

# Testing and Imposing Slutsky Symmetry in Nonparametric Demand Systems

Berthold R. Haag; Stefan Hoderlein; Krishna Pendakur\*

November 6, 2007

## Abstract

Maximization of utility implies that consumer demand systems have a Slutsky matrix which is everywhere symmetric. However, previous non- and semi-parametric approaches to the estimation of consumer demand systems do not give estimators that are restricted to satisfy this condition, nor do they offer powerful tests of this restriction. We use non-parametric modeling to test and impose Slutsky symmetry in a system of expenditure share equations over prices and expenditure. In this context, Slutsky symmetry is a set of nonlinear cross-equation restrictions on levels and derivatives of consumer demand equations. The key insight is that due to the differing convergence rates of levels and derivatives and due to the fact that the symmetry restrictions are linear in derivatives, both the test and the symmetry restricted estimator behave asymptotically as if these restrictions were (locally) linear. We establish large and finite sample properties of our methods, and show that our test has advantages over the only other comparable test. All methods we propose are implemented with Canadian micro-data. We find that our nonparametric analysis yields statistically significantly and qualitatively different results from traditional parametric estimators and tests.

**Keywords:** Demand System, Slutsky Symmetry, Rationality, Nonparametric Regression, Nonparametric Testing.

**JEL Classifications:** D12, C14, C13, C31, C52, D11.

---

\*University of Mannheim; University of Mannheim; Simon Fraser University. Corresponding author is Stefan Hoderlein, stefan\_hoderlein@yahoo.com, Department of Economics (VWL), University of Mannheim, L7 3-5, 68131 Mannheim, Germany. We acknowledge Martin Browning, Richard Blundell, Hide Ichimura, Arthur Lewbel, Oliver Linton and Enno Mammen for their help on this work. We also acknowledge the support of the Social Sciences and Humanities Research Council of Canada and Landesstiftung Baden-Württemberg.

# 1 Introduction

This paper considers the rationality restriction of Slutsky symmetry on a consumer demand system in a nonparametric setting. We provide a nonparametric test for Slutsky symmetry and derive its asymptotic distribution. In addition, we provide a local polynomial estimator of the Slutsky matrix under this restriction. Micro-level consumer surplus computed from a demand system is unique (path-independent) if and only if the Slutsky matrix is everywhere symmetric. Thus, Slutsky matrices which satisfy symmetry may be used for welfare analysis, such as consumer surplus and cost-of-living index calculations. We use nonparametric methods in order to avoid bias due to parametric misspecification, and to avoid confounding the test of symmetry with a test of parametric specification. This point can hardly be overemphasized, as we find evidence of both parametric misspecification and confounding of the test in our empirical work. The obvious price one has to pay for the benefit of not having to rely on functional form assumptions is the curse of dimensionality. However, in our application we do not find this to be a major problem. In addition, the analysis here can be seen as a building block for testing of and imposing symmetry in semiparametric and structured nonparametric models.

The Slutsky symmetry restriction is comprised of a set of nonlinear cross-equation restrictions on the levels and derivatives of expenditure-share equations. In parametric demand system estimation, the complexity of these restrictions and the complexity of share equations seem to go together — in complex demand systems, Slutsky symmetry is similarly complex.<sup>1</sup> Surprisingly, we find that in the nonparametric case, where share equations can be arbitrarily complex, Slutsky symmetry is relatively easy to test and impose. Because the estimated derivatives converge more slowly than estimated levels, and because the symmetry restrictions are linear in the derivatives, Slutsky symmetry behaves asymptotically like a linear restriction on the derivatives. Based on this insight, we provide a new test for symmetry and its asymptotic distribution, show the validity of the bootstrap for this application, and develop an approach to restricted estimation under symmetry.

Our test statistic is based on unrestricted estimators. It is the sum of  $L_2$  distances between all the Slutsky terms which are equal under symmetry. When we turn to our empirical example, we find that symmetry is strongly rejected by this test, which contrasts sharply with a rather weak

---

<sup>1</sup>In demand systems with nonlinear Engel curves, Slutsky symmetry is usually imposed with nonlinear cross-equation restrictions. Some demand systems, such as the Almost Ideal and its quadratic extension, have approximate forms in which symmetry is a set of linear restrictions. In addition, Blundell and Robin (1999) show that these same nonlinear demand systems may be estimated exactly under symmetry by iterative linear methods.

rejection from a parametric quadratic almost ideal (QAI) test. In addition, we are able to reject the QAI parametric model of expenditure share equations against a nonparametric alternative. Taken together, these results suggest the parametric test of symmetry does indeed confound the test of parametric specification with the test of the symmetry restriction.

We show that the nonlinear Slutsky symmetry restriction may be imposed via a Lagrangian, and develop an easy-to-implement two-step locally linear procedure (where the restriction is linear) that is asymptotically equivalent to the Lagrangian approach. As with the imposition of homogeneity, the imposition of symmetry does reduce the number of local parameters to be estimated (see Kim and Tripathi 2003). However, in contrast to the imposition of homogeneity, the imposition of symmetry does not reduce the dimensionality of the problem, increase the speed of convergence or change the required smoothness assumptions for estimation.

In the empirical part, we estimate expenditure-share equations over prices and expenditure, and compensated (Hicksian) price semi-elasticities of these share equations under the restriction of Slutsky symmetry. Estimated nonparametric expenditure-share equations are statistically significantly different from, but qualitatively similar to, QAI parametric estimates. However, estimated nonparametric compensated price semi-elasticities are statistically and qualitatively different from QAI parametric estimates. For example, the nonparametric estimate of the compensated rent share own-price semi-elasticity is strongly increasing with expenditure, but the QAI estimate lies outside the confidence band of the nonparametric estimate and is essentially independent of expenditure.

Although techniques for testing and estimating under the Slutsky symmetry conditions are well-developed in parametric consumer demand models, they are less well-developed in nonparametric models. Lewbel (1995) develops a nonparametric test for symmetry in a demand system. The test proposed in our paper invokes much weaker smoothness assumptions. Blundell, Duncan and Pendakur (1998) and Pendakur (1999) estimate nonparametric Engel-curve systems under symmetry, but nonparametric estimation of the full consumer demand system under symmetry has not been considered.

A closely related literature considers testing and imposing other rationality restrictions that are relevant for the nonparametric analysis of demand. Tripathi (1998) considers testing homogeneity, while Kim and Tripathi (2003) focus on imposing homogeneity. Both of these papers apply to more general settings whereas in an earlier version of this paper (Haag, Hoderlein and Pendakur (2004)) we discuss testing and imposing homogeneity in the same setting as in this paper. Pointwise testing of the symmetry hypothesis is discussed in Hoderlein (2007), which also covers the case of endogenous regressors. A bootstrap procedure for testing Slutsky negative semidefiniteness is proposed in Hoderlein (2002). Finally, Epstein and Yatchew (1985) use nonparametric least

squares and Gallant and Souza (1991) use series estimators to consider testing of rationality restrictions in a consumer demand setting.

In order to keep the exposition simple and concise, in this paper we focus on testing and imposing symmetry in isolation. Of course, the same analysis can be performed in connection with homogeneity, but at the cost of a more involved analysis. See again Haag, Hoderlein and Pendakur (2004) for some guidance on this issue.

The paper proceeds as follows. In the second section, we start out by introducing the economic concepts of this paper on individual level. Moreover, since we work with data from a heterogeneous population, we explain how unobserved heterogeneity may be handled in this paper, and we also discuss the important issue of endogeneity in income or prices. In the third section, we propose a test for symmetry and analyze its asymptotic and finite sample behavior. In the fourth section, we discuss the imposition of symmetry on a locally linear model. Finally, we conclude by implementing this model using Canadian price and expenditure data.

## 2 The Broader Picture: Individual Rationality, Preference Heterogeneity and the Market Perspective

**Individual Rationality:** The economic background on individual level may be summarized as follows. For every individual, define the cost function  $C(p, u)$  to give the minimum cost to attain utility level  $u$  facing the  $M$ -vector of log-prices  $p$ . Similarly, define the Marshallian demand function for individual  $a$  as  $w = \varphi(p, x)$ , where  $w$  denote budget shares,  $x$  denotes log total expenditure and  $z = (p', x)'$ . We pick the budget share formulation as it is common in applications. For any demand function  $\varphi$ , the Slutsky matrix  $\Sigma(z) = (\Sigma^{jk}(z))_{1 \leq j \leq M-1, 1 \leq k \leq M-1}$  is defined as the Hessian of the cost function with respect to (unlogged) prices. The elements may be expressed in terms of log-price and log-expenditure (rather than utility) as

$$\Sigma^{jk}(z) = \partial_k \varphi^j(z) + \varphi^k(z) \partial_x \varphi^j(z) + \varphi^j(z) \varphi^k(z) + \delta^j(k) \varphi^k(z), \quad (2.1)$$

where  $\delta^j(k)$  denotes the Kronecker function to indicate a diagonal element and  $\partial_x = \frac{\partial}{\partial x}$  and  $\partial_k = \frac{\partial}{\partial p^k}$  are used for abbreviation (Mas-Colell, Whinston and Green 1995). If the Slutsky matrix is continuously symmetric over a region of the  $z$  space, then Young's Theorem guarantees the existence of a cost function whose derivatives could produce the observed demand system over this region (see, e.g., Mas-Colell, Whinston and Green 1995). Moreover, if  $\Sigma$  comes from constraint utility maximization subject to a linear budget constraint, then it is also negative semidefinite.

Including Slutsky negative semidefiniteness would be lengthy and is simply beyond the scope

of this paper. On this issue we refer again to Hoderlein (2007) where a pointwise test for negative semidefiniteness is proposed and implemented.

**Preference Heterogeneity:** In a heterogeneous population, all the quantities introduced above would vary across individuals, and would depend on observable household characteristics  $J$  (the “observed part of preference heterogeneity”), and unobservables  $A$  (the “unobserved part”). Consequently, the Slutsky matrix  $\Sigma(z, j, a) = (\Sigma^{jk}(z, j, a))_{1 \leq j \leq M-1, 1 \leq k \leq M-1}$  and the demand function  $\varphi(p, x, j, a)$  would also depend on  $j$  and  $a$ . For clarity of exposition, in the theoretical part of this paper we do not consider observable heterogeneity variables  $J$ . However, note that in our application we do control for a great deal of this observable heterogeneity by concentrating on a relatively homogeneous subpopulation defined by a longer vector of characteristics. In addition, we show that after concentrating on a homogeneous subsample, the remaining observable household information does not have a major effect. Since it would be easy to include  $J$  in our analysis in a completely general fashion by simply conditioning all analysis on it, we omit it only to make the notation less cumbersome.

In contrast, unobserved heterogeneity remains an issue. Omitting, for a fixed subpopulation defined by  $J = j$ , the explicit dependence on  $j$ , our theoretical Slutsky matrix is defined by  $\Sigma(Z, A)$ . In our data, we are given observations on the  $2M + 1$ -dimensional random vector  $Y = (W, Z)'$  where  $W \in \mathbb{R}^M$  is an  $M$ -vector of expenditure shares and  $Z = (P^1, \dots, P^M, X)'$  is the vector of log-prices  $P = (P^1, \dots, P^M)'$  and household log-expenditure  $X$ . The question is now what kind of information we may obtain from  $F_Y$ , the joint distribution of all observable variables (and in particular from the regression function  $m(z) = (m^1(z), \dots, m^M(z))' = \mathbb{E}(W \mid Z = z)$ ), to draw inference on  $\Sigma$ ? In the theorem below, we give an answer to this question, as well as to the following question.

**Endogenous Regressors:** In empirical work on consumer demand, total expenditure is often assumed to be endogenous, due to problems of infrequent purchase or due to correlation with unobserved preference heterogeneity  $A$ . Some research in this area also treats prices as endogenous, due to correlation between unobserved heterogeneity in constraints (ie, prices) and unobserved heterogeneity in preferences (see, eg, Deaton and Tarozzi 2000). Furthermore, industrial organization models which treat heterogeneous prices and preferences in a general equilibrium setting suggest that prices paid by individuals for goods may be endogenous as firms set the price according to unobserved characteristics of the individual. The standard approach to such problems in empirical work on consumer demand is to use instrumental variables to correct for the endogeneity of regressors.

To handle this complication, we require a vector of instruments  $Q$  (containing exogenous regressors and additional instruments) such that  $Z = \vartheta(Q) + V$ , with  $\mathbb{E}[V|Q] = 0$ . The residuals

in this regression,  $V$  may now act as control functions, i.e. they are additional regressors we may include in our regression to control for endogeneity. The following result is an implication from Hoderlein (2007) and answers the question how the property of Symmetry of  $\Sigma(z, a)$  for each  $(z, a) \in \mathcal{Z} \times \mathcal{A}$  in our subpopulation translates to similar properties of the observable Slutsky matrix  $S(z, v) = (S^{jk}(z, v))_{1 \leq j \leq M-1, 1 \leq k \leq M-1}$ , where  $S^{jk}(z, v) = \partial_k m^j(z, v) + m^k(z, v) \partial_x m^j(z, v) + m^j(z, v) m^k(z, v) + \delta^j(k) m^k(z, v)$ , if we allow for regressors to be endogenous.

**Theorem 1.** *Let all quantities be as defined above. Suppose that  $Z \perp A|V$  and some technical regularity conditions hold. Then, if  $\text{Cov}[\partial_x \varphi, \varphi'|Z, V]$  is negative semidefinite and symmetric, then  $\Sigma \text{ nsd} \Rightarrow S \text{ nsd}$ , and  $\Sigma \text{ symmetric} \Rightarrow S \text{ symmetric}$ , except possibly on a set of probability zero.*

This theorem gives sufficient conditions for the empirical Slutsky matrix to “inherit” negative semidefiniteness and symmetry, allowing for potentially endogenous regressors (for a proof see Hoderlein (2007)). Other than regularity conditions, we require that a certain (unobservable) covariance matrix has to be negative semidefinite and symmetric, and  $Z \perp A|V$  (together with a restriction that defines  $V$ , e.g., mean independence and additive separability of  $V$  from  $S$ ). The leading special case is of course if the population is identical up to an additive shift term, i.e.,  $\varphi(X, A) = \phi(X) + \eta(A)$ . While the independence assumptions hold provided that instruments are jointly independent of unobservables  $A$  and  $V$  as is commonly the case in applied work, the covariance condition is less well motivated. It is discussed in more detail in Hoderlein (2007), see also Lewbel (2001) for related work. For the purposes of this paper we remark that the same covariance condition would arise, even if we were to restrict ourselves to a parametric form with heterogeneity in parameters only. Since correcting for endogeneity in this setup requires only the use of control function residuals as additional regressors, the same remark as the one about inclusion of additional regressors applies: We neglect  $V$  in the following, but solely for notational convenience.

