) on this issue. Finally, we reduce the remaining household variation to one approximately continuously distributed principal component. We provide summary statistics of the data in the appendix.

Econometric Issues and Empirical Results: We prove that the rejections obtained in the literature thus far are due to wrong specification of the functional form. First we apply a parametric model used often in consumer demand, the Almost Ideal Demand System, and show that the hypothesis of homogeneity is rejected. Second, we relax the functional form assumption and apply a nonparametric demand model and our test. Once the restrictive assumption is removed, homogeneity is broadly accepted. Then, we assess the validity of the functional form assumption. We apply the test described in section 2.6.2 to compare nonparametric versus parametric regression fits and conclude that the parametric model is misspecified.

1. Testing for homogeneity in a parametric model. Our parametric example will be the workhorse model in the literature, the Almost Ideal Demand System (AIDS) of Deaton and Muellbauer (1980). The parameters of the model are H vectors α and δ, and a $H \times H$ matrix γ. The model consists of theoretical demand functions of the form

$$Y = m(P, Z) + U \tag{3.1}$$

where $\mathbb{E}(U \mid P, Z) = 0_H$ and

$$m(P, Z) = \alpha + \gamma P + \delta \left[Z - g\left(P\right)\right] \tag{3.2}$$

where

$$g\left(P\right) = d + \alpha P + \frac{1}{2} P' \gamma P \tag{3.3}$$

is the log of a price index that deflates total expenditures.

When estimating the standard AIDS model defined through equations (3.1) and (3.2), there are a number of issues to consider. The most important one is endogeneity. In consumer demand, total expenditure is taken as income concept, which is justified by assuming intertemporal separability of preferences. Since the categories of goods considered are broad and they frequently constitute a large part of total expenditure, the latter is believed to be endogenous. We follow the demand literature that usually employs as instrument labor income whose determinants are thought to be exogenous to the unobserved preferences determining, say, food consumption. To allow for endogeneity, we estimate equation (3.1) using three stage least squares (specifically, GMM with a weighting matrix that is efficient under homoscedasticity). Estimation is based on the moment conditions

$$\mathbb{E}\left[\left(Y - \alpha - \gamma P - \delta\left[Z - \alpha' P - \frac{1}{2} P' \gamma P\right]\right) R_l\right] = 0_n, \quad l = 1, ..., L \tag{3.4}$$

where R_1,...,R_L is the set of instruments.

In this parametric setup, the adding up constraint is that budget shares sum to one and requires $1'_H \alpha = 1$, $1'_H \delta = 0$, and $1'_H \gamma = 0'_H$. Homogeneity requires $\gamma 1_H = 0_H$.

By analogy with the common approximate AID model, we first linearly regress Y on a constant, P, and on $Z - P'Y$. This is not a consistent estimator, but yields reasonable starting values $\widehat{\alpha}$, $\widehat{\delta}$, and $\widehat{\gamma}$ for the calculation of $g(P)$. We can obtain new estimates exploiting the conditional linearity of equation (3.2) given $g(P)$. That is, given $g(P)$, the system is linear in parameters, and this suggests a natural iterative procedure conditioning on an updated $g(P)$ at each iteration.

A second remark concerns the possibility of the estimation of d. When the total price index and the demand system are estimated simultaneously, the magnitude of d will have an impact on the estimated coefficients via the change of $g(P)$. In order to avoid the difficulties of such simultaneous estimation, we choose to estimate d using the coefficients α, and γ from the estimation of Y, to calculate the translog price index, and re-estimate factor demand equations Y, until convergence is achieved, i.e. until α, and γ are the same in $g(P)$ and Y. However, the convergence properties of this procedure in finite samples are very poor. Therefore we adopt the simplification carried out in most of the applied studies and we set the value of d to zero.

When estimating the system we only impose the adding up constraints by omitting one good from the system. We estimate the model without the homogeneity restriction $\gamma 1_H = 0_H$, so this restriction can be tested. To allow for endogeneity we estimate the model a second time using GMM where the set of instruments consists of a constant, log labor income, the first principal component of the demographic characteristics, time trend and all elements of P.

Table 1 presents Wald tests of the homogeneity restriction. The systemwide test reveals that homogeneity is rejected, both under exogeneity and endogeneity. The table also lists test statistics and p-values of the homogeneity test separately in each demand equation. Homogeneity is rejected for food under exogeneity and for food and energy under endogeneity at the 10 % significance level.

Table 1: Parametric Model Homogeneity Test

	Under Exogeneity		Under Endogeneity	
	test statistic	p-value	test statistic	p-value
systemwide test				
	68.5427	(0.0000)	48.2073	(0.0000)
equation by equation				
food	55.1441	(0.0000)	34.0508	(0.0000)
housing	0.5528	(0.4571)	0.5210	(0.4704)
energy	1.0137	(0.3140)	2.9403	(0.0864)

2. Testing for homogeneity in a nonparametric model. When testing this hypothesis, we use the following specifications: The basic specification entails a regression of budget shares on log prices, deflated total expenditures and a principal component of demographic household characteristics. We also estimate the model under endogeneity, using labor income as instrument and employing a control function residuals approach as in Hoderlein (2009).

As already mentioned in the second section, we employ a local constant kernel regression combined with a higher order kernel. To test homogeneity when the number of goods is four, we choose the order of the kernel to be 4 and we set $\delta = 7$ and $\tilde{\delta} = 12$, which satisfies assumption 3 in section 2.

We choose the bandwidths h and $\tilde{h}$ according to the rule that $h = h_0 n^{-1/\delta} \widehat{sd}(\tilde{P})$, and $\tilde{h} = h_0 n^{-1/\tilde{\delta}} \widehat{sd}(\tilde{P}, X)$, and we choose h_0 by cross validation (i.e. a data-driven method). For the homogeneity test, we construct the bootstrap version of $\hat{\Gamma}$ as discussed in section 2 above with 199 bootstrap iterations.

Table 2 shows the result of our test statistic for homogeneity of degree zero. For the purpose of checking robustness we display the results at various values of the bandwidth constant h_0 in a neighborhood of the cross validated parameter. The p-values are in brackets.

Table 2: Nonparametric Model Homogeneity Test for different Choice of Bandwidth Constant.

	Under Exogeneity		Under Endogeneity	
	test statistic	p-value	test statistic	p-value
h_{cv}	0.0026	(0.5327)	0.0052	(0.5628)
0.7*h_{cv}	0.0004	(0.5327)	0.0029	(0.4020)
1.3*h_{cv}	0.0025	(0.2714)	0.0050	(0.1608)

We conjecture that homogeneity is accepted, as the p-values are large both under exogeneity and under endogeneity. The result is also robust to deviation from the cross validated bandwidth. The obvious question that arises now is the following: Why is homogeneity rejected by the parametric tests, but not by our nonparametric test in this application? Does this result reflect lack of power of the nonparametric testing procedure, or is it due to parametric misspecification?

3. Testing the functional form assumption. At the end we assess the validity of the functional form assumption. In table 3 we present the results of the test described in section 2.6.2. Again, we display the results at various values of the bandwidth constant h_0 in a neighborhood of the cross validated parameter. The hypothesis that consumer behavior can be described by the specific functional form of AID is generally rejected, as there is only one positive p-value. Both under exogeneity and endogeneity the test rejects, when the bandwidth is obtained by cross validation. Reducing the bandwidth does not alter the result. The only case of acceptance is under exogeneity and a larger bandwidth. From the rejection at the cross validated bandwidth, we can generally infer the wrong specification of the function form of the demand system.

Table 3: Testing nonparametric versus parametric regression fits.

	Under Exogeneity		Under Endogeneity	
	test statistic	p-value	test statistic	p-value
h_{cv}	0.0125	(0.0000)	0.0146	(0.0000)
0.7*h_{cv}	0.0179	(0.0000)	0.0245	(0.0000)
1.3*h_{cv}	0.0101	(0.1307)	0.0079	(0.0000)

Taken together, these findings cast some doubts on previously obtained rejections of homogeneity using parametric models. Clearly it is the parametric specification - not homogeneity of degree zero - which is rejected.