## 3 Testing Symmetry

### 3.1 The Test Statistic and the Asymptotic Distribution

To test for symmetry, we propose a test using the  $L_2$  distance of those elements of the Slutsky matrix which are the same under symmetry. That is, we integrate and add up the squared distance between  $S(z)$  and  $S(z)'$  (the other option would be to use a pointwise test). The null hypothesis

is

$$H_0: \mathbb{P}(s^{jk}(Z) = s^{kj}(Z), \forall j \neq k) = 1$$

and the alternative is that there is at least one pair  $(j, k)$  with  $s^{jk}(z) \neq s^{kj}(z)$  over a significant range. We may express the alternative as

$$H_1: \mathbb{P}(s^{jk}(Z) = s^{kj}(Z), \forall j \neq k) < 1.$$

The null hypothesis is equivalent to the condition that the  $L_2$ -distance of these functions is zero. Using a nonnegative and bounded weighting function  $a(z)$  this can be written as

$$\Gamma_S = \mathbb{E}\left(\sum_{j < k} (s^{jk}(Z) - s^{kj}(Z))^2 a(Z)\right) = 0.$$

Of course this is only equivalent to the null hypothesis if  $a(z)$  is nonzero over the whole support of  $Z$ . However the weighting function allows to restrict the test to certain regions of the explanatory variables. Secondly, assuming that the support of  $a(z)$  is strictly contained in the support of the density, boundary problems can be avoided.

A test statistic may be constructed by the analogy principle. Observing a sample of  $n$  independent and identically distributed random vectors  $(W_1, Z_1), \dots, (W_n, Z_n)$  we replace the unknown functions  $m^j(z)$  by their Nadaraya-Watson estimators  $\hat{m}_h^j(z) = \sum_i K_h(z - Z_i) W_i^j / \sum_i K_h(z - Z_i)$ , where  $K(\cdot)$  is a  $M + 1$ -variate kernel function and  $K_h(u) = (\det H)^{-1/2} K(H^{-1/2}u)$  with a bandwidth matrix  $H^{1/2}$ . For simplicity of notation we assume that the bandwidth matrix is diagonal with identical bandwidth  $h$  in each direction and that the kernel is a product kernel with properties defined in detail in appendix 1 below. The derivatives of the estimator  $\partial_k \hat{m}_h^j(z)$  are used as estimators for the derivatives  $\partial_k m_j(z)$ . We then obtain

$$\hat{\Gamma}_S = \frac{1}{n} \sum_{j=1}^{M-2} \sum_{k=j+1}^{M-1} \sum_{i=1}^n (\partial_k \hat{m}_h^j(Z_i) + \hat{m}_h^k(Z_i) \partial_x \hat{m}_h^j(Z_i) - \partial_j \hat{m}_h^k(Z_i) - \hat{m}_h^j(Z_i) \partial_x \hat{m}_h^k(Z_i))^2 A_i \quad (3.1)$$

where  $A_i = a(Z_i)$ . This test statistic is a nonlinear combination of the function and its derivatives. However, since the estimator of the derivative converges more slowly than the estimator of the function, the asymptotic distribution is dominated by the derivative estimator and the function can be treated as if it were known.

To define the expected value and variance of the test statistic we define the covariance matrix  $(\sigma^{ij}(z))_{1 \leq i, j \leq M-1} = \mathbb{E}((W - m(Z))(W - m(Z))' \mid Z = z)$  and introduce the following notation:

$$\sigma_S^{jkj'k'} = \tilde{\sigma}^{jj'jj'} + \tilde{\sigma}^{kk'kk'} + \tilde{\sigma}^{kj'kj'} + \tilde{\sigma}^{jk'jk'} - 2\tilde{\sigma}^{jj'jk'} - 2\tilde{\sigma}^{jj'kj'} + 2\tilde{\sigma}^{jj'kk'} + 2\tilde{\sigma}^{jk'kj'} - 2\tilde{\sigma}^{jk'kk'} - 2\tilde{\sigma}^{kj'kk'}$$

with

$$\begin{aligned}\tilde{\sigma}^{l_1 l_2 l_3 l_4} &= \int \sigma^{l_1 l_2}(z) \sigma^{l_3 l_4} a(z)^2 K^{l_1 l_2 l_3 l_4}(z) dz \\ K^{j k j' k'}(z) &= \iint K^k(z, v) K^j(z, v - u) dv \cdot \int K^{k'}(z, w) K^{j'}(z, w - u) dw du\end{aligned}$$

and

$$K^j(z, v) = \frac{\partial K}{\partial v^j}(v) + m^j(z) \frac{\partial K}{\partial x}(v).$$

Similarly define

$$\begin{aligned}b_S^{jk} &= \int \sigma^{jj}(z) a(z) \int (K^k(z, u))^2 dudz + \int \sigma^{kk}(z) a(z) \int (K^j(z, u))^2 dudz \\ &\quad - 2 \int \sigma^{jk}(z) a(z) \int K^j(z, u) K^k(z, u) dudz.\end{aligned}$$

The asymptotic distribution is given in the following

**Theorem 2.** *Let the model be as defined above and let assumptions 1– 5 hold. Under  $H_0$ ,*

$$\sigma_S^{-1}(nh^{(M+5)/2}\hat{\Gamma}_S - h^{-(M+1)/2}B_S) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad (3.2)$$

where

$$\sigma_S^2 = 2\left(\sum_{j < k} \sigma_S^{j k j k} + 2 \sum_{j < k} \sum_{\substack{j' < k' \\ (j, k) < (j', k')^2}} \sigma_S^{j k j' k'}\right) \quad B_S = \sum_{j < k} b_S^{jk}$$

**Remarks.** 1. Simplifying the proofs in the appendix to one line, the test statistic can be written as

$$\hat{\Gamma}_H = \Gamma_S + U_n + \Delta_n$$

where  $\Gamma_S$  is 0 under  $H_0$ ,  $\Delta_n$  depends upon the uniform rate of convergence of the estimators and  $U_n$  is a degenerate U-statistic. This U-statistic converges at the rate  $nh^{(M+1)/2}$ , which might be faster than  $n^{1/2}$  depending on the choice of the bandwidth sequence. Under the alternative  $\Gamma_S$  is a positive constant and after multiplying the test statistic with  $nh^{(M+1)/2}$  this term tends to infinity. Therefore we obtain consistency of the test against alternatives with  $\Gamma_S > 0$ .

2. Another test for Slutsky symmetry based on kernel regression has been proposed by Lewbel (1995). This test procedure is based on the integrated conditional moment (ICM) test of Bierens (1982), which uses the fact that the alternative hypothesis is equivalent to

$$\mathbb{E}(s^{jk}(Z) - s^{kj}(Z))w(\Xi'Z) \mid \Xi = \xi) \neq 0$$

---

<sup>2</sup>The ordering is in a lexicographic sense.



for a set of  $\xi$  with nonzero Lebesgue-measure, where the weighting function has to be chosen appropriately (see Bierens and Ploberger, 1997). Lewbel (1995) uses a Kolmogorov-Smirnov-type test-statistic and derives the asymptotic distribution under stringent smoothness assumptions. Assuming that the unknown function and the density are  $r+1$ -times continuously differentiable (see assumption in the appendix), Lewbel (1995) requires  $r > 2(M+1)$  whereas our test requires only  $r > \frac{3}{4}(M+1)$ . Although the smoothness class of an unknown function is difficult to establish in practice, this is a substantial relaxation of assumptions.

Fan and Li (2000) discuss in detail the question under which circumstances the ICM-test or the kernel based test of our type has greater power. This has to be done by investigating the power of the tests against local alternatives, which can be written as

$$s^{jk}(z) = s^{kj}(z) + \lambda_n \delta_n(z)$$

Low frequency alternatives are characterized by  $\lambda_n \rightarrow 0$  and  $\delta_n(z) = \delta(z)$  independent of  $n$ . In contrast, high frequency alternatives also allow  $\int \delta_n^2(z) dz \rightarrow 0$  as  $n$  tends to infinity. The main difference is that ICM tests have higher power for low frequency alternatives and kernel based tests have higher power for high frequency alternatives. These results should carry over to our situation and show that the two tests can be regarded as complements rather than opponents.

More importantly, the test of Lewbel (1995) is not exactly a full test of symmetry, because it only tests every single restriction at a time (e.g.,  $s_{12} = s_{21}$ , then  $s_{23} = s_{32}$  etc), and it is not clear how a test of all restrictions holding jointly. In addition, it requires selecting a large number of parameters and choosing a number of objects (eg, bandwidths, the test functions and squeezing functions), and it is not mentioned how to determine those. Our test just requires the selection of bandwidths whose selection we discuss below. One of the required object is an infinite order kernel which is known to work poorly in practise. In contrast our test require only standard kernels. Finally, Lewbel (1995) derives critical values for his test from the asymptotic distribution, which has to be simulated, because the distribution depends on nuisance parameters. In contrast, we suggest in the next section to implement a bootstrap procedure, which allows to better approximate the unknown distribution in a finite sample.

3. In nonparametric regression analysis the advantages of local polynomial estimators over Nadaraya-Watson estimators are well known, especially in derivative estimation (see Fan and Gijbels, 1996). The asymptotic properties of the test statistic using Nadaraya-Watson

estimators are derived, because this makes the proofs more tractable. We sketch major differences in the proof in the appendix and thereby justify the extension to the local polynomial case in the application. If we use higher order local polynomial estimators for  $m$  and its first partial derivatives, our results continue to hold when  $K(\cdot)$  is replaced by its equivalent kernel. The rate of convergence of the test statistic remains the same; only kernel constants arising in bias and variance expressions will change. In our application the advantages of the local polynomial based test is clear. For  $M = 4$ , we need  $r > 3.75$ . Using Nadaraya-Watson estimators, a kernel of order 4 has to be implemented. Using a local quadratic estimator, for example, the order only has to be 2. However, the smoothness assumptions remain unchanged.

### 3.2 Implementation and The Bootstrap

Implementation of our test statistic requires the choice of a kernel and a bandwidth. Although, we discuss these issues in detail in the Appendix, we briefly state those results here. We require the order of the kernel,  $r$ , to satisfy

$$r > \frac{3}{4}(M + 1).$$

Consequently, for  $M = 4$ , a fourth-order kernel, for example a local quadratic estimator with a Gaussian kernel, is required.

In addition, as  $n \rightarrow \infty$ , we require that the bandwidth sequence  $h = O(n^{-1/\delta})$  satisfies

$$2(M + 1) < \delta < (M + 1)/2 + 2r$$

Consequently, the optimal rate for estimation, given by  $\delta_{opt} = (M + 1) + 2r$  is excluded as a bandwidth for testing. Here, a smaller bandwidth is needed to obtain the asymptotic distribution. To implement, we calculate a data-driven bandwidth by cross-validation, and adjust it by  $n^{1/\delta_{opt}-1/\delta}$ , letting  $\delta$  equal the midpoint of its bounds.

Given the choice of kernel and bandwidth, the direct way to implement the test is to estimate the expected value  $B_S$  and the variance  $\sigma_S^2$ . This requires the estimation of integrals like

$$\int \sigma^{jj}(z)a(z) \int (K^k(z, u))^2 dudz \quad (3.3)$$

or even more complex combinations in the variance parts. Therefore estimators of the conditional variances and covariances are needed. A Nadaraya-Watson-type estimator is given by

$$\hat{\sigma}_h^{jj'}(z) = \frac{\sum_{i=1}^n K_h(z - Z_i)(W_i^j - \hat{m}_h^j(Z_i))(W_i^{j'} - \hat{m}_h^{j'}(Z_i))}{\sum_{i=1}^n K_h(z - Z_i)}$$

Given the large number of bias and variance components in theorem 2, the asymptotic approach to the test is difficult to implement. Moreover, these asymptotic approximations can work very poorly in a finite sample, as was pointed out by Hjellvik and Tjøstheim (1995).

To avoid the problems noted above, one might instead use a bootstrap procedure to derive critical values. To bootstrap the test statistic, note that the estimator of the derivative can be written as a weighted average

$$\partial_k \hat{m}_h^j(z) = \sum_{i=1}^n \tilde{V}_{ni}^k(z) W_i^j \quad (3.4)$$

where  $\tilde{V}_{ni}^k$ ,  $i = 1, \dots, N$  is a set of weights giving the  $k$ 'th price derivative of the  $j$ 'th expenditure share at  $z$  when applied to the data  $W_i^j$ . Using this in the definition of  $\hat{\Gamma}_S$  we obtain

$$\hat{\Gamma}_S = \frac{1}{n} \sum_{j=1}^{M-2} \sum_{k=j+1}^{M-1} \sum_{i=1}^n \left( \sum_{l=1}^n V_{nl}^k(Z_i) W_l^j - V_{nl}^j(Z_i) W_l^k \right)^2 A_i$$

with  $V_{nl}^k(Z_i) = \tilde{V}_{nl}^k(Z_i) + \hat{m}_h^k(Z_i) \tilde{V}_{nl}^x(Z_i)$ .

Next we exploit the fact that the estimator of the function converges faster than the estimator of the derivative. Plugging in  $W_l^j = m^j(Z_l) + \varepsilon_l^j$  and noting that for large  $n$  it holds under the assumption of symmetry that

$$\begin{aligned} \sum_{l=1}^n V_{nl}^{jk}(Z_i) m^j(Z_l) - V_{nl}^{kj}(Z_i) m^k(Z_l) &\approx \\ \partial_k m^j(Z_i) + m^k(Z_i) \partial_x m^j(Z_i) - \partial_j m^k(Z_i) - m^j(Z_i) \partial_x m^k(Z_i) &= 0 \end{aligned} \quad (3.5)$$

almost surely. Therefore the test statistic can under  $H_0$  be approximated by

$$\hat{\Gamma}_S \approx \frac{1}{n} \sum_{j=1}^{M-2} \sum_{k=j+1}^{M-1} \sum_{i=1}^n \left( \sum_{l=1}^n V_{il}^{jk} \varepsilon_l^j + V_{il}^{kj} \varepsilon_l^k \right)^2 A_i.$$

The bootstrap is based on this equation and is described as follows

1. Construct (multivariate) residuals  $\hat{\varepsilon}_i = W_i - \hat{m}_h(Z_i)$ .
2. For each  $i$  randomly draw  $\varepsilon_i^* = (\varepsilon_i^{1*}, \dots, \varepsilon_i^{M-1,*})'$  from a distribution  $\hat{F}_i$  that replicates the first three moments of  $\hat{\varepsilon}_i$ , i.e.  $\mathbb{E}_{\hat{F}_i} \varepsilon_i^* = 0$ ,  $\mathbb{E}_{\hat{F}_i} \varepsilon_i^* (\varepsilon_i^*)' = \hat{\varepsilon}_i (\hat{\varepsilon}_i)'$  and  $\mathbb{E}_{\hat{F}_i} \varepsilon_i^{j,*} = \hat{\varepsilon}_i^j$  for  $j = 1, \dots, M-1$ .
3. Calculate  $\hat{\Gamma}_S^*$  from the bootstrap sample  $(\varepsilon_i^*, Z_i), i = 1, \dots, n$  by

$$\hat{\Gamma}_S^* = \frac{1}{n} \sum_{j=1}^{M-2} \sum_{k=j+1}^{M-1} \sum_{i=1}^n \left( \sum_{\substack{l=1 \\ l \neq i}}^n V_{il}^{jk} \varepsilon_l^{j*} - V_{il}^{kj} \varepsilon_l^{k*} \right)^2 A_i.$$

4. Repeat this often enough to obtain critical values.

To approximate the distribution by the bootstrap, usually the restriction of the null hypothesis is imposed in the construction of the residuals. Because symmetry implies a complicated restriction to the demand function and its derivatives, this is not directly possible. Therefore the restriction is imposed in the construction of the test statistic by using equation 3.5.

The theoretical result concerning the bootstrap procedure is given in the following

**Theorem 3.** *Let the model be as defined above and let assumptions 1–5 and 6 hold. Under  $H_0$ , conditional on the data  $(W_i, Z_i), i = 1, \dots, n$ ,*

$$\sigma_S^{-1}(nh^{(M+5)/2}\hat{\Gamma}_S^* - h^{-(M+5)/2}B_S) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

**Remark.** 1. It is well known in the nonparametric testing literature that the convergence of the finite sample distribution of the test statistic to the asymptotic distribution is rather slow. That the bootstrap overcomes this problem is supported by various theoretical results, as well as simulation studies, e. g. Härdle and Mammen (1993), Hjellvik and Tjøstheim (1995) or Li and Wang (1998).

2. The set of admissible distributions  $\hat{F}_i$  is very general. One may use a wild bootstrap (Härdle and Mammen, 1993) or smooth conditional moment bootstrap (Gozalo, 1997) suitably modified to account for cross-equation correlations.
3. To prove the asymptotic result of theorem 3 it is sufficient to assume that the bootstrap distribution  $\hat{F}_i$  replicates the first two moments of  $\hat{e}_i$ . Matching the first three moments as suggested in the algorithm we propose potentially yields to higher order approximation in the Edgeworth expansion.

### 3.3 Possible Generalizations

The technique we have used to test for symmetry can be extended to a more general setting. Let  $m(z) = \mathbb{E}(W \mid Z)$  be an unknown function and  $\mathcal{F}$  be an operator that can be applied to  $m(z)$ . Assume that we want to test the hypothesis

$$H_0: \mathbb{P}(\mathcal{F}m = 0) = 1.$$

For the problem of testing symmetry,  $\mathcal{F}$  maps from the space of  $M + 1$ -variate demand functions to the space of  $(M + 1)M/2$ -dimensional vector-valued functions with the specification  $\mathcal{F}m = (s^{jk}(Z) - s^{kj}(Z))_{j=1, \dots, M+1, k > j}$ . A simpler example is when  $\mathcal{F}$  is the derivative operator applied

to a scalar valued function. This example has been considered by Hjellvik, Yao and Tjøstheim (1998). The null hypothesis is equivalent to the fact that the  $L_2(\mathbb{P})$ -norm of  $\mathcal{F}m$  is equal to zero, i. e.

$$\mathbb{E}(\mathcal{F}m)'(\mathcal{F}m) = 0.$$

Since the operator is assumed to be known, as test statistic serves an empirical version of the left-hand side of this equation. If we replace the expectation by the sample mean and use a kernel estimator  $\hat{m}_h$  for the unknown function, this leads to

$$\hat{T} = \frac{1}{n} \sum_{i=1}^n (\mathcal{F}\hat{m}_h(Z_i))'(\mathcal{F}\hat{m}_h(Z_i))$$

For a specific operator, the asymptotic properties of this test statistic can be derived by using similar steps as we do in the proof. For the same reasons as mentioned above one may wish to avoid to use the asymptotic distribution of the test statistic and rely on a bootstrap version. As mentioned above, the kernel estimator can be written as a weighted sum  $\hat{m}_h(z) = \sum_{i=1}^n V_{hi}(z)W_i$ . Assume that the operator allows to write

$$\mathcal{F}\hat{m}_h(z) = \tilde{\mathcal{F}} \sum_{i=1}^n \tilde{V}_{hi}(z)W_i = \tilde{\mathcal{F}} \sum_{i=1}^n \tilde{V}_{hi}(z)m(Z_i) + \tilde{\mathcal{F}} \sum_{i=1}^n \tilde{V}_{hi}(z)\epsilon_i$$

and the first term on the right hand side converges to zero under  $H_0$ . This is the case if  $\tilde{\mathcal{F}} = \mathcal{F}$  and  $\tilde{V}_{hi} = V_{hi}$ , i. e. the operator is linear. In our more complex situation it holds that

$$\tilde{\mathcal{F}} \sum_{i=1}^n \tilde{V}_{hi}(z)m(Z_i) = \mathcal{F}m(z) + \sum_{i=1}^n (\mathcal{F} - \tilde{\mathcal{F}}\tilde{V}_{ni})m(Z_i), \quad (3.6)$$

where the second term on the right hand side converges to zero for all  $z$ , because the estimator of the function converges faster than the estimator of the derivative. So the non-linear part of the operator can be transferred to the kernel weights  $\tilde{V}_{hi}$ . For the bootstrap to work, the null hypothesis has to be imposed to the test statistic. This is done by observing that the first term on the right hand side of (3.6) is approximately zero for large  $n$ . Then the test statistic is under  $H_0$  approximately equal to

$$\hat{T} \approx \frac{1}{n} \sum_{i=1}^n (\tilde{\mathcal{F}} \sum_{i=1}^n \tilde{V}_{hi}(Z_i)\epsilon_i)'(\tilde{\mathcal{F}} \sum_{i=1}^n \tilde{V}_{hi}(Z_i)\epsilon_i).$$

Then, the distribution of the test statistic is approximated by bootstrapping from the residuals using a wild bootstrap procedure.

### 3.4 Monte Carlo Analysis

Up to this point, our analysis was motivated by asymptotic arguments. In this subsection we analyze the behavior of our test statistic in a controlled experiment which is designed to replicate features of our real world application below. We focus on the following issues that are important for applications: 1) we compare our test with the only other existing alternative, Lewbel's (1995) test; 2) we explore how sensitive our test is with respect to the choice of bandwidth, which is always a sensitive issue in nonparametric econometrics; 3) we explore how sensitive our test is with respect to the sample size of the data in order to get an idea of the data requirements in a typical application; and 4) finally, we perform some robustness checks by considering other model specifications.

Starting with the first issue, although both our  $L_2$  test and Lewbel's (1995) sup-norm test have power against all alternatives, our symmetry test is far easier to implement than Lewbel's in part because it does not require the use of complex squeezing functions. In fact, the implementation of his test is very cumbersome, and thus has never been performed. Furthermore, our test imposes less stringent assumptions on the smoothness of the underlying demand functions. For a direct small sample comparison, we perform a small Monte Carlo exercise assessing the power of these tests. Our data generating process is a 3-good budget-share system given by a 3-good QAI model based on our data, where the 3 goods are food (pooling both -in and -out), rent and clothing. Our model for QAI demands (based on estimates on our data for food, rent and clothing) is given by (see, eg, Banks, Blundell and Lewbel 1997)

$$\begin{aligned} W_i^j &= a^j + \sum_{k=1}^3 a^{jk} P_i^k + b^j \tilde{X}_i + q^j \bar{X}_i + \varepsilon_i^j \\ \tilde{X}_i &= X_i - \sum_{k=1}^3 a^k P_i^k - \frac{1}{2} \sum_{k=1}^3 \sum_{l=1}^3 a^{kl} P_i^k P_i^l \\ \bar{X}_i &= \exp \left( - \sum_{k=1}^3 b^k P_i^k \right) \left( \tilde{X}_i \right)^2, \end{aligned}$$

where  $j = 1, 2, 3$ ,  $\sum_k \varepsilon_i^k = 0$ ,  $\sum_k a^k = 1$ ,  $\sum_k a^{jk} = 0 \forall j$ ,  $\sum_k b^k = 0$ ,  $a^1 = 0.33$ ,  $a^2 = 0.54$ ,  $a^{11} = 0.05$ ,  $a^{12} = -0.02$ ,  $a^{22} = 0.06$ ,  $b^1 = -0.01$ ,  $b^2 = -0.09$ ,  $q^1 = 0.05$ ,  $q^2 = -0.11$ . This 3-good system satisfies homogeneity and adding-up, so that there is only 1 Slutsky restriction:  $a^{12} = a^{21}$ . Our Monte Carlo  $P_i, X_i$  data are standard normals which replicates those in our empirical application: the standard deviation of log-expenditure is 0.55, the standard deviations of the log-prices of food, rent and clothing are all 0.45. The disturbances  $\varepsilon_i^1, \varepsilon_i^2$  are mean-zero multivariate normal with variances equal to 0.13 and covariance equal to  $-0.02$ , again chosen

to replicate the data in our application. We use a sample size of  $n = 2000$ , although we have many more observations in our empirical application below. For our test, we use a local quadratic estimator with a gaussian kernel and for Lewbel’s test, we use a local mean estimator with a gaussian kernel.

We assess power by setting  $a^{21} = a^{12} + d$  where  $d$  is a deviation from symmetry. In this context,  $d$  has a natural scale: in the QAI model  $a^{jk}$  is equal to the  $jk$ ’th compensated cross-price semi-elasticity for  $\tilde{X}_i = 0$ . So, given the sizes of  $a^{jk}$  in our exercise and those observed in empirical work on demand,  $d$  starts to ‘seem large’ at an absolute value of 0.05 or more.

Figure 1 gives the results for Monte Carlo power analysis of both our test and Lewbel’s (1995) test using cross-validated bandwidths and  $n = 2000$  in each simulation, and averaging over 100 simulations<sup>3</sup>. For simplicity, we use the same bandwidth in each equation. Note that for our test, we shrink the cross-validated bandwidth by  $n^{1/\delta_{opt}-1/\delta}$ , which equals 0.81 for this Monte Carlo problem. For Lewbel’s test, we use cross-validated bandwidths. Here we see that although both tests get the size right on the Null Hypothesis, our test has much greater power to detect small violations of symmetry. For violations greater in absolute value than 0.06, our test rejects in all simulations. In contrast, Lewbel’s test rejects only at a rate of about 60% for violations of  $\pm 0.10$ .

Figure 2 gives results for our test with bandwidths that are either too large or too small relative to our recommended bandwidth selection rule. The thick black line is reproduced from Figure 1. The grey line uses a bandwidth that is one-third smaller than our recommended choice. Note that in a 4-dimensional problem, an one-third reduction in the bandwidth in every direction is a very substantial change. We find that the use of the smaller bandwidth affects the rejection rate of the test on the Null Hypothesis, with a 10% rejection rate when the size of the test is 5%. That is, using the wrong bandwidth results in size distortion. However, it does not much affect the power of the test off the Null Hypothesis. As with the test using our recommended bandwidth selection rule, the smaller bandwidth test rejects in all the simulations with violations greater in absolute value than 0.06. Similarly, the thin grey line shows rejection rates for our test using a bandwidth one-half larger than our recommended choice. Here, we see the same pattern, with over-rejection on the Null Hypothesis, but power roughly unaffected off the Null Hypothesis.

Figure 3 gives results for Monte Carlo power analysis of our test with varying sample sizes. We use samples sizes of 2000, 1000 and 500. Here, we see that sample sizes substantially reduce the power of the test off the Null Hypothesis, and may distort the size of the test on the Null

---

<sup>3</sup>Because each simulation requires about 30 minutes of computer time, we used only 100 simulations for each experiment in Figures 1 to 3. However, we conduct Monte Carlo experiments with 500 simulations as well for  $d = 0$  and  $d = \pm 0.06$  and the results are identical to the 100 simulation case for our test (rejection rates of 5.0% and 100.0% respectively), and very similar to the 100 simulation case for Lewbel’s test.

Hypothesis. In particular, if  $n = 1000$ , the test rejects in all simulations only for violations larger than  $\pm 0.08$ . For  $n = 500$ , this occurs only for violations of  $\pm 0.10$ . Turning to the size of the test, the rejection rate is 6% on the Null Hypothesis for  $n = 1000$  and 9% for  $n = 500$ . In order to get greater precision on the estimated size of the test, we rerun Monte Carlo's on the Null Hypothesis averaging over 500 simulations. This exercise gives rejection rates of 5.2%, 6.4% and 9.4% on the Null Hypothesis for  $n = 2000$ ,  $n = 1000$  and  $n = 500$ , respectively. Thus, we conclude that a moderately large sample is needed (ie, at least  $n = 1000$  in this case) to get acceptable behavior for our test.

Turning now to how the model specification affects the performance of the test, we conducted an experiment comparing the results with QAI model to results with an Almost Ideal (AI) model and to results with homothetic demands. These models were generated by restricting the QAI model with  $q_1 = q_2 = 0$  to generated AI demands, and  $b_1 = b_2 = q_1 = q_2 = 0$  to generate homothetic demands. We do not present a Figure showing these results for the sake of brevity, and because we found that neither the size of the test nor the power of the test were affected by changing the model.

Our Monte Carlo exercises may be summarized as follows. First, although both our test and Lewbel's (1994) test get the size right, the Monte Carlo evidence suggests that our test has greater power to detect small deviations from symmetry. Second, our recommended bandwidth selection rule for testing, which shrinks the cross-validated bandwidth, performs well, with larger or smaller bandwidths resulting in size distortions, but relatively little effect on the power. Third, although our test has reasonable power with smaller sample sizes, it appears that small samples over-reject on the Null Hypothesis, and that a larger sample is necessary to get the size right. Fourth, changing the data-generating process to Almost Ideal or to homothetic preferences did not change the power or size of the test.

## 4 Imposing Symmetry

In order to impose symmetry on a nonparametric regression function, we first introduce the basic setup, including notations. Moreover, we briefly state the main result in exactly the form given by our multivariate setup, with the intention to serve as reference in the latter part of this section when we discuss the restricted estimator.



## 4.1 The System Locally Linear Estimator

Unrestricted estimators for the coefficients of levels and slopes in a nonparametric demand system may be obtained by considering the locally linear least squares criterion

$$\min_{\theta(z) \in \Theta} n^{-1} \sum_{i=1}^n K_i \epsilon'_i \epsilon_i \quad (4.1)$$

with local polynomial residual  $\epsilon_i$  given by

$$\begin{aligned} \epsilon_i^j &= W_i^j - m^j(z) - \sum_{k=1}^M \beta_k^j(z) (P_i^k - p^k) - \beta_x^j(z) (X_i - x) \\ &= W_i^j - m^j(z) - \sum_{k=1}^M (h\beta_k^j(z)) (\bar{P}_i^k - \bar{p}^k) - (h\beta_x^j(z)) (\bar{X}_i - \bar{x}), \end{aligned}$$

where overbars denote division by  $h$  so that, for example,  $\bar{P}_i^k = P_i^k/h$ ;  $\epsilon_i = (\epsilon_i^1, \dots, \epsilon_i^{M-1})'$ ;  $K_i = K_h(z - Z_i)$  and the functions  $m^j(z)$  and  $\beta_k^j(z)$  give the levels and slopes of the vector-valued regression at  $z$ . Denote the subset of parameters giving levels and slopes as  $\theta(z) = (m(z)', \theta_\beta(z)')'$ , where  $m(z) = (m^1(z), \dots, m^{M-1}(z))'$ , and  $\theta_\beta(z) = (\theta_\beta^1(z)', \dots, \theta_\beta^{M-1}(z)')'$ ,  $\theta_\beta^j(z) = \{\theta_k^j(z), \theta_x^j(z)\} = \{h\beta_k^j(z), h\beta_x^j(z)\}$ ,  $j = 1, \dots, M-1$ ,  $k = 1, \dots, M$ , so that  $\theta_\beta(z)$  denotes the level of  $h$  scaled slope coefficients. Finally, denote estimators for these quantities by a hat.