4 Conclusion

Homogeneity of degree zero is a core property of rational consumer behavior. Since it corresponds to a linear budget constraint, which is a plausible assumption in a market economy without frictions, one would expect to find it widely accepted in the literature. Yet the contrary is true - many parametric demand studies find homogeneity rejected, see Lewbel (1999) for an overview.

This paper offers a convenient explanation for this phenomenon. Since tests of the homogeneity property based on parametric demand systems actually test the joint hypothesis of parametric specification and homogeneity, it is not always clear what is being rejected. This paper proposes a nonparametric, distribution free test of the homogeneity hypothesis, and establishes its large sample properties. We also discuss important extensions for applications like the use of higher order polynomials. When we implement the test using British consumer data, we find that homogeneity is not any more rejected, suggesting that it was misspecified parametricity assumptions that led to previous rejections. Can we provide evidence to support this conjecture? Indeed, when we perform a parametric test of the same hypothesis, using the most commonly employed demand system (the almost ideal demand system) we find that homogeneity is rejected with exactly the same data. While this could be due to the better power properties of the test based on the parametric demand system, it may also be indicative of the effect of parametric misspecification. To this end, we perform a test of the parametric specification based on the same nonparametric testing principle that we employed when testing homogeneity. This time, however, we do find rejection. We conclude that the parametric model is rejected by the data, not homogeneity of degree zero. More general nonparametric tests like the one we propose may allow us to get a more accurate picture of real consumer behavior, and a better assessments of the merits of the theory of rational consumer behavior. Hence we advocate their use in practise.

References

[1] Aït-Sahalia, Y., Bickel, P. and T. Stoker. 2001. Goodness-of-fit tests for kernel regression with an application to option implied volatilities, *Journal of Econometrics* **105**, 363–412.

[2] Davidson, R. and E. Flachaire, 2008. The wild bootstrap, tamed at last, *Journal of Econometrics*, vol. 146(1), 162-169.

[3] Deaton, A., and J. Muellbauer 1980. An almost ideal Demand System, *American Economic Review,* 70, 312-26.

[4] Dette, H. 1999. A consisten test for the functional form of a regression based on a difference of variance estimators, *Annals of Statistics* **27**, 1012–1050.

[5] Fan, J. and I. Gijbels. 1996. *Local Polynomial Modelling and Its Applications*, Chapman and Hall, London.

[6] Fan, J. and Q. Yao. 1998. Efficient estimation of conditional variance functions in stochastic regression, *Biometrika* **85**, 645–660.

[7] Fan, J. , C.M. Zhang and J. Zhang. 2001. Generalized likelihood ratio statistics and Wilks phenomenon, *Annals of Statistics* **29**, 153–193.

[8] Fan, Y. and Q. Li. 1996. Consistent model specification tests: omitted variables and semiparametric functional forms, *Econometrica* **64**, 865–890.

[9] Härdle, W. and E. Mammen. 1993. Comparing nonparametric vs. parametric regression fits, *Annals of Statistics* **21**, 1926–1947.

[10] Hall, P. 1984. Central limit theorems for integrated squared errors of multivariate nonparametric density estimators, *Annals of Statistics* **11**, 1156–1174.

[11] Hjellvik, V. and D. Tjøstheim. 1995.. Nonparametric tests of linearity for time series. *Biometrika* **82**, 351–368.

[12] Hoderlein, S. 2009. How Many Consumers are Rational, *Working Paper*, Brown University..

[13] de Jong, P. 1987. A central limit theorem for generalized quadratic forms, *Probability Theory and Related Fields* **75**, 261–275.

[14] Jorgensen, D., Lau, L. and T. Stoker 1982; The Transcedental Logarithmic Model of Individual Behavior, in: BASMAN, R. and G. RHODES (Eds.), Advances in Econometrics, Vol 1. JAI Press.

[15] Lavergne, P. and Q. Vuong. 1996. Nonparametric selection of regressors: the nonnested case, *Econometrica* **64**, 207–219.

[16] Lewbel, A. 1999; Consumer Demand Systems and Household Expenditure, in Pesaran, H. and M. Wickens (Eds.), *Handbook of Applied Econometrics*, Blackwell Handbooks in economics.

[17] Masry, E. 1996. Multivariate local polynomial regression for time series: uniform strong consistency, *Journal of Time Series Analysis* **17**, 571–599.

[18] Stoker, T 1989. Tests of Additive Derivative Constraints, *Review of Economic Studies,* **56**, 535-52

[19] Yatchew, A. and L. Bos 1997. Nonparametric Least Squares Estimation and Testing of Economic Models, *Journal of Quantitative Economics*, 13, 81-131.

Appendix: Data

Table 4: Descriptive statistics

	10% quantile	mean	median	90% quantile	stdev
Budget share food	0.1145	0.2134	0.2017	0.3320	0.0829
Budget share housing	0.1641	0.3227	0.3094	0.5038	0.1282
Budget share energy	0.0583	0.1789	0.1578	0.3342	0.1059
Budget share others	0.0335	0.1515	0.1250	0.3115	0.1091
Priceindex food	4.5914	4.6218	4.6194	4.6593	0.0256
Priceindex housing	4.5653	4.7195	4.7415	4.8507	0.1072
Priceindex energy	4.6270	4.6747	4.6724	4.7309	0.0394
Priceindex others	4.3258	4.4205	4.4196	4.5220	0.0733
Log total expenditure	4.2344	5.2124	5.2804	6.0603	0.6858
Log income	4.4064	5.3328	5.3773	6.1726	0.6616

Appendix: Proofs

Proof of Theorem 1

For abbreviation we introduce $V_i = (X_i, Z_i)$ and $W_i = (Y_i, X_i, Z_i)$ and decompose the statistic in the following way

$$\begin{aligned}\hat{\Gamma}_{\mathcal{K}} &= \frac{1}{n}\sum_{k=1}^{d_Y}\sum_{i=1}^{n}\Big(\frac{1}{n}\sum_{j=1}^{n}\frac{K_h(V_i - V_j)}{\hat{f}_h(V_i)}(Y_j^k - \hat{m}_{\tilde{h}}^k(X_j))\Big)^2 A_i \\ &= \hat{\Gamma}_{\mathcal{K}1} + \hat{\Gamma}_{\mathcal{K}2} + \hat{\Gamma}_{\mathcal{K}3} + \hat{\Gamma}_{\mathcal{K}4}, \qquad (5.1)\end{aligned}$$

where

$$\begin{aligned}\hat{\Gamma}_{\mathcal{K}1} &= \frac{1}{n}\sum_{k=1}^{d_Y}\sum_{i=1}^{n}\Big(\frac{1}{n}\sum_{j=1}^{n}K_h(V_i - V_j)\frac{Y_j^k - \mu^k(V_j)}{\hat{f}_h(V_i)}\Big)^2 A_i \\ \hat{\Gamma}_{\mathcal{K}2} &= \frac{1}{n}\sum_{k=1}^{d_Y}\sum_{i=1}^{n}\Big(\frac{1}{n}\sum_{j=1}^{n}K_h(V_i - V_j)\frac{\mu^k(V_j) - m^k(X_j)}{\hat{f}_h(V_i)}\Big)^2 A_i \\ \hat{\Gamma}_{\mathcal{K}3} &= \frac{1}{n}\sum_{k=1}^{d_Y}\sum_{i=1}^{n}\Big(\frac{1}{n}\sum_{j=1}^{n}K_h(V_i - V_j)\frac{m^k(X_j) - \hat{m}_{\tilde{h}}^k(X_j)}{\hat{f}_h(V_i)}\Big)^2 A_i\end{aligned}$$