## 4.2 The Asymptotic Distribution of the Unrestricted System Local Linear Estimator (SLLE)

In the single equation case, the asymptotic behavior of the parameters maximizing the local polynomial criterion is well understood, e.g. Härdle and Tsybakov (1997). We give the following result which establishes the asymptotic behavior of the locally linear estimator in a system of equations.

**Theorem 4.** *Let the model be as defined above, and let assumptions 1–3 hold. Then,*

$$\sqrt{nh^{M+1}} \left( \hat{\theta}(z) - \theta(z) - h^2 \text{bias}(z) \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, f(z)^{-1} \begin{pmatrix} \Sigma(z) \kappa_0^{M+1} & \Sigma(z) \otimes b' \\ \Sigma(z) \otimes b & \Sigma(z) \otimes A \end{pmatrix} \right),$$

where,  $b$  is a fixed  $M+1$  vector given by  $b = \mu_2^{-1} \kappa_0^M \kappa_1 \iota_{M+1}$ , with  $\iota_{M+1}$  a vector of ones of length  $M+1$ , and  $A$  is a fixed  $(M+1) \times (M+1)$  matrix given by  $A = \mu_2^{-2} \kappa_0^M \kappa_2 \iota_{M+1} \iota_{M+1}'$ , with  $\kappa_l = \int u^l K^2(u) du$ ,  $l = 0, 1, \dots, 4$ . Moreover,  $\text{bias}(z) = (\text{bias}_{\hat{\alpha}}(z)', \text{bias}_{\hat{\theta}_\beta(z)}(z)')'$  where  $\text{bias}_{\hat{\alpha}}(z) = \kappa_2 (\partial_x^2 m(z) + D_p^2 m(z)' \iota_M)$ , and  $\text{bias}_{\hat{\theta}_\beta(z)}(z) = \text{vec}(\text{bias}_{\hat{\theta}_k^j(z)}(z))_{\substack{j=1, \dots, M-1, \\ k=1, \dots, M, x}}$ , with  $\text{bias}_{\hat{\theta}_k^j(z)}(z) = h\kappa_4 (\partial_k^3 m^j(z) + \sum_{\substack{l=1, \dots, M, x \\ l \neq k}} \partial_k \partial_l^2 m^j(z))$

Because this is a straightforward extension of existing results, we transfer the proof to the author's webpage: <http://webrum.uni-mannheim.de/vwl/stefanho>.

### 4.3 The Asymptotic Distribution of the Symmetry-Restricted SLLE

Theorem 4 is a building block in the asymptotic distribution of the restricted estimator. Specifically, we make use of the fact that the symmetry restrictions are linear in the derivative terms, and that the derivative terms converge more slowly than the level terms. As a consequence, the restriction of Slutsky symmetry behaves asymptotically like a set of linear restrictions on the slope parameters. This result parallels the discussion of the nonparametric test statistic, where the speed of convergence of the test statistics is also governed by the slower speed of convergence of estimators for the derivatives.

Let  $G(\cdot)$  denote the mapping from  $\mathbb{R}^{(M+2)(M-1)}$  into  $\mathbb{R}^{(M-1)(M-2)/2}$  which maps  $\theta(z)$  into  $(M-1)(M-2)/2$  symmetry restrictions

$$(\theta_k^j(z) + \theta_x^j(z)m^k(z)) - (\theta_j^k(z) + \theta_x^k(z)m^j(z)) = 0, \quad 1 \leq j < k \leq M-1$$

A one-step symmetry restricted estimator is then given by the solution to the following Lagrangian,

$$\min_{\theta(z) \in \Theta} n^{-1} \sum_{i=1}^n K_i \epsilon_i' \epsilon_i \quad s.t. \quad G(\theta(z)) = 0. \quad (4.2)$$

In this minimization problem, the first order conditions are nonlinear functions of the parameters admitting only an implicit solution, denoted as  $\tilde{\theta}(z) = (\tilde{m}(z)', \tilde{\theta}_\beta(z)')'$ , which is difficult to analyze. Moreover, while it is clear that a minimum exists as we are minimizing a continuous function over a compact set, uniqueness is not obvious, as the restriction is not convex.

Hence, to determine the asymptotic distribution of the nonlinearly restricted Lagrangian estimator, we establish that a “linearized” auxiliary two-step procedure yields an estimator which is asymptotically equivalent in the sense that its distribution differs from the solution to problem (4.2) only by terms of lower order than the leading bias term. Like in the case of the test statistic, the result is again driven by the slower convergence of the derivative estimators.

The auxiliary two step procedure has two computational advantages: First, it is clear that an unique minimum exists. Second, it is computationally straightforward, and hence easy to implement by applied people. Therefore, this procedure is interesting in its own right. Indeed, in our application there are virtually no differences between the two estimators, as may be seen from the graphs in the appendix. Finally, note that the procedure could be iterated, to yield a

numerical solution to the original nonlinear problem<sup>4</sup>. But since the various restricted estimators share the same limiting distribution (only differing by terms of higher order) we do not elaborate on this.

The auxiliary two step procedure looks as follows:

1. Step 1: Solve the unrestricted optimization problem 4.1 to obtain the vector of unrestricted levels,  $\hat{m}(z) = (\hat{m}_1(z), \dots, \hat{m}_{M-1}(z))'$ .
2. Step 2: Solve the linearly restricted problem

$$\min_{\theta_\beta(z) \in \Theta_\beta} n^{-1} \sum_{i=1}^n K_i \epsilon_i(\hat{m}(z))' \epsilon_i(\hat{m}(z)) \quad s.t. \quad R(\hat{m}(z))\theta_\beta(z) = 0, \quad (4.3)$$

over the local slope parameters  $\theta_\beta(z) = (h\beta_1(z), \dots, h\beta_{M-1}(z))'$ , where the unrestricted first step estimators  $\hat{m}(z)$  are plugged in for  $m(z)$ . Let the solution to problem (4.3) be denoted by  $\check{\theta}_\beta(z)$ .

Here, the  $((M-1)(M-2)/2) \times (M-1)(M+1)$  matrix  $R(\hat{m}(z))$  expresses the linearized symmetry restrictions. It has as first column

$$\left( \underbrace{(0, 1, 0, \dots, 0, \hat{m}_2(z), 0)}_{M+1}, \underbrace{(-1, 0, \dots, 0, -\hat{m}_1(z))}_{M+1}, \mathbf{0}_{(M-3)(M+1)} \right),$$

where  $\mathbf{0}_{(M-3)(M+1)}$  is a  $(M-3)(M+1)$  vector of zeros, and  $\epsilon_i(\hat{m}(z)) = (\epsilon_i^1(\hat{m}(z)), \dots, \epsilon_i^{M+1}(\hat{m}(z)))'$ , with

$$\epsilon_i^j(\hat{m}(z)) = W_i^j - \hat{\alpha}^j(z) - \sum_{k=1}^M \theta_k^j(z) (P_i^k - p^k) h^{-1} - \theta_x^j(z) (X_i - x) h^{-1}, \quad j = 1, \dots, M-1.$$

As mentioned, this two step procedure has the advantage that at each step an unique optimum exists. In step 1 this is trivial to see, in step 2 it follows from the fact that we optimizing a convex function under a convex constraint.

The following theorem, whose proof may be found in the appendix, establishes the sense in which the original nonlinear restriction behaves asymptotically like a linear restriction on the slopes: the two step estimator have the same asymptotic distribution as the nonlinearly restricted Lagrangian estimator up to terms of higher order in  $h$  than the leading bias term.

---

<sup>4</sup>In particular, in a third step an estimator for the levels could be obtained, keeping the derivatives fixed. We are grateful to a referee for this suggestion.

**Theorem 5.** *Let the model be as defined above, and let symmetry as well as assumptions 1–3 hold. Then,*

$$\begin{pmatrix} \tilde{m}(z) \\ \tilde{\theta}_\beta(z) \end{pmatrix} = \begin{pmatrix} \hat{m}(z) + o_p(h^2) \\ \check{\theta}_\beta(z) + o_p(h^3) \end{pmatrix}.$$

Now we are in the position to obtain an explicit distribution for the nonlinearly restricted estimator  $\tilde{\theta}(z)$ . To this end, let  $H(m(z)) = I - Q^{-1}R(m(z))'(R(m(z))Q^{-1}R(m(z))')^{-1}R(m(z))$ , where  $R$  is as defined above and

$$Q = \begin{pmatrix} I_{M-1} \otimes 1 & I_{M-1} \otimes \mu_2 \iota'_{M+1} \\ I_{M-1} \otimes \mu_2 \iota_{M+1} & I_{M-1} \otimes \mu_2 \iota_{M+1} \iota'_{M+1} \end{pmatrix}.$$

The following theorem summarizes our result:

**Theorem 6.** *Let the model be as defined above, and symmetry as well as assumptions 1–3 hold. Then,*

$$\begin{aligned} \sqrt{nh^{M+1}} \begin{pmatrix} \tilde{m}(z) - m(z) - h^2 \text{bias}_{\hat{\alpha}}(z) \\ \tilde{\theta}_\beta(z) - \theta_\beta(z) - h^2 H(m(z)) \text{bias}_{\hat{\theta}_\beta}(z) \end{pmatrix} \\ \xrightarrow{\mathcal{D}} \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, f(z)^{-1} \begin{pmatrix} \Sigma(z) \kappa_0^{M+1} & (\Sigma(z) \otimes b') \Xi' \\ \Xi (\Sigma(z) \otimes b) & \Xi (\Sigma(z) \otimes A) \Xi' \end{pmatrix} \right) \end{aligned}$$

where  $\Xi = (I + Q^{-1}R(m(z))'(R(m(z))Q^{-1}R(m(z))')^{-1}R(m(z)))$ , and the bias expressions,  $A, b$  as well as  $\Sigma(z)$  are defined as in theorem 5 above.

**Remark.** Note that the difference between the covariance matrix of the unrestricted and the restricted estimator is positive semidefinite, so that the estimator is closer to the true parameter in a MSE sense, at least if undersmoothing is performed.

## 5 Empirical Implementation

### 5.1 Data

The data used in this paper come from the following public use sources: (1) the Family Expenditure Surveys 1969, 1974, 1978, 1982, 1984, 1986, 1990, 1992 and 1996; (2) the Surveys of Household Spending 1997, 1998 and 1999; and (3) Browning and Thomas (1999), with updates and extensions to rental prices from Pendakur (2002). Price and expenditure data are available for 12 years in 5 regions (Atlantic, Quebec, Ontario, Prairies and British Columbia) yielding 60 distinct price vectors. Prices are normalized so that the price vector facing residents of Ontario in 1986 is

$(\ln 100, \dots, \ln 100)$ . The data come with weights, which are incorporated into the nonparametric estimation and testing via multiplication of data weights with kernel weights.

Table 1 gives (unweighted) summary statistics for 6952 observations of rental-tenure unattached individuals aged 25-64 with no dependents. Figure 4 gives a kernel estimate of the density of log-expenditure, and Figure 5 gives kernel estimates of the densities of the four log-prices. Analysis is restricted to these households to minimize demographic variation in preferences. Demographic variation could be added to the model by conditioning all levels, log-price derivatives and log-expenditure derivatives on demographic covariates. Rather than pursue this strategy, we use a sample with very limited demographic variation.

The empirical analysis uses annual expenditure in four expenditure categories: Food at home, Food Out, Rent and Clothing. This yields three independent expenditure share equations which depend on 4 log-prices and log-expenditure. These four expenditure categories account for about half the current consumption of the households in the sample. Estimation of this sub-demand system is only valid under the assumption of weak separability of the included 4 goods from all the excluded goods. As is common in the estimation of consumer demand, we invoke weak separability for the estimation that follows, but do not test it.

Table 1: The Data		Min	Max	Mean	Std Dev
Expenditure Shares	Food at Home	0.02	0.84	0.23	0.11
	Food Out	0.00	0.75	0.11	0.10
	Rent	0.01	0.97	0.54	0.14
	Clothing	0.00	0.61	0.12	0.09
Total Expenditure		640	40270	8596	4427
Log Total Expenditure		6.46	10.60	8.86	0.57
log-Prices	Food at Home	3.19	4.94	4.54	0.45
	Food Out	3.15	5.14	4.62	0.50
	Rent	3.29	4.97	4.45	0.43
	Clothing	3.67	4.95	4.62	0.35

## 5.2 Bandwidth Selection

The bandwidth matrices are diagonal and apply to standardized data, so that local weights for the unrestricted model for observations  $i = 1, \dots, n$  for equation  $j = 1, \dots, M - 1$  are given by

$$K_h(Z_i - z) = K\left(\frac{X_i - x}{h^j \sigma_x}\right) \prod_{k=1}^M K\left(\frac{P_i^k - p^k}{h^j \sigma_{p^k}}\right).$$

where  $\sigma_z$  are the standard deviations of each element of  $z$ , and where  $h^j$  are equation-specific scalar bandwidth parameters for  $j = 1, \dots, M - 1$ . The  $\sigma_z$  are given in Table 1. We cross-validate to find ‘optimal’ bandwidth parameters:  $\hat{h}_1 = 1.4$ ,  $\hat{h}_2 = 3.3$  and  $\hat{h}_3 = 2.0$ . For (unrestricted and symmetry-restricted) estimation of Slutsky terms, we use cross-validated bandwidths to estimate levels, and bandwidths larger by a factor of 1.25 to estimate derivatives.

### 5.3 Estimation

All estimates and tests are based on local quadratic estimators which are equivalent to Nadaraya-Watson estimators with 4th order kernels. Each locally weighted estimate is a quadratic form in log-prices and log-expenditure, and the locally estimated constant terms give the estimated levels. Estimation of Slutsky terms takes constant terms as given and uses the estimates local derivatives either imposing symmetry or not.

Figures 6–8 show the nonparametric estimated levels of expenditure shares in thick black lines as well as parametric quadratic almost ideal (QAI) estimates (see Banks, Blundell and Lewbel, 1997) in thin black lines for comparison. Figures 9-14 show nonparametric estimated Slutsky terms in thick lines as well as parametric QAI estimates of Slutsky terms in thin black lines for comparison. Here, symmetry-restricted Slutsky terms are grey and symmetry-unrestricted Slutsky terms are black. For own-price effects (Figures 9–11), there is one restricted and one unrestricted Slutsky term displayed. For cross-price effects (Figures 12–14), there are two symmetry-unrestricted Slutsky terms for each symmetry-restricted Slutsky term.

Simulated confidence bands are shown in the Figures with crosses, and are given for estimated expenditure shares and for symmetry-restricted estimated Slutsky terms, see also the appendix on simulation below<sup>5</sup>. Pointwise 90 % confidence bands are provided at every 5th percentile from the 5th to 95th percentile of the observations.

All estimated expenditure shares and Slutsky terms are displayed for the set of observations who live in Ontario 1986. Although the results are presented for only one price situation, and thus seem like Engel curves, the model incorporates price variation explicitly through the local quadratic estimators. All 6952 observations are used in estimating each point. This is quite different from the standard approach of estimating the Engel curve in each price situation separately.

The estimated expenditure shares shown in Figures 6 to 8 seem plausible, and are similar to the global quadratic estimates. This is consistent with the literature on nonparametric Engel curves (see e.g., Blundell, Duncan and Pendakur, 1998). However, it is notable that large portions

---

<sup>5</sup>Simulated confidence bands for symmetry-unrestricted Slutsky terms are slightly larger than those for symmetry-restricted Slutsky terms, and are suppressed to reduce clutter in the Figures.

of the Food-Out and Rent estimated QAI expenditure share equations lie outside the pointwise confidence intervals of the nonparametric estimates. This is due to the fact that the QAI expenditure share equations have to fit all 60 price vectors in the data with only a few parameters to accommodate the price variation. Quadratic Engel curves estimated separately for each price regime fit the data better than the QAI demand system, but have the disadvantage of not satisfying symmetry by construction. The nonparametric demand system estimated here has the best of both worlds — it fits the data better, and may be restricted so that its Slutsky matrix satisfies symmetry.

The estimated own-price Slutsky terms shown in Figures 9 to 11 are broadly reasonable. The estimated Food-In and Rent own-price Slutsky terms are negative over the entire range of expenditure, which is implied by (but not sufficient for) concavity of the cost function. The point estimates for the Food-Out own-price Slutsky term are positive below the 15th percentile of expenditure, but nowhere are the point estimates significantly greater than zero. These patterns are evident regardless of whether or not symmetry is imposed.

Although the QAI estimates of expenditure shares are tolerably close to the nonparametric estimates, the QAI estimates of own-price Slutsky terms are quite distant from the nonparametric estimates. For some of three own-price Slutsky terms, they lie outside the 90 % confidence bands of the symmetry-restricted nonparametric estimates. Slutsky terms given QAI are third order polynomials in log-expenditure, but the higher order terms are very constrained by estimated price effects in the QAI so that the QAI Slutsky terms in the Figures are very close to linear over the middle of the expenditure distribution. This lack of curvature in the QAI fails to capture the curvature shown by nonparametric estimates. For example, the nonparametric estimate of the rent share compensated own-price semi-elasticity is strongly increasing with expenditure but the QAI estimate is nearly independent of expenditure.

Figures 12 to 14 show estimated symmetry-unrestricted cross-price Slutsky terms and symmetry-restricted cross-price Slutsky terms. Again, the estimates given QAI are quite distant from the nonparametric estimates. Only for the Food-Out-Rent cross-price Slutsky term does the QAI estimate lie within the confidence band of the nonparametric estimate. As with own-price Slutsky terms, the estimates given QAI do not allow for as much curvature as the nonparametric estimates suggest. This is especially evident in the Food-in-Food-out cross-price Slutsky term. Here, the nonparametric estimates are strongly increasing with expenditure, but the QAI estimate is essentially independent of expenditure.

In Figures 12 to 14, there are two estimated unrestricted terms and one estimated restricted term for each cross-price effect. We test symmetry by asking whether or not the two estimated unrestricted terms are very different from each other, and given the size of estimated confidence

bands, one might expect the hypothesis that symmetry holds everywhere to be rejected, and it is indeed massively rejected.

## 5.4 Testing

For all tests, we use a weighting function to focus the test on the middle of the distribution of data. The weighting function  $a(p, x)$  is equal to 1 for all observations which: (1) are in the middle 95 % of the expenditure distribution for some price vector; and (2) face a price vector wherein each price lies in the middle 95 % of its respective marginal distribution. The weight function  $a(p, x)$  is equal to zero otherwise. Since the prices exhibit covariation, deleting outliers in the marginal price distributions reduces the sample by about 10 %, leaving  $a(p, x)$  equal to 1 for about 85 % of observations in the sample. Thus, test statistics are summed over only 85 % of the 6952 observations used in the estimation.

### 5.4.1 Testing the QAI demand system

One may test the hypothesis that the nonparametric estimates are different from the QAI estimates using standard methods (see Härdle and Mammen, 1993). We compute a test statistic which sums the squared difference between the estimated QAI and nonparametric expenditure shares across all observations across all three equations.

The sample value of this test statistic for the QAI restriction is 4.27. We bootstrap the sampling distribution of this test statistic under the null that expenditure shares are given by the estimated QAI equations. The bootstrap 5 % and 1 % critical values are 1.99 and 5.46, and the bootstrap p-value for our sample test statistic is 0.013. So, we may reject the hypothesis that preferences are QAI against a nonparametric alternative at reasonable sizes. That we do not find an overwhelming rejection is consistent with other work testing QAI demands (eg, Banks, Blundell and Lewbel 1997).

### 5.4.2 Testing Symmetry

To test symmetry when the number of goods  $M$  is four, we need the order of the local polynomial equivalent kernel to exceed 3.75 ( $r > \frac{3}{4}(M+1)$ ). The local quadratic with kernel of order 2 satisfies this condition. As we note in the remark after assumption 5, we calculate a data-driven bandwidth (by cross-validation) and adjust it by  $n^{1/\delta_{opt}-1/\delta}$  where  $\delta_{opt} = (M+1) + 2r$  and  $\delta = \frac{5}{4}(M+1) + r$  (the midpoint of its bounds). Consequently, in our application, the bandwidths are adjusted to  $h = 0.83 \times \hat{h}$ .



For the symmetry test, we construct  $\hat{\Gamma}_S$  using  $a(p, x)$  as above. The value of the test statistic is 1943. To find critical values the bootstrap procedure described in section 3.2 is applied, using the multivariate smooth conditional moment bootstrap with the number of bootstrap iterations  $B = 300$ . The simulated distribution of the test statistic is shown in Figure 15. Clearly, the value of the test statistic is way out in the tail of its simulated distribution under the null of symmetry. The simulated 5 % and 1 % critical values are 313 and 886, respectively, so symmetry is decisively rejected.

We also applied Lewbel’s (1994) test to our data. Because Lewbel’s approach is quite memory-intensive, we were able to apply it only to a 75% random sample of our data. (Our test also rejects decisively on this sample at the 1% critical value.) Since Lewbel does not give guidance as to bandwidth and kernel choice, we implement his test with cross-validated bandwidths and normal kernels. We use Lewbel’s preferred method for calculating critical values, which is via simulation. Lewbel’s test applies only to tests of symmetry across pairs of goods, rather than the to entire set of  $M(M - 1)/2$  joint restrictions which are together sufficient for Slutsky symmetry. His test 6 statistics are 0.63, 3.98, 3.35, 0.80, 1.15 and 1.73 and their associated 5% critical values are 5.40, 6.00, 5.36, 6.13, 4.98 and 5.33. Consequently, at the 5% level, Lewbel’s test rejects none of the 6 restrictions which together comprise Slutsky symmetry<sup>6</sup>. As we demonstrated in our Monte Carlo simulation, this may be due to the low power of Lewbel’s sup-norm test in finite samples.

One may also test symmetry under the restriction that the demand system is QAI. The likelihood ratio test statistic for the hypothesis that the QAI demand system satisfies symmetry against an asymmetric QAI alternative is 13.2 with 3 degrees of freedom, and has two-sided 5 % and 1 % critical values of 9.3 and 12.8. Thus, the QAI demand system picks up the violation of symmetry, but the additional restrictions of the QAI result in a much less decisive rejection than the nonparametric approach.

Why is symmetry rejected by our test in this application? Of course, it might simply be wrong, or the grouping of goods may violate the separability assumption. Another remaining explanation

---

<sup>6</sup>The details on the test are as follows.

Detailed Results for Lewbel’s test

Pair of Goods	Test Stat	5% crit	1% crit	p-value
foodin, foodout	0.63	5.40	10.09	0.33
foodin, rent	3.98	6.00	8.12	0.85
foodin, cloth	3.35	5.36	8.07	0.86
foodout, rent	0.80	6.13	8.06	0.40
foodout, cloth	1.15	4.98	7.55	0.56
rent, cloth	1.73	5.33	8.40	0.71

is unobserved heterogeneity<sup>7</sup>, whose impact is noted in Brown and Walker (1989), Lewbel (2001) and Hoderlein (2005). If the maintained assumptions about unobservables are false then the symmetry test may reject even if symmetry is true for all members of the underlying heterogeneous population, which may happen here. An obvious alternative is measurement error, e.g., there might not be one price for all consumers, but prices might vary. The latter are fundamental problems that limit the scope of econometric analysis.

## 6 Conclusions

The Slutsky symmetry restriction is comprised of a set of nonlinear cross-equation restrictions on the levels and derivatives of the expenditure share vector function. In parametric demand system estimation, the complexity of these restrictions and the complexity of share equations seem to go together — only in very simple demand systems is Slutsky symmetry a linear restriction. Surprisingly, we find that in the nonparametric case, where share equations can be arbitrarily complex, Slutsky symmetry is easy to impose. Because the estimated derivatives converge slower than estimated levels, and because the restrictions are linear in the derivatives, the Slutsky symmetry restriction and test become linear in our context. Based on this insight, we provide a new test of symmetry, its asymptotic distribution, guidance on kernels and bandwidths in implementation, and a methodology for bootstrap simulation of the sampling distribution of the test statistic. Monte Carlo analysis suggests that our test has better power than the Lewbel (1995) nonparametric test of symmetry. Further, we provide a two-step linearized nonparametric approach to estimating the matrix-valued function of Slutsky terms under the restriction of symmetry. Finally, we implement the model with Canadian data and find plausible results.

## 7 Appendix

### 7.1 General Assumptions

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

---

<sup>7</sup>As noted above, in this empirical work, we use a relatively homogeneous subsample with very limited demographic variation. However, the age of the single childless individuals in the sample is observable and does vary. But, it is only weakly correlated with the nonparametric residuals. Linear regressions of the residuals from the nonparametric regression on age, log-prices, log-expenditures and all two-way interactions do not have much explanatory power, with  $R^2$  values of 0.04, 0.04 and 0.02 for food-in, food-out and rent shares, respectively. Consequently, we believe that the assumption of no unobserved preference heterogeneity taken in our application is no more heroic than usual.

This assumption can be relaxed to dependent data. All proofs can be extended to  $\alpha$ -mixing processes in the case of estimation and  $\beta$ -mixing processes in the case of testing. Changes in the proofs are briefly discussed below. The validity of the bootstrap is not affected by dependent data, if we assume that  $\mathbb{E}(\varepsilon_t \mid Y_{t-1}, \dots, Y_1) = 0$ . Then, the bootstrap works because the residuals are mean independent and we only use the residuals for resampling (see also Kreiss, Neumann and Yao, 2002).

**Assumption 2.** *For the data generating process*

1.  $f(y)$  is  $r + 1$ -times continuously differentiable ( $r \geq 2$ ).  $f$  and its first partial derivatives are bounded and square-integrable.
2.  $m(z)$  is  $r + 1$ -times continuously differentiable.
3.  $f(z) = \int f(w, z)dw$  is bounded from below on the compact support  $A$  of  $a(z)$ , i. e.  $\inf_{z \in A} f(z) = b > 0$ .
4. The covariance matrix

$$\Sigma(z) = (\sigma^{ij}(z))_{1 \leq i, j \leq M-1} = \mathbb{E}((W - m(Z))(W - m(Z))' \mid Z = z) \quad (7.1)$$

is square-integrable (elementwise) on  $A$ .

**Assumption 3.** *For the kernel regression*

The kernel is a  $M + 1$ -dimensional function  $K: R^{M+1} \rightarrow R$ , symmetric around 0 with  $\int K(u)du = 1$ ,  $\int |K(u)|du < \infty$  and of order  $r$  (i. e.  $\int u^k K(u)du = 0$  for all  $k < r$  and  $\int u^r K(u)du < \infty$ ). Further  $\|x\|^{M+1} K(x) \rightarrow 0$  for  $\|x\| \rightarrow \infty$ .

**Remark.** For estimation in isolation, the boundedness on compact support assumption A.4 could be relaxed. However, this assumption is needed for testing, and hence will be maintained throughout the paper. Our results hold for arbitrary kernel functions, but since we use product kernels in our application, we will establish some of the theorems under this additional assumption. In this case, we use the notation  $\mu_2 = \int u^2 K(u)du$ , and  $\kappa_l = \int u^l K^2(u)du$  ( $l = 0, 1, 2$ ) for one-dimensional kernel functions.

## 7.2 Assumptions For Testing

For the testing part, we shall need additional assumptions, in particular on the smoothness conditions, order of kernels and rate of bandwidth.

**Assumption 4.** *For the kernel regression*

1. For the order  $r$  of the kernel, we require

$$r > \frac{3}{4}(M + 1). \quad (7.2)$$

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

$$2(M+1) < \delta < (M+1)/2 + 2r$$

**Assumption 5.**  $E((W^j - m^j(Z))^2(W^k - m^k(Z))^2) < \infty$  for every  $1 \leq j, k \leq M-1$ .

**Remark.** The asymptotic distribution of the test statistic is derived under the above conditions on the bandwidth sequence. It is important to note that the optimal rate for estimation, given by

$$\delta_{opt} = (M+1) + 2r$$

is excluded. Here, a smaller bandwidth is needed to obtain the asymptotic distribution. In practice we calculate a data-driven bandwidth (by cross-validation) and adjust it by  $n^{1/\delta_{opt}-1/\delta}$ . Although we do not formally address the issue of data-driven bandwidths  $\hat{h}$  we assume that our results will hold if  $\hat{h}/h \xrightarrow{P} 1$ .

**Assumption 6.** On the bootstrap distribution

The bootstrap residuals  $\varepsilon_i^*, i = 1, \dots, n$  are drawn independently from distributions  $\hat{F}_i$ , such that  $E_{\hat{F}_i} \varepsilon_i^* = 0$ ,  $E_{\hat{F}_i} \varepsilon_i^* (\varepsilon_i^*)' = \hat{\varepsilon}_i \hat{\varepsilon}_i'$  and that Cramer's conditions are fulfilled marginally, i. e. there exists a constant  $c > 0$  such that  $E|\varepsilon_i^{j*}|^p \leq c^{p-2} p! E(\varepsilon_i^{j*})^2 < \infty$  for all  $p = 3, 4, \dots$  and for all  $j = 1, \dots, M-1$ .

**Remark.** Cramer's conditions are fulfilled for discrete distributions, distributions with compact support and among others for the normal distribution, which are the most often used distributions in practice.

### 7.3 Proof of Theorems 2 and 3

The proof of Theorem 2 follows the same strategy as the proof of Theorem 1 in Aït-Sahalia, Bickel and Stoker (2001). In essence, this consists of applying a functional expansion to  $\hat{\Gamma}_S$ . This leads to a von Mises expansion where the first order term is zero under  $H_0$ . The second order term is an infinite weighted sum of chi-squared distributed random variables. Here, a Feller-type condition is fulfilled which ensures the asymptotic negligibility of all summands. This condition is stated as the central limit theorem for degenerate U-statistics by de Jong (1987) (see Lemma 7.1). By employing this result to derive Theorem 2, we proceed differently than Aït-Sahalia, Bickel and Stoker (2001) to alleviate the proof of the bootstrap result.