and $\hat{\Gamma}_{\mathcal{K}4}$ contains all cross terms. Note that under H_0 we have that $\hat{\Gamma}_{\mathcal{K}2} = 0$ almost surely. We start by investigating $\hat{\Gamma}_{\mathcal{K}1}$, which yields the asymptotic distribution and show that $\hat{\Gamma}_{\mathcal{K}3}$ and $\hat{\Gamma}_{\mathcal{K}4}$ are of lower order afterwards.
First, we write

$$\begin{aligned}\hat{\Gamma}_{\mathcal{K}1} &= \frac{1}{n}\sum_{k=1}^{d_Y}\sum_{i=1}^{n}\Big(\frac{1}{n}\sum_{j=1}^{n}K_h(V_i - V_j)\frac{Y_j^k - \mu^k(V_j)}{f(V_i)}\Big)^2\Big(\frac{f(V_i)}{\hat{f}_h(V_i)}\Big)^2 A_i \\ &= (I_{\mathcal{K}n} + \Delta_{\mathcal{K}n})(1 + o_P(1)),\end{aligned}$$

where we have defined

$$\begin{aligned}I_{\mathcal{K}n} &= \int\sum_{k=1}^{d_Y}\Big(\frac{1}{n}\sum_{j=1}^{n}K_h(v - V_j)\frac{Y_j^k - \mu^k(V_j)}{f(v)}\Big)^2 a(v)f(v)\,\mathrm{d}v \qquad (5.2)\\ \Delta_{\mathcal{K}n} &= \int\sum_{k=1}^{d_Y}\Big(\frac{1}{n}\sum_{j=1}^{n}K_h(v - V_j)\frac{Y_j^k - \mu^k(V_j)}{f(v)}\Big)^2 a(v)(\hat{f}_e(v) - f(v)\,\mathrm{d}v, \qquad (5.3)\end{aligned}$$

and $\hat{f}_e = \frac{1}{n}\sum_{i=1}^{n}\delta_{(V_i)}(v)$ denotes the empirical distribution of the sampled data (where $\delta_{(V_i)}$ is the Dirac-measure at V_i).

Starting with the leading term, we rearrange $I_{\mathcal{K}n}$ to obtain

$$\begin{aligned} I_{\mathcal{K}n} &= \frac{1}{n^2}\sum_{i<j}\sum_{k=1}^{d_Y}\int K_h(v-V_i)\frac{Y_i^k-\mu^k(V_i)}{f(v)}K_h(v-V_j)\frac{Y_j^k-\mu^k(V_j)}{f(v)}a(v)f(v)\,\mathrm{d}v \\ &\quad+\frac{1}{n^2}\sum_{i=1}^{n}\sum_{k=1}^{d_Y}\int\big(K_h(v-V_i)\frac{Y_i^k-\mu^k(V_i)}{f(v)}\big)^2a(v)f(v)\,\mathrm{d}v \\ &= I_{\mathcal{K}n,1}+I_{\mathcal{K}n,2}. \end{aligned} \tag{5.4}$$

Now it remains to show

$$nh^{(d_X+d_Z)/2}I_{\mathcal{K}n,1}\xrightarrow{\mathcal{D}}\mathcal{N}(0,\Sigma_{\mathcal{K}}^2) \tag{5.5}$$

$$nh^{(d_X+d_Z)/2}I_{\mathcal{K}n,2}-h^{-(d_X+d_Z)/2}B_{\mathcal{K}}\xrightarrow{P}0 \tag{5.6}$$

$$nh^{(d_X+d_Z)/2}\Delta_{\mathcal{K}n}\xrightarrow{P}0. \tag{5.7}$$

From this the statement of the theorem follows.

Proof of (5.5) Write

$$I_{\mathcal{K}n,1}=\sum_{i<j}h_n(W_i,W_j)$$

as U-statistic with kernel

$$\begin{aligned} h_n(W_i,W_j) = \frac{2}{n^2h^{d_X+d_Z}}&\sum_{k=1}^{d_Y}(Y_i^k-\mu^k(V_i))(Y_j^k-\mu^k(V_j)) \\ &\times\int K(u)K(u+(V_i-V_j)/h)\frac{a(V_i+uh)}{f(V_i+uh)}\,\mathrm{d}u. \end{aligned}$$

where a change of variables has been applied. Asymptotic normality is shown by using a central limit theorem for generalized U-statistics (see de Jong, 1987). Under the conditions

$$\frac{\max_{1\le i\le n}\sum_{j=1}^n\mathbb{E}\,h_n(W_i,W_j)}{\mathbf{var}\,I_{\mathcal{K}n,1}}\xrightarrow{P}0 \qquad\text{and}\qquad \frac{\mathbb{E}\,I_{\mathcal{K}n,1}^4}{(\mathbf{var}\,I_{\mathcal{K}n,1})^2}\xrightarrow{P}3 \tag{5.8}$$

it follows that

$$\sqrt{2}\frac{I_{\mathcal{K}n,1}}{\sqrt{\mathbf{var}\,I_{\mathcal{K}n,1}}}\xrightarrow{\mathcal{D}}\mathcal{N}(0,1).$$

It is immediate to see that the kernel is degenerate, symmetric and centered. Now, we introduce $\sigma_n^2=\mathbb{E}\,h_n(W_i,W_j)^2$. As we have independent and identically distributed data we can write

$$\max_{1\le i\le n}\sum_{\substack{j=1\\ j\neq i}}^n\mathbb{E}\,h_n(W_i,W_j)^2=(n-1)\sigma_n^2$$

and

$$
\begin{aligned}
\mathbf{var}\, I_{\mathcal{K}n,1} &= \sum_{i_1<i_2} \mathbf{var}\, h_n(W_{i_1}, W_{i_2}) \\
&\qquad + \sum_{i_1<i_2} \sum_{\substack{i_3<i_4 \\ (i_3,i_4)\neq(i_1,i_2)}} \mathbf{cov}(h_n(W_{i_1}, W_{i_2}), h_n(W_{i_3}, W_{i_4})) \\
&= \frac{n(n-1)}{2}\sigma_n^2
\end{aligned}
$$

because $h_n(\cdot,\cdot)$ is centered. From these two results the first condition in equation (5.8) is established. For the second calculate

$$
\begin{aligned}
\mathbb{E}\, I_{\mathcal{K}n,1}^4 = \sum_{i_1<i_2} \mathbb{E}\, h_n(W_{i_1}, W_{i_2})^4 &+ 3 \sum_{i_1<i_2} \sum_{\substack{i_3<i_4 \\ (i_3,i_4)\neq(i_1,i_2)}} \mathbb{E}\, h_n(W_{i_1}, W_{i_2})^2 h_n(W_{i_3}, W_{i_4})^2 \\
&+ 24 \sum_{i_1<i_2} \sum_{i_3\neq i_1,i_2} \mathbb{E}\, h_n(W_{i_1}, W_{i_2})^2 h_n(W_{i_1}, W_{i_3}) h_n(W_{i_2}, W_{i_3}) \\
+ 3 \sum_{i_1} \sum_{i_2\neq i_1} \sum_{i_3\neq i_1,i_2} \sum_{i_4\neq i_1,i_2,i_3} & \mathbb{E}\, h_n(W_{i_1}, W_{i_2}) h_n(W_{i_2}, W_{i_3}) h_n(W_{i_3}, W_{i_4}) h_n(W_{i_4}, W_{i_1}) \quad (5.9)
\end{aligned}
$$

where all vanishing terms (with $\mathbb{E}\, h_n(W_{i_1}, W_{i_2}) = 0$) are omitted. To show the second condition, the remaining terms have to be calculated. Starting with the denominator, we have to calculate

$$
\sigma_n^2 = \mathbb{E}\, h_n(W_1, W_2)^2. \qquad (5.10)
$$

Resolving the square and changing variables[4] to $\tilde{v} = (v - v_1)/h$ together with expanding $a(\cdot)$ and $f(\cdot)$ yields

$$
\begin{aligned}
\sigma_n^2 = \frac{4}{n^4 h^{2(d_X+d_Z)}} \sum_{k,k'} \iint K(\tilde{v}) \frac{y_1^k - \mu^k(v_1)}{f(v_1)} \\
\times K(\tilde{v} + (v_1 - v_2)/h) \frac{y_2^k - \mu^k(v_2)}{f(v_1)} a(v_1) f(v_1)\, \mathrm{d}\tilde{v} \\
\times \int K(\tilde{v}) \frac{y_1^{k'} - \mu^{k'}(v_1)}{f(v_1)} K(\tilde{v} + (v_1 - v_2)/h) \frac{y_2^{k'} - \mu^{k'}(v_2)}{f(v_1)} a(v_1) f(v_1)\, \mathrm{d}\tilde{v} \\
\times f(y_1, v_1) f(y_2, v_2)\, \mathrm{d}y_1\, \mathrm{d}v_1\, \mathrm{d}y_2\, \mathrm{d}v_2 (1 + O(h))
\end{aligned}
$$

[4]Here the notation is simplified. As v_1 is $d_X + d_Z$-dimensional one has to apply $d_X + d_Z$ substitutions.

Now substitute $\tilde{\tilde{v}} = (v_1 - v_2)/h$ to obtain

$$
\begin{aligned}
&= \frac{4(\kappa_*)^{d_X+d_Z}}{n^4 h^{d_X+d_Z}} \sum_{k,k'} \int (y_1^k - \mu^k(v_1))(y_2^k - \mu^k(v_1))(y_1^{k'} - \mu^{k'}(v_1))(y_2^{k'} - \mu^{k'}(v_1)) \\
&\qquad \times \Big(\frac{a(v_1)}{f(v_1)}\Big)^2 f(y_1, v_1) f(y_2, v_1)\, \mathrm{d}y_1\, \mathrm{d}y_2\, \mathrm{d}v_1 (1 + O(h)) \\
&= \frac{4(\kappa_*)^{d_X+d_Z}}{n^4 h^{d_X+d_Z}} \sum_{k,k'} \int \Big(\int (y_1^k - \mu^k(v_1))(y_1^{k'} - \mu^{k'}(v_1)) \frac{f(y_1, v_1)}{f(v_1)}\, \mathrm{d}y_1\Big)^2 a(v_1)^2\, \mathrm{d}v_1 \\
&\qquad\qquad \times (1 + O(h)) \\
&= \frac{2}{n^4 h^{d_X+d_Z}} \Sigma_{\mathcal{K}}^2 (1 + O(h)).
\end{aligned}
$$

Similar calculations show that

$$
\begin{aligned}
\mathbb{E}\, h_n(W_1, W_2)^4 &= O(n^{-8} h^{-3(d_X+d_Z)}) \\
\mathbb{E}\, h_n(W_1, W_2)^2 h_n(W_1, W_3)^2 &= O(n^{-8} h^{-2(d_X+d_Z)}) \\
\mathbb{E}\, h_n(W_1, W_2)^2 h_n(W_1, W_3) h_n(W_2, W_3) &= O(n^{-8} h^{-2(d_X+d_Z)}) \\
\mathbb{E}\, h_n(W_1, W_2) h_n(W_2, W_3) h_n(W_3, W_4) h_n(W_1, W_4) &= O(n^{-8} h^{-(d_X+d_Z)}).
\end{aligned}
$$

Using combinatorial arguments it can be established from equation (5.9) that $\mathbb{E}\, I_{\mathcal{K}n,1}^4$ is asymptotically dominated by terms with $\mathbb{E}\, h_n(W_1, W_2)^2 h_n(W_3, W_4)^2 = (\mathbb{E}\, h_n(W_1, W_2)^2)^2$. Therefore the second condition in equation 5.8 is fulfilled as

$$
\frac{\mathbb{E}\, I_{\mathcal{K}n,1}^4}{(\mathbf{var}\, I_n)^2} = \frac{12 n^{-4} h^{-2(d_X+d_Z)} \Sigma_{\mathcal{K}}^4 (1 + o(1))}{(2 n^{-2} h^{-(d_X+d_Z)} \Sigma_{\mathcal{K}}^2 (1 + o(1)))^2} \longrightarrow 3
$$

and weak convergence of $I_{\mathcal{K}n,1}$ is established.

Proof of (5.6) The expected value of the test statistic is given by

$$
\mathbb{E}\, I_{\mathcal{K}n,2} = \frac{1}{n} \sum_{k=1}^{d_Y} \iint \Big(K_h(v - v_1) \frac{y_1^k - \mu^k(v_1)}{f(v)}\Big)^2 a(v) f(v)\, \mathrm{d}v f(y_1, v_1)\, \mathrm{d}y_1\, \mathrm{d}v_1.
$$

Changing variables and expanding yields

$$
\begin{aligned}
&= \frac{\kappa_0^2}{n h^{d_X+d_Z}} \sum_{k=1}^{d_Y} \int (y_1^k - \mu^k(v_1))^2 \frac{a(v_1)}{f(v_1)} f(y_1, v_1)\, \mathrm{d}v_1 (1 + O(h)) \\
&= n^{-1} h^{-(d_X+d_Z)} B_{\mathcal{K}} (1 + O(h^r)).
\end{aligned}
$$

Convergence in probability follows from Markov's inequality with second moments, which requires to calculate

$$
\frac{1}{n^4} \Big(\int \sum_{k=1}^{d_Y} \big(K_h(v - v_1)(y_1^k - \mu^k(v_1))\big)^2 \frac{a(v)}{f(v)}\, \mathrm{d}v\Big)^2 f(y_1, v_1)\, \mathrm{d}y_1\, \mathrm{d}v_1.
$$

Changing variables as before results in

$$\frac{\kappa_0^2}{n^4 h^{2(d_X+d_Z)}} \sum_{k,k'} \int (y_1^k - \mu^k(v_1))^2 (y_1^{k'} - \mu^{k'}(v_1))^2 \frac{a(v_1)^2}{f(v_1)^2} f(y_1, v_1)\, \mathrm{d}y_1\, \mathrm{d}v_1 (1 + o(1)),$$

which is bounded by Assumption 2. In total this yields

$$\mathbb{E}\, I_{\mathcal{K}n,2}^2 = O(n^{-3} h^{-2(d_X+d_Z)}) = o(n^{-2} h^{-(d_X+d_Z)})$$

and convergence in probability of $I_{\mathcal{K}n,2}$ follows.

Proof of (5.7) For this statement we will restrict to the case when $d_Y = 1$. Then convergence in probability has to be shown for

$$\Delta_{\mathcal{K}n} = \frac{1}{n^3} \sum_{i,j,k} \gamma_n(W_i, W_j, W_k),$$

where

$$\gamma_n(W_i, W_j, W_k) = \tilde{\gamma}_n(W_i, W_j, W_k) - \int \tilde{\gamma}_n(W_i, W_j, w) f(w)\, \mathrm{d}w,$$

with

$$\tilde{\gamma}_n(W_i, W_j, W_k) = K_h(V_k - V_i) \frac{Y_i - \mu^1(V_i)}{f(V_k)} a(V_k) K_h(V_k - V_j) \frac{Y_j - \mu^1(V_j)}{f(V_k)} a(V_k).$$

First we show that the expectation tends to zero

$$\mathbb{E}\, \Delta_{\mathcal{K}n} = \frac{1}{n^3} \sum_{i,j,k} \mathbb{E}\, \gamma_n(W_i, W_j, W_k) = o(n^{-1} h^{-(d_X+d_Z)/2}),$$

where only the cases $i = k \neq j, j = k \neq i$ and $i = j = k$ have to be considered, all others have expectation zero. In the remaining cases, two (resp. one) substitution can be applied and their total contribution is $O(n^{-1} h^{2(d_X+d_Z)} + n^{-2} h^{d_X+d_Z})$.
Then, Markov's inequality is applied with the second moments and we have to investigate

$$\mathbb{E}\, \Delta_{\mathcal{K}n}^2 = \frac{1}{n^6} \sum_{ijk} \mathbb{E}\, \gamma_n(W_i, W_j, W_k)^2 + \frac{2}{n^6} \sum_{ijk} \sum_{i'j'k'} \mathbb{E}\, \gamma_n(W_i, W_j, W_k) \gamma_n(W_{i'}, W_{j'}, W_{k'}).$$