The extension to  $\beta$ -mixing random variables follows by using Theorem 2.1 in Fan and Li (1999) to obtain the asymptotic distribution. Apart from this, the difference consists in tedious calculations of covariances where essentially a summability condition of the  $\beta$ -coefficients is necessary.

#### Preliminary Lemmata

We denote marginal densities the list of the arguments and with a superscript indicating the element of  $w$  which is part of the argument. Kernel density estimators are defined in the same way.

For a density  $f$  we define the seminorms

$$\begin{aligned}\|f^k\|_f &= \max\left\{\sup_{z \in \mathcal{A}} |f(z)|, \sup_{z \in \mathcal{A}} \left| \int w^k f^k(w^k, z) dw^k \right| \right\} \\ \|f^k\|_d &= \max\left\{ \max_{p=1, \dots, M+1} \sup_{z \in \mathcal{A}} |\partial_p f(z)|, \max_{p=1, \dots, M+1} \sup_{z \in \mathcal{A}} \left| \partial_p \int w^k f^k(w^k, z) dw^k \right| \right\}\end{aligned}$$

for density functions.

**Lemma 7.1 (de Jong, 1987).** *Let  $Y_1, \dots, Y_n$  be a sequence of independent and identically distributed random variables. Suppose that the  $U$ -statistic  $U_n = \sum_{1 \leq i < j \leq n} h_n(Y_i, Y_j)$  with a symmetric function  $h_n$  is centered (i. e.  $Eh_n(Y_1, Y_2) = 0$ ) and degenerate (i. e.  $E(h_n(Y_1, Y_2) | Y_1) = E(h_n(Y_1, Y_2) | Y_2) = 0$ ,  $\mathbb{P}$ -a.s.). Then if*

$$\frac{\max_{1 \leq i \leq n} \sum_{j=1, j \neq i} \mathbb{E} h_n(Y_i, Y_j)^2}{\text{var } U_n} \longrightarrow 0 \quad \text{and} \quad \frac{\mathbb{E} U_n^4}{(\text{var } U_n)^2} \longrightarrow 3 \quad (7.3)$$

we have that

$$\frac{U_n}{\sqrt{\text{var } U_n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$$

**Lemma 7.2.** *Under the assumptions we have that for any  $k = 1, \dots, M-1$*

$$\begin{aligned}\|\hat{f}_h^k - f^k\|_d &= O_P(h^{r-1} + (\log n / (nh^{M+3}))^{1/2}) \\ \|\hat{f}_h^k - f^k\|_f &= O_P(h^r + (\log n / (nh^{M+1}))^{1/2})\end{aligned}$$

**Proof** Noting that

$$\sup_{z \in \mathcal{A}} \left| \int y (\hat{f}_h^k - f^k) dy \right| \leq \|\hat{f}_h^k(z)\|_\infty \|\hat{m}_h^k(z) - m^k(z)\|_\infty + \|m^k(z)\|_\infty \|\hat{f}_h^k(z) - f^k(z)\|_\infty \quad (7.4)$$

where  $\|\cdot\|_\infty$  denotes the supremum-norm. Then the lemma follows from the uniform rates of convergence for kernel estimators (see Härdle, 1990, or Masry, 1996).  $\square$

## Proof of Theorem 2

For simplification, denote  $\partial_x = \partial_{M+1}$ . We start by expanding the statistic in order to replace  $\hat{m}_h^k(z)$  with  $m(z)$

$$\begin{aligned}\hat{\Gamma}_S &= \frac{1}{n} \sum_{j=1}^{M-2} \sum_{k=1}^{M-1} \sum_{i=1}^n (\partial_k \hat{m}_h^j(Z_i) + m^k(Z_i) \partial_x \hat{m}_h^j(Z_i) - \partial_j \hat{m}_h^k(Z_i) - m^j(Z_i) \partial_x \hat{m}_h^k(Z_i))^2 A_i \\ &\quad + \frac{1}{n} \sum_{j=1}^{M-2} \sum_{k=1}^{M-1} \sum_{i=1}^n ((\hat{m}_h^k(Z_i) - m^k(Z_i)) \partial_x \hat{m}_h^j(Z_i) - (\hat{m}_h^j(Z_i) - m^j(Z_i)) \partial_x \hat{m}_h^k(Z_i))^2 A_i \\ &\quad + \frac{1}{n} \sum_{j=1}^{M-2} \sum_{k=1}^{M-1} \sum_{i=1}^n (\partial_k \hat{m}_h^j(Z_i) + m^k(Z_i) \partial_x \hat{m}_h^j(Z_i) - \partial_j \hat{m}_h^k(Z_i) - m^j(Z_i) \partial_x \hat{m}_h^k(Z_i)) \\ &\quad \quad \times ((\hat{m}_h^k(Z_i) - m^k(Z_i)) \partial_x \hat{m}_h^j(Z_i) - (\hat{m}_h^j(Z_i) - m^j(Z_i)) \partial_x \hat{m}_h^k(Z_i)) A_i \\ &= \hat{\Gamma}_{S1} + \hat{\Gamma}_{S2} + \hat{\Gamma}_{S3}\end{aligned} \quad (7.5)$$

By Chebychev's inequality,  $\hat{\Gamma}_{S2} = O_p(h^{2r} + n^{-1}h^{-(M+1)}) = o_p(n^{-1}h^{-(M+5)/2})$  and an application of Cauchy-Schwarz shows that the third term is also of  $o_p(n^{-1}h^{-(M+5)/2})$  provided that  $\hat{\Gamma}_{S1}$  has the limiting distribution of the theorem. So it is left to derive the asymptotic distribution of  $\hat{\Gamma}_{S1}$  has.

To apply the functional expansion, we consider the theoretical version

$$\Gamma_{S1} = \sum_{j < k} \Gamma_{S1}^{jk}.$$

For the beginning it suffices to investigate the case  $j = 1, k = 2$  and to note that the other terms can be treated in the same way.

Consider  $\Gamma_{S1}^{12}$  as a functional of two  $M + 2$ -dimensional density functions  $f^1(w^1, z), f^2(w^2, z)$ , two  $M + 1$ -dimensional functions  $m^1(z)$  and  $m^2(z)$  and a  $M + 1$ -dimensional density  $f_3(z)$ .

$$\begin{aligned} \Gamma_{S1}^{12}(f_1, f_2, c_1, c_2, f_3) &= \int (\partial_2 \frac{\int w^1 f_1(w^1, z) dw^1}{f_1(z)} + m^2(z) \partial_{M+1} \frac{\int w^1 f_1(w^1, z) dw^1}{f_1(z)} \\ &\quad - \partial_1 \frac{\int w^2 f_2(w^2, z) dw^2}{f_2(z)} - m^1(z) \partial_{M+1} \frac{\int w^2 f_2(w^2, z) dw^2}{f_2(z)})^2 a(z) f_3(z) dz \end{aligned}$$

Then the following functional expansion holds (analogous to Lemma A.1 in Aït-Sahalia, Bickel and Stoker, 2001)

**Lemma 7.3.** *Let  $|g_1(w^1, z)|, |g_2(w^2, z)| < b/2$  be bounded functions and  $G^b(R^{M+2}, R)$  the set of all such functions. Then under  $H_0$  and assumption 2,  $\Gamma_{S1}^{12}$  has an extension on  $G^b \times G^b$  around  $(f^1, f^2)$  given by*

$$\begin{aligned} \Gamma_{S1}^{12}(f^1 + g_1, f^2 + g_2, m^1, m^2, f_3) &= \Gamma_{S1}^{12}(f^1, f^2, m^1, m^2, f_3) + \\ &\quad \int \left( \partial_2 \int \alpha^1(w, z) g_1(w^1, z) dw + m_2(z) \partial_{M+1} \int \alpha^1(w, z) g_1(w^1, z) dw^1 \right. \\ &\quad \left. - \partial_1 \int \alpha^2(w, z) g_2(w^2, z) dw^2 - m^1(z) \partial_{M+1} \int \alpha^2(w, z) g_2(w^2, z) dw^2 \right)^2 a(z) f_3(z) dz \\ &\quad + O(\|g_1\|_d^2 \|g_1\|_f + \|g_2\|_d^2 \|g_2\|_f) \end{aligned} \quad (7.6)$$

with  $\alpha^k(w, z) = (w^k - m^k(z))/f(z)$ .

Let  $\hat{f}_e(z) = \frac{1}{n} \sum_{i=1}^n 1_{\{z=Z_i\}}(z)$  denote the empirical distribution of the sampled data and extend the test statistic in the following way

$$\begin{aligned} \hat{\Gamma}_{S1}^{12} &= \Gamma_{S1}^{12}(\hat{f}_h^1, \hat{f}_h^2, m^1, m^2, \hat{f}_e) \\ &= \Gamma_{S1}^{12}(f^1 + \hat{f}_h^1 - f^1, f^2 + \hat{f}_h^2 - f^2, m^1, m^2, f) \\ &\quad + \Gamma_{S1}^{12}(f^1 + \hat{f}_h^1 - f^1, f^2 + \hat{f}_h^2 - f^2, m^1, m^2, \hat{f}_e - f) \end{aligned}$$

Applying Lemma 7.3 to  $g_i = \hat{f}_h^i - f^i, i = 1, 2$  allows to write

$$= I_{S_n}^1 + \Delta_{S_n}^1 + O_p(\|\hat{f}_h^1 - f^1\|_d^2 \|\hat{f}_h^1 - f^1\|_f + \|\hat{f}_h^2 - f^2\|_d^2 \|\hat{f}_h^2 - f^2\|_f) \quad (7.7)$$

using that  $\Gamma_{S_1}^{12}(f^1, f^2, m^1, m^2, f) = 0$  under  $H_0$  and where

$$\begin{aligned} I_{S_n}^{12} &= \int \left( \sum_{i=1}^n r_{S_n}^{12}(W_i, Z_i; z) \right)^2 a(z) f(z) dz \\ \Delta_{S_n}^{12} &= \int \left( \sum_{i=1}^n r_{S_n}^{12}(W_i, Z_i; z) \right)^2 a(z) (\hat{f}_e(z) - f(z)) dz \end{aligned}$$

with

$$\begin{aligned} r_{S_n}^{12}(W_i, Z_i; z) &= \partial_2 \alpha^1(W_i, z) K_h(z - Z_i) + m^2(z) \partial_{M+1} \alpha^1(W_i, z) K_h(z - Z_i) \\ &\quad - \partial_1 \alpha^2(W_i, z) K_h(z - Z_i) - m^1(z) \partial_{M+1} \alpha^2(W_i, z) K_h(z - Z_i) \end{aligned} \quad (7.8)$$

Here it has also been used that under  $H_0$  it holds that  $\int \alpha^k(w, z) f^k(w, z) dw^k = 0$ .

The lower order terms in the extension 7.7 are bounded by Lemma 7.2 and the following

**Lemma 7.4.** *Under the assumptions we have that*

$$\Delta_{S_n}^{12} = o_p(n^{-1} h^{-(M+5)/2}).$$

To obtain the asymptotic distribution we combine the results for  $\hat{\Gamma}_{S_1}^{jk}$  for all  $1 \leq j < k \leq M-1$ , which leads to

$$\hat{\Gamma}_{S_1} = \sum_{j < k} \hat{\Gamma}_{S_1}^{jk} = \sum_{j < k} I_{S_n}^{jk} + o_p(n^{-1} h^{-(M+5)/2}) \quad (7.9)$$

Introducing the centered random variables

$$\tilde{r}_{S_n}^{jk}(W_i, Z_i; z) = r_{S_n}^{jk}(W_i, Z_i; z) - \mathbb{E} r_{S_n}^{jk}(W_i, Z_i; z) \quad (7.10)$$

the following decomposition applies

$$\begin{aligned}
I_{Sn} &= \sum_{j < k} I_{Sn}^{jk} \\
&= \sum_{j < k} \int \left( \frac{1}{n} \sum_{i=1}^n r_{Sn}^{jk}(W_i, Z_i; z) \right)^2 a(z) f(z) dz \\
&= \frac{2}{n^2} \sum_{j < k} \sum_{i_1 < i_2} \int \tilde{r}_{Sn}^{jk}(W_{i_1}, Z_{i_1}; z) \tilde{r}_{Sn}^{jk}(W_{i_2}, Z_{i_2}; z) a(z) f(z) dz \\
&\quad + \frac{1}{n^2} \sum_{j < k} \sum_{i=1}^n \int (r_{Sn}^{jk}(W_i, Z_i; z))^2 a(z) f(z) dz \\
&\quad + \frac{2(n-1)}{n^2} \sum_{j < k} \sum_{i=1}^n \int \tilde{r}_{Sn}^{jk}(W_i, Z_i; z) \mathbb{E} r_{Sn}^{jk}(W_i, Z_i; z) a(z) f(z) dz \\
&\quad - \frac{n(n-1)}{n^2} \sum_{j < k} \int \left( \mathbb{E} r_{Sn}^{jk}(W_i, Z_i; z) \right)^2 a(z) f(z) dz \\
&= I_{Sn1} + I_{Sn2} + I_{Sn3} + I_{Sn4}
\end{aligned} \tag{7.11}$$

$$\begin{aligned}
&\quad + \frac{2(n-1)}{n^2} \sum_{j < k} \sum_{i=1}^n \int \tilde{r}_{Sn}^{jk}(W_i, Z_i; z) \mathbb{E} r_{Sn}^{jk}(W_i, Z_i; z) a(z) f(z) dz \\
&\quad - \frac{n(n-1)}{n^2} \sum_{j < k} \int \left( \mathbb{E} r_{Sn}^{jk}(W_i, Z_i; z) \right)^2 a(z) f(z) dz \\
&= I_{Sn1} + I_{Sn2} + I_{Sn3} + I_{Sn4}
\end{aligned} \tag{7.12}$$

These terms are analyzed by the following

**Lemma 7.5.** *Under the assumptions we have that under  $H_0$*

$$\begin{aligned}
nh^{(M+5)/2} I_{Sn1} &\xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_S^2) \\
nh^{(M+5)/2} I_{Sn2} - h^{-(M+1)/2} B_S &\xrightarrow{P} 0 \\
nh^{(M+5)/2} I_{Sn3} &\xrightarrow{P} 0 \\
nh^{(M+5)/2} I_{Sn4} &\longrightarrow 0
\end{aligned}$$

Together with equation 7.9 this states the asymptotic result of the theorem.  $\square$

## Extension to Local Polynomial Estimators

Define  $\hat{\Gamma}_S^{LP}$  by replacing the Nadaraya-Watson estimators in the definition of  $\hat{\Gamma}_S$  with local polynomial estimators of order  $p$ . Recall that any local polynomial estimator (of a function and of a derivative) can be written as

$$\hat{m}_h^{LP}(z) = \frac{Z_n(z)}{N_n(z)}$$

with  $Z_n(z)$  and  $N_n(z)$  appropriately defined. These quantities are explicitly given using the Laplacian extension of the determinant of the estimator's density matrix. The main difference in the case of local polynomials is the proof of 7.3. It is no longer sufficient to derive an expansion of third order of the functional expansion of  $\hat{\Gamma}_{S1}^{12,LP}$  around  $\Gamma_{S1}^{12}$ . Now,  $\Gamma_{S1}^{12,LP}$  has to be regarded



as a function of  $\tilde{K} = 1 + \sum_{i=0}^{2p} \binom{i+M}{M} + \sum_{i=0}^{2p} \binom{i+M-1}{M-1}$  different functions  $f_1 \dots, f_{\tilde{K}}$  which are given by the probability limits of the entries in the density matrix. For the proof of Lemma ??, the functional has to be expanded up to order  $p+1$ , where  $p$  is the order of the local polynomial (if  $p$  is even, up to order  $p+2$ ). Tedious calculations of the higher order derivatives and an investigation of the resulting terms, show that the leading term is again given by the second derivative. It can be shown that all other terms are of lower order. The other Lemmata follow by using equivalent kernel arguments.