The covariance parts vanish, whenever $k \neq k'$. If $k = k'$ the covariance terms are zero by the conditional independence of the error terms, in all cases where $i \neq i'$ or $j \neq j'$. For the remaining cases we have to distinguish if the number of different indices is $N = 2, 3$. Then, the overall contribution of these terms is $O(n^{N-6} h^{-4(d_X+d_Z)} h^{N(d_X+d_Z)}) = o(n^{-2} h^{-(d_X+d_Y)})$.
Next, consider the variance terms. If there are three different indices, two changes of variables can be applied and the overall contribution is $O(n^{-3} h^{-2(d_X+d_Z)}) = o(n^{-2} h^{-(d_X+d_Z)})$. If

there are two different indices, one change of variables can be applied and we obtain terms of order $O(h^{-3(d_X+d_Z)})$ with a total contribution of $O(n^{-4}h^{-3(d_X+d_Z)}) = o(n^{-2}h^{-(d_X+d_Z)})$. If $i = j = k$ one change of variables is still possible and the contribution is $O(n^{-5}h^{-3(d_X+d_Z)}) = o(n^{-2}h^{-(d_X+d_Z)})$. This completes the proof of equation (5.7).

Convergence in Probability of $\hat{\Gamma}_{\mathcal{K}3}$ For the third term in (5.1) it holds that

$$\begin{aligned} |\hat{\Gamma}_{\mathcal{K}3}| &\leq \max_{k=1,\dots,d_Y} \sup_{x\in\mathcal{A}} |m^k(X_j) - \hat{m}^k_{\tilde{h}}(X_j)|^2 \sup_{v\in\mathcal{A}} |a(v)| \\ &= O_P(\tilde{h}^{2r} + \frac{\log n}{n\tilde{h}^{d_X}}) \\ &= o_P\big(n^{-1}h^{-(d_X+d_Z)/2}\big) \end{aligned}$$

under Assumption 4.3.

Convergence in Probability of $\hat{\Gamma}_{\mathcal{K}4}$ The non-zero parts are given by

$$\begin{aligned} \hat{\Gamma}_{\mathcal{K}4} = \frac{1}{n}\sum_{k=1}^{d_Y}\sum_{i=1}^{n}\Big(\frac{1}{n}\sum_{j=1}^{n} K_h(V_i - V_j)\frac{Y_j^k - \mu^k(V_j)}{\hat{f}_h(V_i)}\Big) \\ \times \Big(\frac{1}{n}\sum_{j'=1}^{n} K_h(V_i - V_{j'})\frac{m^k(X_{j'}) - \hat{m}^k_{\tilde{h}}(X_{j'})}{\hat{f}_h(V_i)}\Big) A_i \\ = \sum_{i,j,j'} \gamma_{ijj'}. \end{aligned}$$

Because

$$\mathbb{E}\,\varepsilon_j^k\big(m^k(V_{j'}) - \hat{m}_{\tilde{h}}(V_{j'})\big) \mid V_1,\dots,V_n) = n^{-1}K_{\tilde{h}}(V_j - V_{j'})\sigma^2(V_j)$$

we have that

$$\mathbb{E}\,\hat{\Gamma}_{\mathcal{K}4} = O(n^{-1}) = o(n^{-1}h^{-(d_X+d_Z)/2}).$$

It follows from similar considerations as done to show (5.7) that

$$\mathbb{E}\,\hat{\Gamma}^2_{\mathcal{K}4} = o(n^{-2}h^{-(d_X+d_Z)}).$$

This completes the proof of the theorem □

Proof of Theorem 2

Under H_{1n} the decomposition (5.1) remains valid and the asymptotic analysis of $\hat{\Gamma}_{\mathcal{K}1}$ and $\hat{\Gamma}_{\mathcal{K}3}$ is unchanged. However $\hat{\Gamma}_{\mathcal{K}2}$ is not zero any longer. If it holds that $\mu(x, z) = m(x) + \varepsilon_n(x, z)$, we have that

$$\hat{\Gamma}_{\mathcal{K}2} = \frac{1}{n}\sum_{k=1}^{d_Y}\sum_{i=1}^{n}\Big(\frac{1}{n}\sum_{j=1}^{n} K_h(V_i - V_j)\frac{\varepsilon_n^k(V_j)}{f(V_i)}\Big)^2 A_i(1+o_P(1))$$
$$= \sum_{k=1}^{d_Y}\int\Big(\frac{1}{n}\sum_{j=1}^{n} K_h(v - V_j)\frac{\varepsilon_n^k(V_j)}{f(v)}\Big)^2 a(v)\,\mathrm{d}v + o_P(n^{-1}h^{(d_X+d_Z)/2}).$$

This follows from similar calculations as to show (5.7). Omitting the lower order terms it holds that

$$nh^{(d_X+d_Z)/2}\hat{\Gamma}_{\mathcal{K}2} = \int K(u)^2\,\mathrm{d}u\sum_{k=1}^{d_Y}\frac{1}{n}\sum_{j=1}^{n}\Big(\frac{\varepsilon_n^k(V_j)}{f(V_j)}\Big)^2 a(V_j) + o_P(1)$$
$$\xrightarrow{P} h^{-(d_X+d_Z)/2}B_L + o_P(1).$$

The last convergence holds by assumption if $\lambda_n = O(nh^{(d_X+d_Z)/2}$. In particular for any fixed alternative, the convergence does not apply and $nh^{(d_X+d_Z)/2}\hat{\Gamma}_{\mathcal{K}2} = O(n)$ and diverges. This yields consistency of the test statistic. □

Proof of Theorem 3

In the proof of this theorem we use the notation $\mathbb{E}^*$ and $\mathbf{var}^*$ to denote expectation and variance conditional on the data. Decompose

$$\hat{\Gamma}^*_{\mathcal{K}} = \frac{1}{n}\sum_{k=1}^{d_Y}\sum_{i=1}^{n}\Big(\frac{1}{n}\sum_{j=1}^{n} K_h(X_i - X_j, Z_i - Z_j)\frac{Y_j^{k,*} - \hat{m}_{\tilde{h}}^{k,*}(X_i)}{\hat{f}_h(X_j, Z_j)}\Big)^2 A_i$$
$$= \frac{1}{n}\sum_{k=1}^{d_Y}\sum_{i=1}^{n}\Big(\frac{1}{n}\sum_{j=1}^{n} K_h(X_i - X_j, Z_i - Z_j)$$
$$\times\Big(\frac{\varepsilon^{k,*}}{\hat{f}_h(X_j, Z_j)} + \frac{\hat{m}_{\tilde{h}}^{k}(X_i) - \hat{m}_{\tilde{h}}^{k,*}(X_i)}{\hat{f}_h(X_j, Z_j)}\Big)\Big)^2 A_i$$
$$= (I^*_{\mathcal{K}n} + \Delta^*_{\mathcal{K}n})(1+o_P(1)) + \Gamma^*_{\mathcal{K}3} + \Gamma^*_{\mathcal{K}4},$$

where $I^*_{\mathcal{K}n}$ and $\Delta^*_{\mathcal{K}n}$ are defined as in (5.2) and (5.3) by replacing $Y_j^k - \mu^k(X_j)$ with $\varepsilon_j^{k,*}$. $\Gamma^*_{\mathcal{K}3}$ can be bounded by showing that

$$\sup_{x\in\mathcal{A}}|\hat{m}_{\tilde{h}}^{k}(x) - \hat{m}_{\tilde{h}}^{k,*}(x)| = O_P\Big(\tilde{h}^r + \Big(\frac{\log n}{nh^{d_X}}\Big)^{1/2}\Big). \tag{5.11}$$

Decomposing $I^*_{\mathcal{K}n}$ as in equation (5.4) into $I^*_{\mathcal{K}n,1}$ and $I^*_{\mathcal{K}n,2}$ it remains to show that

$$nh^{(d_X+d_Z)/2}I^*_{\mathcal{K}n,1} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma^2_{\mathcal{K}}), \tag{5.12}$$

conditional on the data with probability tending to one and

$$nh^{(d_X+d_Z)/2}I^*_{\mathcal{K}n,2} - h^{-(d_X+d_Z)/2}B_{\mathcal{K}} \xrightarrow{P} 0 \tag{5.13}$$

$$nh^{(d_X+d_Z)/2}\Delta^*_{\mathcal{K}n} \xrightarrow{P} 0. \tag{5.14}$$

Then the statement of the theorem follows.

Proof of (5.11) First note that

$$\begin{aligned}\sup_{x\in\mathcal{A}}|\hat{m}^k_{\tilde{h}}(x) - \hat{m}^{k,*}_{\tilde{h}}(x)| &= \sup_{x\in\mathcal{A}}|(\hat{f}^k_{\tilde{h}}(x))^{-1}\frac{1}{n}\sum_{i=1}^n K_{\tilde{h}}(x-X_i)(Y^k_i - Y^{k,*}_i)| \\ &\le \sup_{x\in\mathcal{A}}|m^k(x) - \hat{m}^k_{\tilde{h}}(x)| + \sup_{x\in\mathcal{A}}|(\hat{f}^k_{\tilde{h}}(x))^{-1}\frac{1}{n}\sum_{i=1}^n K_{\tilde{h}}(x-X_i)\varepsilon^{k,*}_i| \\ &\quad + \sup_{x\in\mathcal{A}}|(\hat{f}^k_{\tilde{h}}(x))^{-1}\frac{1}{n}\sum_{i=1}^n K_{\tilde{h}}(x-X_i)\varepsilon^k_i|.\end{aligned}$$

The first term already has the desired rate. Because $\hat{f}^k_{\tilde{h}}(x)$ is consistent and $f(x)$ is bounded from below on $\mathcal{A}$ further analysis can be restricted to the numerator. Since the analysis of the second and the third term in analogous, we concentrate on the second term. First, a truncation argument is applied. Define $\tilde{\varepsilon}^{k,*}_i = \mathbf{1}_{\{\varepsilon^{k,*}_i \le n\tilde{h}^{d_X}\}}$, which allows to decompose

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n K_{\tilde{h}}(x-X_i)\varepsilon^{k,*}_i &= \frac{1}{n}\sum_{i=1}^n K_{\tilde{h}}(x-X_i)\tilde{\varepsilon}^{k,*}_i \\ &\quad + \frac{1}{n}\sum_{i=1}^n K_{\tilde{h}}(x-X_i)\varepsilon^{k,*}_i\mathbf{1}_{\{\varepsilon^{k,*}_i > n\tilde{h}^{d_X}\}}.\end{aligned} \tag{5.15}$$

Starting with the second term, note that it holds that $\mathbb{E}\,|\varepsilon^{k,*}_i\mathbf{1}_{\{\varepsilon^{k,*}_i > n\tilde{h}^{d_X}\}}| = O(n^{-2}\tilde{h}^{-2d_X})$, because the forth moment of $\varepsilon^{k,*}_i$ is finite. Then, the second term on the right side of (5.15) can be bounded with Markov's inequality with first moments

$$\begin{aligned}\mathbb{E}\Big|\frac{1}{n}\sum_{i=1}^n K_{\tilde{h}}(x-X_i)\varepsilon^{k,*}_i\mathbf{1}_{\{\varepsilon^{k,*}_i > n\tilde{h}^{d_X}\}}\Big| &\le \mathbb{E}\,|K_{\tilde{h}}(x-X_1)\varepsilon^{k,*}_1\mathbf{1}_{\{\varepsilon^{k,*}_1 > n\tilde{h}^{d_X}\}}| \\ &\le \sup_u |K(u)|\,\mathbb{E}\,|\varepsilon^{k,*}_i\mathbf{1}_{\{\varepsilon^{k,*}_i > n\tilde{h}^{d_X}\}}|(1+O(\tilde{h})) \\ &= O(n^{-2}\tilde{h}^{-2d_X}),\end{aligned}$$

from which the desired rate follows.

Finally, we turn to the first term in (5.15). Covering the compact set $\mathcal{A}$ with N cubes $\mathcal{A}_l =$

$\{x \mid \|x - x_l\| < \eta_N\}, l = 1, \dots, N, \eta_N = O(N^{-1/d_X})$ we write

$$\sup_{x \in \mathcal{A}} \Big| \frac{1}{n} \sum_{i=1}^n K_{\tilde{h}}(x - X_i) \tilde{\varepsilon}_i^{k,*} \Big| \le \max_l \Big| \frac{1}{n} \sum_{i=1}^n K_{\tilde{h}}(x_l - X_i) \tilde{\varepsilon}_i^{k,*} \Big| + \max_l \sup_{x \in \mathcal{A}_l} \Big| \frac{1}{n} \sum_{i=1}^n (K_{\tilde{h}}(x - X_i) - K_{\tilde{h}}(x_l - X_i)) \tilde{\varepsilon}_i^{k,*} \Big|. \quad (5.16)$$

Using the Lipschitz-continuity of the kernel, one directly obtains that the second term on the right hand side is of $O_P(\eta_N \tilde{h}^{-d_X/2} n^{-1/2}) = o_P(n^{-1/2} \tilde{h}^{d_X/2} (\log n)^{1/2})$. The first term on the is bounded using Bonferroni's inequality first and then Bernstein's inequality

$$\mathbb{P}\Big(\Big| \frac{1}{n} \sum_{i=1}^n K_{\tilde{h}}(x_l - X_i) \tilde{\varepsilon}_i^{k,*} \Big| > \Big(\frac{\log n}{n \tilde{h}^{d_X}} \Big)^{1/2} \frac{c_1}{2} \Big) \le 2 \exp\Big(- \frac{c_1^2 (\log n)/(4 n \tilde{h}^{d_X})}{4 \sum_{i=1}^n \mathbb{E}(\frac{1}{n} K_{\tilde{h}}(x - X_i) \tilde{\varepsilon}_i^{k,*})^2 + c_2 (\frac{\log n}{n^3 \tilde{h}^{3d_X}})^{3/2}} \Big),$$

where c_2 is the constant arising from Cramer's conditions on the distribution of $\tilde{\varepsilon}_i^{k,*}$. It follows from standard arguments that $\sum_{i=1}^n \mathbb{E}(\frac{1}{n}(K_{\tilde{h}}(x - X_i)\tilde{\varepsilon}_i^{k,*}$ and so we get that

$$\mathbb{P}\Big(\Big| \frac{1}{n} \sum_{i=1}^n K_{\tilde{h}}(x - X_i) \tilde{\varepsilon}_i^{k,*} \Big| > \Big(\frac{\log n}{n \tilde{h}^{d_X}} \Big)^{1/2} \frac{c}{2} \Big) \le O(n^{-1}).$$

Then, for $N = o(n)$ the desired rate of convergence is obtained.