### Proof of Theorem 3

Analogously to equation 7.5 we get

$$\hat{\Gamma}_S^* = \hat{\Gamma}_{S1}^* + \hat{\Gamma}_{S2}^* + \hat{\Gamma}_{S3}^*$$

where the last two terms are of lower order (use Lemma 7.6). Again we decompose

$$\hat{\Gamma}_{S1}^* = \sum_{j < k} \hat{\Gamma}_{S1}^{jk*}$$

and investigate wlog the case  $j = 1, k = 2$ . Let now  $\hat{f}_h^{1*}$  denote the kernel density estimator of  $(\varepsilon_i^{1*}, Z_i)_{i=1, \dots, n}$ . Replacing  $W^1, W^2$  with  $\varepsilon^{1*}, \varepsilon^{2*}$  in the definition of  $\Gamma_{S1}^{12}$  and applying Lemma 7.3 with  $g^i = \hat{f}_h^{i*} - f^i, i = 1, 2$  we can decompose

$$\begin{aligned} \hat{\Gamma}_{S1}^{12*} &= \Gamma_{S1}^{12}(\hat{f}_h^{1*}, \hat{f}_h^{2*}, m^1, m^2, \hat{f}_e) \\ &= \Gamma_{S1}^{12}(f^1, f^2, m^1, m^2, f) + I_{Sn}^{12*} + \Delta_{Sn}^{12*} \\ &\quad + O_P(\|\hat{f}_h^{1*} - f^1\|_d^2 \|\hat{f}_h^{1*} - f^1\|_f + \|\hat{f}_h^{2*} - f^2\|_d^2 \|\hat{f}_h^{2*} - f^2\|_f) \end{aligned}$$

As  $f^i(\varepsilon^*, z) = f^1(\varepsilon^*)f(z)$ , we have that  $\Gamma_{S1}^{12}(f^1, f^2, m^1, m^2, f) = 0$ . Note that this property allows to construct the distribution of  $\hat{\Gamma}_S$  under  $H_0$  by the bootstrap. Here

$$\begin{aligned} I_{Sn}^{12*} &= \int \left( \sum_{i=1}^n r_{Sn}^{12*}(\varepsilon_i^*, Z_i; z) \right)^2 a(z) f(z) dz \\ \Delta_{Sn}^{12*} &= \int \left( \sum_{i=1}^n r_{Sn}^{12*}(\varepsilon_i^*, Z_i; z) \right)^2 a(z) (\hat{f}_e(z) - f(z)) dz \end{aligned}$$

and

$$\begin{aligned} r_{Sn}^{12*}(\varepsilon_i^*, Z_i; z) &= \varepsilon_i^{1*} \partial_2 \frac{K_h(z - Z_i)}{f(z)} + \varepsilon_i^{1*} m^2(z) \partial_{M+1} \frac{K_h(z - Z_i)}{f(z)} \\ &\quad - \varepsilon_i^{2*} \partial_j \frac{K_h(z - Z_i)}{f(z)} - \varepsilon_i^{2*} m^j(z) \partial_{M+1} \frac{K_h(z - Z_i)}{f(z)}. \end{aligned}$$

Next, lower order terms are bounded.

**Lemma 7.6.** *Under the assumptions we have that for any  $k = 1, \dots, M - 1$*

$$\begin{aligned}\|\hat{f}_h^{k*} - f^k\|_d &= O_P(h^{r-1} + (\log n/(nh^{M+3}))^{1/2}) \\ \|\hat{f}_h^{k*} - f^k\|_f &= O_P(h^r + (\log n/(nh^{M+1}))^{1/2})\end{aligned}$$

Analogous to Lemma 7.4 it holds that

$$\Delta_{S1}^{12*} = o_P(n^{-1}h^{-(M+5)/2}).$$

Together we have that

$$\hat{\Gamma}_{S1}^* = \sum_{j < k} I_{Sn}^{jk*} + o_P(n^{-1}h^{-(M+5)/2}).$$

Next, the same decomposition as in 7.12 applies

$$I_{Sn}^* = \sum_{j < k} I_{Sn}^{jk*} = I_{Sn1}^* + I_{Sn2}^* + I_{Sn3}^* + I_{Sn4}^*.$$

where  $I_{Sni}^*$  and  $\tilde{r}_{Sn}^{jk*}$  are defined as in equations 7.10–7.12 with the difference that all expectations are replaced with expectations conditional on the data.

**Lemma 7.7.** *Under the assumptions we have that under  $H_0$*

$$nh^{(M+5)/2} I_{Sn1}^* \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_S^2)$$

with probability tending to one and

$$\begin{aligned}nh^{(M+5)/2} I_{Sn2}^* - h^{-(M+1)/2} B_S &\xrightarrow{P} 0 \\ nh^{(M+5)/2} I_{Sn3}^* &\xrightarrow{P} 0 \\ nh^{(M+5)/2} I_{Sn4}^* &\xrightarrow{P} 0\end{aligned}$$

This completes the proof. □

## Proof of the Lemmata

**Proof of Lemma 7.3** Consider  $\Psi(t) = \Gamma_S^{12}(f^1 + tg_1, f^2 + tg_2, m^1, m^2, f_3)$  as a function of  $t$  and write the Taylor expansion around  $t = 0$

$$\Psi(t) = \Psi(0) + t\Psi'(0) + t^2\Psi''(0)/2 + t^3\Psi'''(\vartheta(t))/6$$

Analogously to Lemma 1 in A"it-Sahalia, Bickel and Stoker (2001) we can show that  $\Psi(0) = 0$  and  $\Psi'(0) = 0$  under  $H_0$ . The leading term is given by

$$\Psi''(0) = 2 \int \partial_t \psi(0, z)^2 a(z) f_3(z) dz$$

with

$$\psi(t) = \varphi_1^2(t, z) + m^2(z)\varphi_{M+1}^1(t, z) - \varphi_2^1(t, z) - m^1(z)\varphi_{M+1}^2$$

and

$$\varphi_k^i(t, z) = \partial_k \frac{\int w^i (f^i(w, z) + tg_i(w, z)) dw^i}{f^i(z) + tg_i(z)}.$$

Finally,

$$\Psi'''(\vartheta(t)) = O(\|g_1\|_d^2 \|g_1\|_f + \|g_2\|_d^2 \|g_2\|_f).$$

Evaluating the Taylor expansion at  $t = 1$  gives the statement of the lemma.  $\square$

**Proof of Lemma 7.4** This Lemma is shown using the same combinatoric considerations as in the proof of Lemma 7 in Ait-Sahalia, Bickel and Stoker (2001) and standard kernel methods. Therefore we omit the details here.

**Proof of Lemma 7.5** Before showing the statements of this lemma, we start by investigating  $\tilde{r}_{Sn}(\cdot)$ . Calculating the derivatives we obtain

$$\begin{aligned} r_{Sn}^{jk}(W_i, Z_i; z) &= \alpha^j(W_i, z) h^{-(M+2)} K^k(z, (z - Z_i)/h) \\ &\quad - \alpha^k(W_i, z) h^{-(M+2)} K^j(z, (z - Z_i)/h) + \alpha^j(W_i, z) K_h(z - Z_i) \left( \frac{\partial_k f(z)}{f(z)} + m^k(z) \frac{\partial_{M+1} f(z)}{f(z)} \right) \\ &\quad - \alpha^k(W_i, z) K_h(z - Z_i) \left( \frac{\partial_j f(z)}{f(z)} - m^j(z) \frac{\partial_{M+1} f(z)}{f(z)} \right) \end{aligned} \quad (7.13)$$

Because the sum over the third terms in ?? is zero under  $H_0$ . Here the last two terms converge faster, as the inner derivative of the kernel brings an extra  $h$  to the first two terms. Therefore, further analysis can be restricted to the first two terms and we will omit the lower order terms.

**Asymptotic Normality of  $I_{Sn1}$**   $I_{Sn1}$  can be written as U-Statistic by

$$I_{Sn1} = \sum_{i_1 < i_2} h_n(Y_{i_1}, Y_{i_2})$$

with

$$h_n(Y_{i_1}, Y_{i_2}) = \frac{2}{n^2} \sum_{j < k} \int \tilde{r}_{Sn}^{jk}(W_{i_1}, Z_{i_2}; z) \tilde{r}_{Sn}^{jk}(W_{i_2}, Z_{i_2}; z) a(z) f(z) dz$$

which is by construction centered and degenerated.

Asymptotic normality is established by verifying the conditions of Lemma 7.1. We start with the denominator and note first that since we have independent and identically distributed data we can define  $\sigma_n^2 = \mathbb{E} h_n(Y_i, Y_j)^2$ . Because  $h_n(\cdot, \cdot)$  is centered, we have

$$\text{var } I_{Sn1} = \frac{n(n-1)}{2} \sigma_n^2$$

Direct and tedious calculations, which employ standard techniques of change of variables show that

$$\sigma_n^2 = \mathbb{E}h_n(Y_1, Y_2)^2 = \frac{4}{n^4 h^{M+5}} \sigma_S^2 (1 + O(h)).$$

By the assumption of iid data we have that

$$\max_{1 \leq i \leq n} \sum_{\substack{k=1 \\ k \neq i}}^n \mathbb{E}h_n(Y_i, Y_k)^2 = n\sigma_n^2.$$

Together with the result for the denominator, the first condition of Lemma 7.1 follows.

To verify the second condition, we need to analyze

$$\begin{aligned} \mathbb{E}I_{Sn1}^4 &= \sum_{i_1 < i_2} \mathbb{E}h_n^4(Y_{i_1}, Y_{i_2}) + 3 \sum_{i_1 < i_2} \sum_{\substack{i_3 < i_4 \\ (i_3, i_4) \neq (i_1, i_2)}} \mathbb{E}h_n(Y_{i_1}, Y_{i_2})^2 h_n(Y_{i_3}, Y_{i_4})^2 \\ &\quad + 24 \sum_{i_1 < i_2} \sum_{i_3 \neq i_1, i_2} \mathbb{E}h_n(Y_{i_1}, Y_{i_2})^2 h_n(Y_{i_1}, Y_{i_3}) h_n(Y_{i_2}, Y_{i_3}) \\ &\quad + 3 \sum_{i_1} \sum_{i_2 \neq i_1} \sum_{i_3 \neq i_1, i_2} \sum_{i_4 \neq i_1, i_2, i_3} \mathbb{E}h_n(Y_{i_1}, Y_{i_2}) h_n(Y_{i_2}, Y_{i_3}) h_n(Y_{i_3}, Y_{i_4}) h_n(Y_{i_4}, Y_{i_1}) \quad (7.14) \end{aligned}$$

Again, it follows from direct calculations that

$$\begin{aligned} \mathbb{E}h_n(Y_1, Y_2)^4 &= O(n^{-8} h^{-3M-7}) \\ \mathbb{E}h_n(Y_1, Y_2)^2 h_n(Y_1, Y_3)^2 &= O(n^{-8} h^{-2M-6}) \\ \mathbb{E}h_n(Y_1, Y_2)^2 h_n(Y_1, Y_3) h_n(Y_2, Y_3) &= O(n^{-8} h^{-2M-6}) \\ \mathbb{E}h_n(Y_1, Y_2) h_n(Y_2, Y_3) h_n(Y_3, Y_4) h_n(Y_1, Y_4) &= O(n^{-8} h^{-M-5}) \end{aligned}$$

Using some combinatorics one sees from equation 7.14 that the total contribution of terms of these kinds to  $\mathbb{E}I_{Sn1}^4$  is at most  $O(n^{-4} h^{-(M+1)})$ . So  $\mathbb{E}I_{Sn1}^4$  is asymptotically dominated by terms with  $\mathbb{E}h_n(Y_1, Y_2)^2 h_n(Y_3, Y_4)^2 = (\mathbb{E}h_n(Y_1, Y_2^2))^2$ . Therefore the second condition from Lemma 7.1 is fulfilled as

$$\frac{\mathbb{E}I_{Sn1}^4}{(\text{var } I_{Sn1})^2} = \frac{12n^{-4} h^{-2(M+5)} \sigma_H^4 (1 + o(1))}{(2n^{-2} h^{-(M+5)} \sigma_H^2 (1 + o(1)))^2} \longrightarrow 3$$

and the asymptotic normality of  $I_{n1}$  is established.

**Stochastic Convergence of  $I_{Sn2}$**  The expected value of the test statistic is given by

$$\begin{aligned} \mathbb{E}I_{Sn2} &= \frac{1}{n} \sum_{j < k} \iint (\alpha^j(w_1, z) h^{-(M+2)} K^k(z, (z - z_1)/h) \\ &\quad - \alpha^k(w_1, z) h^{-(M+2)} K^j(z, (z - z_1)/h))^2 a(z) f(z) dz f(w_1, z_1) dw_1 dz_1 \\ &= n^{-1} h^{-(M+3)} B_S + o(n^{-1} h^{-(M+5)/2}) \end{aligned}$$

Stochastic convergence follows from an application of Markov's inequality with second moments, which requires to calculate  $\mathbb{E}I_{n2}^2 = O(n^{-3}h^{-(M+3)})$ . This gives the second statement of the lemma.

**Stochastic Convergence of  $I_{Sn3}$**  Because  $\tilde{r}_n^{jk}(W_i, Z_i; z)$  are centered functions, we have that  $\mathbb{E}I_{Sn3} = 0$ . Substituting  $\bar{z} = (z - z_1)/h$  for  $z_1$  gives

$$\begin{aligned}\mathbb{E}r_{Sn}^{jk}(W_1, Z_1; z) &= h^{-1} \int (\alpha^j(w_1, z)K^k(z, \bar{z}) - \alpha^k(w_1, z)K^j(z, \bar{z}))f(w_1, z)dw_1d\bar{z} + O(h^{r-1}) \\ &= O(h^{r-1})\end{aligned}\quad (7.15)$$

for every  $z \in A$  and therefore

$$\mathbb{E}I_{Sn3}^2 = \frac{4(n-1)^2}{n^3} \mathbb{E} \left( \sum_{j < k} \int \tilde{r}_{Sn}^{jk}(W_i, Z_i; z) \mathbb{E}r_{Sn}^{jk}(W_1, Z_1; z) a(z) f(z) dz \right)^2 = O(n^{-1}h^{2(r-1)})$$

which is of  $o(n^{-1}h^{-(M+5)/2})$ .