Proof of (5.12) To derive the asymptotic distribution of

$$I^*_{\mathcal{K}n,1} = \sum_{i<j} h_n(W_i^*, W_j^*),$$

given the data with probability tending to one, again the limit theorem by de Jong (1987) will be applied. This is done by showing that the conditions hold with probability tending to one, i.e.

$$\frac{\max_{1 \le i \le n} \sum_{j=1}^n \mathbb{E}^* h_n(W_i^*, W_j^*)^2}{\mathbf{var}^* I^*_{\mathcal{K}n1}} \xrightarrow{P} 0 \quad \text{and} \quad \frac{E^*(I^*_{\mathcal{K}n,1})^4}{(\mathbf{var}^* I^*_{\mathcal{K}n,1})^2} \xrightarrow{P} 3.$$

Here,

$$h_n(W_i^*, W_j^*) = \frac{2}{n^2} \int K_h(v - V_i) K_h(v - V_j) \frac{a(v)}{f(v)} \, \mathrm{d}v \sum_{k=1}^{d_Y} \varepsilon_i^{k,*} \varepsilon_j^{k,*}.$$

First we analyze

$$\begin{aligned}
\mathbb{E}^* h_n(W_i^*, W_j^*)^2 &= \frac{4}{n^4}\Big(\int K_h(v-V_i)K_h(v-V_j)\frac{a(v)}{f(v)}\,\mathrm{d}v\Big)^2 \\
&\quad\times\Big(\sum_{k=1}^{d_Y}(Y_i^k - \hat{m}_{\tilde{h}}^k(X_i))(Y_j^k - \hat{m}_{\tilde{h}}^k(X_j))\Big)^2 \\
&= \frac{4}{n^4}\Big(\int K_h(v-V_i)K_h(v-V_j)\frac{a(v)}{f(v)}\,\mathrm{d}v\Big)^2 \\
&\quad\times\Big(\sum_{k=1}^{d_Y}(Y_i^k - \mu^k(V_i))(Y_j^k - \mu^k(V_j))\Big)^2(1+O_P\Big(\tilde{h}^r + \Big(\frac{\log n}{n\tilde{h}^{d_X}}\Big)^{1/2}\Big)) \\
&= h_n(W_i, W_j)^2 + o_p(n^{-4}h^{-2(d_X+d_z)}).
\end{aligned} \tag{5.17}$$

This holds because under H_0 we have that $m^k(X_i) = \mu^k(X_i, Z_i)$ almost surely. Starting with the numerator, we utilize the conditional independence of the bootstrap residuals to see that

$$\mathbf{var}^* I_{\mathcal{K}n,1} = \sum_{i<j}\mathbb{E}^* h_n(W_i^*, W_j^*).$$

To bound this in probability, apply Markov's inequality with the first moment

$$\mathbb{E}\Big|\sum_{i<j}\mathbb{E}^* h_n(W_i^*, W_j^*)^2\Big| = \sum_{i<j}\mathbb{E}\, h_n(W_i^*, W_j^*)^2 = n^{-2}h^{-(d_X+d_Z)}2\Sigma_{\mathcal{K}}^2(1+o(1)),$$

from which it follows that

$$\mathbf{var}^* I_{\mathcal{K}n,1} \xrightarrow{P} \mathbf{var}\, I_{\mathcal{K}n,1}.$$

This is now used to show the first condition. Together with (5.8) and (5.17) we obtain

$$\begin{aligned}
&\frac{\max_{i=1,\dots,n}\sum_{j=1,j\neq i}^n \mathbb{E}^* h_n(W_i^*, W_j^*)^2}{\mathbf{var}\, I_{\mathcal{K}n,1}} \\
&= \frac{\max_{i=1,\dots,n}\sum_{j=1,j\neq i}^n h_n(W_i, W_j)^2 + O_P(n^{-3}(\tilde{h}^r + (\log n/(n\tilde{h}^{d_X}))^{1/2}))}{\mathbf{var}\, I_{\mathcal{K}n,1}} \\
&= \frac{\max_{i=1,\dots,n}\sum_{j=1,j\neq i}^n h_n(W_i, W_j)^2}{\mathbf{var}\, I_{\mathcal{K}n,1}} + O_P(n^{-1}(\tilde{h}^r + (\log n/(n\tilde{h}^{d_X}))^{1/2})) \\
&= o_P(1).
\end{aligned}$$

For the second condition we again use the convergence of the denominator. Then using the first moment to bound the probability leads to similar calculations as done in the proof of equation (5.5).

Proof of (5.13) The proof of equation (5.13) consists of using iterated expectations and use there the same calculations as to proof equation (5.12).

Proof of (5.14) As $\mathbb{E}^* \varepsilon_j^{k,*} = 0$ the same arguments as for $\Delta_{\mathcal{K}n}$ remain to hold for $\Delta^*_{\mathcal{K}n}$. $\square$

Proof of Theorem 4

From Masry (1996) it is known that

$$\sup_{v\in\mathcal{A}} |\hat{S}(v) - f(v)M| = O_P\Big(h^2 + \Big(\frac{\log n}{nh^{d_X+d_Y}}\Big)^{1/2}\Big),$$

therefore we can write

$$\begin{aligned}
\hat{\Gamma}^L_{\mathcal{K}} &= \frac{1}{n}\sum_{k=1}^{d_Y}\sum_{i=1}^{n}(\lfloor f(V_i)^{-1}M^{-1}(\hat{T}^k(V_i) - \tilde{T}^k(V_i))\rfloor_1)^2 A_i(1+o_P(1)) \\
&= \frac{1}{n}\sum_{k=1}^{d_Y}\sum_{i=1}^{n}\Big(\frac{1}{n}\sum_{0\le|\mathbf{j}|\le p}\kappa_{\mathbf{j}}^{-1}\sum_{l=1}^{n}\Big(\frac{V_l - V_i}{h}\Big)^{\mathbf{j}} K_h(V_l - V_i)\frac{Y_l^k - \hat{m}^{k,L}_{\tilde{h}}(X_l)}{f(V_i)}\Big)^2 \\
&= \hat{\Gamma}^L_{\mathcal{K}1} + \hat{\Gamma}^L_{\mathcal{K}2} + \hat{\Gamma}^L_{\mathcal{K}3} + \hat{\Gamma}^L_{\mathcal{K}4},
\end{aligned}$$

where we decompose according to $Y_l^k - \hat{m}^{k,L}_{\tilde{h}}(X_l) = Y_l^k - \mu^k(V_l) + \mu^k(V_l) - m^k(V_l) + m^k(X_l) - \hat{m}^{k,L}_{\tilde{h}}(X_l)$ and transfer all cross terms to $\hat{\Gamma}^L_{\mathcal{K}4}$. Then $\hat{\Gamma}^L_{\mathcal{K}2} = 0$ almost surely under H_0. And $\hat{\Gamma}^L_{\mathcal{K}3} = O_P(h^{2r} + \log n/(nh^{d_X}))$ by applying results from Masry (1996) for the density estimator and the local linear estimator. Next, we decompose

$$\hat{\Gamma}^L_{\mathcal{K}1} = I^L_{\mathcal{K}n,1} + I^L_{\mathcal{K}n,2} + \Delta^L_{\mathcal{K}n},$$

where the quantities are given as in (5.2), (5.3) and (5.4) and the kernel is replaced by

$$\tilde{K}_h(u) = \sum_{1\le\mathbf{j}\le p}\Big(\frac{u}{h}\Big)^{\mathbf{j}}\kappa_{\mathbf{j}}^{-1}K_h(u).$$

Since this kernel satisfies the assumptions which are necessary to show (5.5)–(5.7) (note that higher order properties of the kernel are not used there), the statement of the theorem follows.

Proof of Theorem 5

In this case we can decompose the test statistic into $\hat{\Gamma}^S_{\mathcal{K}} = \hat{\Gamma}^S_{\mathcal{K}1} + \hat{\Gamma}^S_{\mathcal{K}2} + \hat{\Gamma}^S_{\mathcal{K}3} + \hat{\Gamma}^S_{\mathcal{K}4} + \hat{\Gamma}^S_{\mathcal{K}5}$, where $\hat{\Gamma}^S_{\mathcal{K}1} = \hat{\Gamma}_{\mathcal{K}1}$,

$$\begin{aligned}
\hat{\Gamma}^S_{\mathcal{K}2} = \frac{1}{n}\sum_{k=1}^{d_Y}\sum_{i=1}^{n}\Big(\frac{1}{n}\sum_{j=1}^{n}&\frac{K_h(X_i - X_j, Z_i - Z_j)}{\hat{f}_h(X_i, Z_i)} \\
&\times(\mu^k(X_j, Z_j) - m^k(X_j) - G^k(Z_j,\theta))\Big)^2 A_i
\end{aligned}$$

$$\hat{\Gamma}^S_{\mathcal{K}3} = \frac{1}{n}\sum_{k=1}^{d_Y}\sum_{i=1}^{n}\Big(\frac{1}{n}\sum_{j=1}^{n}\frac{K_h(X_i - X_j, Z_i - Z_j)}{\hat{f}_h(X_i, Z_i)}(m^k(X_j) - \hat{m}^k_{\tilde{h}}(X_j))\Big)^2 A_i$$
$$\hat{\Gamma}^S_{\mathcal{K}4} = \frac{1}{n}\sum_{k=1}^{d_Y}\sum_{i=1}^{n}\Big(\frac{1}{n}\sum_{j=1}^{n}\frac{K_h(X_i - X_j, Z_i - Z_j)}{\hat{f}_h(X_i, Z_i)}(G^k(Z_j, \theta) - G^k(Z_j, \hat{\theta})\Big)^2 A_i$$

and $\hat{\Gamma}^S_{\mathcal{K}5}$ contains all cross terms. Under H_{0S} we have that $\hat{\Gamma}^S_{\mathcal{K}2} = 0$ almost surely and the other two terms are bounded by uniform convergence rates. Start with

$$\begin{aligned}
|\hat{\Gamma}^S_{\mathcal{K}3}| &\le \max_{k=1,\dots,d_Y}\sup_{x\in\mathcal{A}_X}|m^k(x) - \hat{m}^k_{\tilde{h}}(x)|^2 \sup_{(x,z)\in\mathcal{A}}|a(x,z)| \\
&\le \sup_{(x,z)\in\mathcal{A}}|a(x,z)|\Big(\max_{k=1,\dots,d_Y}\sup_{x\in\mathcal{A}_X}|m^k(x) - \tilde{m}^k_{\tilde{h}}(x)|^2 \\
&\quad + \max_{k=1,\dots,d_Y}\sup_{z\in\mathcal{A}_Z}|G^k(z,\theta) - G^k(z,\hat{\theta})|^2\Big) \\
&= o_P(n^{-1}h^{(d_X+d_Z)/2}).
\end{aligned}$$

The quantity $\tilde{m}^k_{\tilde{h}}(X_j)$ denotes a nonparametric regression of the unobserved variable $Y_i - G(Z_i, \theta)$ on X_i. The standard uniform convergence rate holds for this estimator and by our assumptions it converges faster than the test statistic. For the parametric function $G(z, \theta)$ the convergence rate was assumed. From this assumption we also obtain

$$|\hat{\Gamma}^S_{\mathcal{K}4}| = o_P(n^{-1}h^{(d_X+d_Z)/2}).$$

The asymptotic distribution of $\hat{\Gamma}_{\mathcal{K}1}$ was derived in the proof of Theorem 1. □

Proof of Theorem 6

For dependent data, decomposition (5.1) still applies and under H_0 it holds that $\hat{\Gamma}_{\mathcal{K}2} = 0$. Because β-mixing implies α-mixing, the results in Masry (1996) hold under Assumption 7. This means, we have that

$$\sup_{x\in\mathcal{A}}|m^k(x) - \hat{m}_{\tilde{h}}(x)| = O_P(\tilde{h}^r + \Big(\frac{\log n}{n\tilde{h}^{d_X}}\Big)^{1/2})$$

and the same rate holds for the kernel density estimator. Therefore it remains to analyze $\hat{\Gamma}_{\mathcal{K}1}$ and to show (5.5)–(5.7) for the dependent case.

Proof of (5.5) To derive the asymptotic distribution, we regard $I_{\mathcal{K}n,1}$ still as U-Statistic, and apply Theorem 2.1 in Fan and Li (1999). To apply this central limit theorem a large number of assumptions have to be checked. For brevity we concentrate on those that influence the rates.

Denote with $(\tilde{W}_i)_{i=1,\dots,n}$ a sequence of independent and identically distributed random variables with the same marginal distribution as $(W_i)_{i=1,\dots,n}$

$$\frac{m}{n^{3/2}}\frac{\mathbb{E}\,h_n(\tilde{W}_1,\tilde{W}_2)^4}{(\mathbb{E}\,h_n(\tilde{W}_1,\tilde{W}_2)^2)^2} = O\Big(\frac{m}{n^{3/2}h}\Big)$$
$$m^4\frac{\max_{t>1}\mathbb{E}(\int h_n(w,W_1)h_n(w,W_t)f(w)\,\mathrm{d}w)^2}{(\mathbb{E}\,h_n(\tilde{W}_1,\tilde{W}_2)^2)^2} = O(m^4h)$$
$$n^2\beta(m)^{1-1/\nu}\frac{m^2+n^2\beta(m)^{1-1/\nu}}{(\mathbb{E}\,h_n(\tilde{W}_1,\tilde{W}_2)^2)^2} = O\big(n^6h^2(m^2\beta(m)^{1-1/\nu}+n^2\beta(m)^{2-2/\nu})\big)$$

Together with the assumptions on the number of existing moments of Y and the kernel function ($\mathbb{E}\,Y^{4\nu}<\infty, \kappa_0^{4\nu}$), this yields (5.5).

Proof of (5.6) It is easy to see that $\mathbb{E}\,I_{\mathcal{K}n,2}$ is unchanged. To show convergence in probability using the second moment of $I_{\mathcal{K}n,2}$, the covariances have to be bounded. Writing

$$I_{\mathcal{K}n,2} = \sum_{i=1}^{n} h_n'(X_i),$$

with

$$h_n'(W_i) = \frac{1}{nh^{-(d_X+d_Z)}}\sum_{k=1}^{d_Y}\int\Big(K(u,v)(Y_1^k-m^k(X_1))\Big)^2\frac{a(X_1+uh,Z_1+vh)}{f(X_1+uh,Z_1+vh)}\,\mathrm{d}u\,\mathrm{d}v.$$

We then use the covariance inequality for strongly dependent processes ($\nu>1$)

$$\mathbf{cov}(h_n'(W_i),h_n'(W_j)) \le c\big(\mathbb{E}(h_n'(W_1))^\nu\big)^{2/\nu}\beta(j-i)^{1-2/\nu}.$$

As $\big(\mathbb{E}(h_n'(W_1))^\nu\big)^{2/\nu} = O(n^{-2}h^{-2(d_X+d_Z)})$ (if $\mathbb{E}\,Y^{2\nu}<\infty$ and $\kappa_0^{2\nu}<\infty$) the convergence follows if $\sum_{i=1}^\infty\beta(i)^{1-2/\nu}<\infty$.

Proof of (5.7) To show that the expected value converges we use

$$|\,\mathbb{E}\,\gamma_n(W_i,W_j,W_k)| \le 4M^{1/\nu}\beta(\min\{i-k,j-k\})^{1-1/\nu},$$

where $M=\max\{\mathbb{E}\,\tilde{\gamma}_n(W_i,W_j,W_k)^\nu, \mathbb{E}\int\tilde{\gamma}_n(W_i,W_j,w)^\nu f(w)\,\mathrm{d}w\}$. (Lemma A.1 in Dette and Spreckelsen, 2004). Convergence in probability is shown by using the first absolute moment and Lemma A.0 in Fan and Li (1999) to obtain

$$\mathbb{E}\,|\gamma_n(W_i,W_j,W_k)| \le 4M^{1/\nu}\beta(\min\{i-k,j-k\})^{1-1/\nu},$$

with M as above. Some tedious calculations show that convergence in probability is established if $\mathbb{E}\,Y^{2\nu}<\infty$ and $\sum_{i=1}^\infty\beta(i)^{1-1/\nu}<\infty$, establishing the asymptotic distribution. □