**Convergence of  $I_{Sn4}$**  The convergence of the deterministic part follows from 7.15 and the upper bound of the bandwidth sequence

$$I_{Sn4} = O(h^{2(r-1)}) = o(n^{-1}h^{-(M+5S)/2})$$

which completes the proof of the lemma.  $\square$

**Proof of Lemma 7.6** Note that from equation 7.4 follows that only the first part has to be investigated, since the second part in equation 7.4 concerning the density estimator is unchanged in the bootstrap sample and has the desired rate.

Uniform convergence of

$$\hat{e}_h^{k*}(z) = \frac{\hat{g}_h^{k*}(z)}{\hat{f}_h(z)} = \frac{\sum_{i=1}^n K_h(z - Z_i) \varepsilon_i^{k*}}{\sum_{i=1}^n K_h(z - Z_i)}$$

and its derivative with respect to  $z^k$  is proofed by standard methods. A chaining argument and the application of Bernstein's inequality yield the desired result (See Härdle, 1990, or Masry, 1996).  $\square$

**Proof of Lemma 7.7** Define  $Y_i^* = (\varepsilon_i^*, Z_i)$ . To derive the asymptotic distribution of

$$I_{Sn1}^* = \sum_{i_1 < i_2} h_n(Y_{i_1}^*, Y_{i_2}^*)$$

given the data, again Lemma 7.1 has to be applied. This is done by showing that the conditions hold with probability tending to one, i. e.

$$\frac{\max_{1 \leq i \leq n} \sum_{j=1}^n \mathbb{E}^* h_n(Y_i^*, Y_j^*)^2}{\mathbf{var}^* I_{Sn1}^*} \xrightarrow{P} 0$$

$$\frac{\mathbb{E}^*(I_{Sn1}^*)^4}{(\mathbf{var}^* I_{Sn1}^*)^2} \xrightarrow{P} 3.$$

Note, that by construction  $\mathbb{E}^* h_n(Y_1^*, Y_2^*) = 0$  and  $\mathbb{E}^*(h_n(Y_1^*, Y_2^*) \mid Y_1^*) = \mathbb{E}^*(h_n(Y_1^*, Y_2^*) \mid Y_2^*) = 0$  almost surely. Direct calculations reveal that  $\mathbb{E}^* h_n(Y_1^*, Y_2^*)$  has a structure similar to  $h_n(Y_1, Y_2)^2$ . Especially, if expectation is taken, and the appropriate changes of variables are applied, the same leading term can be derived.

To verify the conditions, we first get

$$\mathbf{var}^* I_{n1}^* = \sum_{i_1 < i_2} \mathbb{E}^* h_n(Y_{i_1}^*, Y_{i_2}^*)^2$$

because the bootstrap residuals are independent and  $h_n(Y_{i_1}^*, Y_{i_2}^*)$  is centered conditional on the data. To bound this in probability, use Markov's inequality with the first moment

$$\mathbb{E} \left| \sum_{i_1 < i_2} \mathbb{E}^* h_n(Y_{i_1}^*, Y_{i_2}^*)^2 \right| = \sum_{i_1 < i_2} \mathbb{E} h_n(Y_{i_1}^*, Y_{i_2}^*)^2 = n^{-2} h^{-(M+5)} \sigma_S^2 (1 + o(1))$$

from which

$$\mathbf{var}^* I_{Sn1}^* \xrightarrow{P} \mathbf{var} I_{Sn1}$$

follows. This is now used to show the first condition. By the iid-assumption on the data sample, for the maximum holds

$$\mathbb{P} \left( \frac{\max_{i=1, \dots, n} \sum_{j=1, j \neq i}^n \mathbb{E}^* h_n(Y_i^*, Y_j^*)^2}{\mathbf{var} I_{Sn1}} > c \right) = n \mathbb{P} \left( \frac{\sum_{j=2}^n \mathbb{E}^* h_n(Y_1^*, Y_j^*)^2}{\mathbf{var} I_{Sn1}} > c \right)$$

And for the right hand side we use the Markov inequality with second moments and similar calculations as in lemma 7.5 to obtain

$$\mathbb{P} \left( \frac{\sum_{j=2}^n \mathbb{E}^* h_n(Y_1^*, Y_j^*)^2}{\mathbf{var} I_{Sn1}} > c \right) = O(n^{-2} h^4) = o(n^{-1})$$

For the second condition we again use the convergence of the denominator. Then bounding the probability with the second moment leads to similar calculations as done in Lemma 7.5. Stochastic convergence of  $I_{Sn2}^*$  and  $I_{Sn3}^*$  consists of using iterated expectations and repeating there the same calculations as in this lemma.  $\square$

## 7.4 Proof of Theorem 5

Consider for simplicity the case of one constraint and two goods only. After analyzing this in detail, we will give arguments for why the general case follows directly. Substituting the restriction, the problem becomes:

$$\min_{\theta(z) \in \Theta} n^{-1} \sum_{i=1}^n K_i((\epsilon_i^1)^2 + (\epsilon_i^2)^2),$$

where

$$\begin{aligned} \epsilon_i^1 &= W_i^1 - m^1(z) - \theta_1^1(z)(\bar{P}_i^1 - \bar{p}^1) - \theta_2^1(z)(\bar{P}_i^2 - \bar{p}^2) - \theta_x^1(z)(\bar{X}_i - \bar{x}) \\ &= W_i^1 - \bar{Z}_i' \theta^1 \\ \epsilon_i^2 &= W_i^2 - m^2(z) - (\theta_2^1(z) + \theta_x^1(z)\alpha^2(z) - \theta_x^2(z)m^1(z))(\bar{P}_i^1 - \bar{p}^1) \\ &\quad - \theta_2^2(z)(\bar{P}_i^2 - \bar{p}^2) - \theta_x^2(z)(\bar{X}_i - \bar{x}) \\ &= W_i^2 - \bar{Z}_i' \theta_r^2, \end{aligned}$$

where the bars denote division by  $h$  and  $\bar{Z}_i = (1, \bar{P}_i^1 - \bar{p}^1, \bar{P}_i^2 - \bar{p}^2, \bar{X}_i - \bar{x})'$ . Moreover, we substituted for  $\theta_2^1(z)$ , and we use an additional subscript  $r$ , to denote parameters and estimators which embody the restriction. Then, the FOC can be written as

$$\begin{aligned} \sum_{i=1}^n K_i(W_i^1 - \bar{Z}_i' \tilde{\theta}^1) \bar{Z}_i + \sum_{i=1}^n K_i(W_i^2 - \bar{Z}_i' \tilde{\theta}_r^2) (\bar{P}_i^1 - \bar{p}^1) \tilde{\psi}_1 &= 0 \\ \sum_{i=1}^n K_i(W_i^2 - \bar{Z}_i' \tilde{\theta}_r^2) \bar{Z}_i + \sum_{i=1}^n K_i(W_i^2 - \bar{Z}_i' \tilde{\theta}_r^2) (\bar{P}_i^1 - \bar{p}^1) \tilde{\psi}_2 &= 0, \end{aligned}$$

where  $\tilde{\psi} = (\tilde{\psi}_1', \tilde{\psi}_2')' = ((-\tilde{\theta}_x^2(z), 0, 1, \tilde{m}^2(z)), (\tilde{\theta}_x^1(z), -1, 0, -\tilde{m}^1(z)))$ . Note that the sixth equation is obtained by the trivial equality

$$\sum_{i=1}^n K_i(W_i^2 - \bar{Z}_i' \tilde{\theta}_r^2) (\bar{P}_i^1 - \bar{p}^1) - \sum_{i=1}^n K_i(W_i^2 - \bar{Z}_i' \tilde{\theta}_r^2) (\bar{P}_i^1 - \bar{p}^1) = 0.$$

Inserting  $W_i^j = \bar{Z}_i' \theta^j + B^j(\bar{Z}_i) + U_i^j$ ,  $j = 1, 2$ , where  $B^j(\bar{Z}_i)$  denotes the second order bias terms, we obtain

$$\begin{aligned} \sum_{i=1}^n K_i \bar{Z}_i \bar{Z}_i' (\theta^j - \tilde{\theta}^j) + \tilde{\psi}_j \sum_{i=1}^n K_i (\bar{P}_i^1 - \bar{p}^1) \bar{Z}_i' (\theta^2 - \tilde{\theta}_r^2) + \\ \sum_{i=1}^n K_i \bar{Z}_i (B^j(\bar{Z}_i) + U_i^j) + \tilde{\psi}_j \sum_{i=1}^n K_i (\bar{P}_i^1 - \bar{p}^1) (B^2(\bar{Z}_i) + U_i^2) = 0, \end{aligned}$$

where  $j = 1, 2$ . This system of equations may be rewritten as

$$\begin{aligned}\tilde{\theta}^1(z) - \theta^1(z) &= A_n^{-1} \sum_{i=1}^n K_i (B^1(\bar{Z}_i) + U_i^1) \bar{Z}_i \\ &\quad + A_n^{-1} \tilde{\psi}_1 \sum_{i=1}^n K_i (\bar{P}_i^1 - \bar{p}^1) \bar{Z}_i' (\theta_r^2(z) - \tilde{\theta}_r^2(z)) \\ &\quad + A_n^{-1} \tilde{\psi}_1 \sum_{i=1}^n K_i (B^2(\bar{Z}_i) + U_i^2) (\bar{P}_i^1 - \bar{p}^1),\end{aligned}\tag{7.16}$$

for  $\theta^1(z) = (m^1(z), \theta_1^1(z), \theta_2^1(z), \theta_x^1(z))'$  and

$$\begin{aligned}\tilde{\theta}_r^2(z) - \theta_r^2(z) &= C_n^{-1} A_n^{-1} \sum_{i=1}^n K_i (B^1(\bar{Z}_i) + U_i^1) \bar{Z}_i \\ &\quad + C_n^{-1} A_n^{-1} \tilde{\psi}_2 \sum_{i=1}^n K_i (B^2(\bar{Z}_i) + U_i^2) (\bar{P}_i^1 - \bar{p}^1),\end{aligned}\tag{7.17}$$

for  $\theta_r^2(z) = (m^2(z), (\theta_2^1(z) + \theta_x^1(z)\alpha^2(z) - \theta_x^2(z)\alpha^1(z)), \theta_2^2(z), \theta_x^2(z))'$ . Here,

$$A_n = \sum_{i=1}^n K_i \bar{Z}_i \bar{Z}_i' \quad \text{and} \quad C_n = I - A_n^{-1} \tilde{\psi}_2 \sum_{i=1}^n K_i (\bar{P}_i^1 - \bar{p}^1) \bar{Z}_i'.$$

In equations 7.16 and 7.17 the first term gives  $\hat{\theta}(z) - \theta(z)$ , while the other terms can be seen as additional bias terms. Consider 7.17 first. Recall that  $\tilde{\psi}_2 = (\tilde{\theta}_x^1(z), -1, 0, -\tilde{m}^1(z))'$  and obviously  $\tilde{\psi}_2 = O_p(1)$ , if symmetry is true. Then,

$$A_n^{-1} \tilde{\psi}_2 \sum_{i=1}^n K_i (\bar{P}_i^1 - \bar{p}^1) \bar{Z}_i' = O_p(1),$$

by standard arguments. Thus,  $C_n = I + O_p(1)$ , and the result that  $\tilde{\theta}_r^2(z) - \theta_r^2(z) = (o_p(1), o_p(h), o_p(h), o_p(h))'$  follows from 7.17 by standard arguments.

We will now analyze the behavior of the components  $\tilde{m}^1(z) - m^1(z)$  and  $\tilde{m}^2(z) - m^2(z)$ . Let

$$g_n = \frac{1}{nh^{M+1}} \tilde{\psi}_1 \sum_{i=1}^n K_i (\bar{P}_i^1 - \bar{p}^1) \bar{Z}_i' (\theta_r^2(z) - \tilde{\theta}_r^2(z)).$$

Recall that  $\tilde{\psi}_1 = (-\tilde{\theta}_x^2(z), 0, 1, \tilde{m}^2(z))'$ , and note that by the above consistency result we have  $\tilde{\psi}_1 = (O_p(h), 0, O(1), O_p(1))'$ . Combining this with the last arguments,  $g_n = (o_p(h^2), 0, o_p(h), o_p(h))'$ .

Similarly for

$$k_n = \tilde{\psi}_2 \frac{1}{nh^{M+1}} \sum_{i=1}^n K_i (\bar{P}_i^1 - \bar{p}^1) \bar{Z}_i' (\theta_r^2(z) - \tilde{\theta}_r^2(z)),$$



we obtain  $k_n = (o_p(h^2), 0, o_p(h), o_p(h))'$ . Analogously,

$$\frac{1}{nh^{M+1}} \sum_{i=1}^n K_i(B^2(\bar{Z}_i) + U_i^2)(\bar{P}_i^1 - \bar{p}^1)\tilde{\psi}_1 = (O_p(h^4), 0, O_p(h^3), O_p(h^3))'.$$

Summarizing,

$$\begin{aligned} \tilde{m}^1(z) - m^1(z) &= e' A_n^{-1} \sum_{i=1}^n K_i(B^1(\bar{Z}_i) + U_i^1)\bar{Z}_i + o_p(h^2) \\ &= \hat{m}^1(z) - m^1(z) + o_p(h^2), \end{aligned}$$

where  $e = (1, 0, 0, 0)'$ . Noting that  $\tilde{\psi}_2 = (O_p(h), O(1), 0, O_p(1))'$ , and not just  $O_p(1)$  as stated above, similar arguments produce  $\tilde{m}^2(z) - m^2(z) = \hat{m}^2(z) - m^2(z) + o_p(h^2)$ , and hence

$$\tilde{m}(z) - m(z) = \hat{m}(z) - m(z) + o_p(h^2), \quad (7.18)$$

but note that the same is not true for  $\tilde{\theta}_\beta(z) - \theta_\beta(z)$ . Substituting 7.18 into the first order conditions, we obtain that

$$\begin{pmatrix} \tilde{\theta}_\beta^1(z, \tilde{m}) - \theta_\beta^1(z) \\ \tilde{\theta}_{\beta,r}^2(z, \tilde{m}) - \theta_{\beta,r}^2(z) \end{pmatrix} = \begin{pmatrix} \tilde{\theta}_\beta^1(z, \hat{m}) - \theta_\beta^1(z) \\ \tilde{\theta}_{\beta,r}^2(z, \hat{m}) - \theta_{\beta,r}^2(z) \end{pmatrix} + o_p(h^2),$$

where  $\tilde{\theta}_\beta^1(z, \tilde{m}) = \tilde{\theta}_\beta^1(z)$  and  $\tilde{\theta}_{\beta,r}^2(z, \tilde{m}) = \tilde{\theta}_{\beta,r}^2(z)$  are the left hand side sub-vectors of  $h$ -scaled local slope coefficients in 7.16 and 7.17, with their dependence on the respective levels made explicit. This is easy to see, as the  $\tilde{m}$  enter linearly or quadratically. Since  $A_n$  is asymptotically diagonal, this means that – up to terms of smaller order than the leading bias term – the same first order conditions on the slopes could have been obtained by the second step of the two step procedure introduced above.

This argumentation continues to hold with more prices and restrictions as they simply add constraints which are linearizable in a similar fashion. In particular, as is easily seen the asymptotic distribution of the  $\hat{m}$  remains unchanged as only additional terms of  $o_p(h^2)$  have been added. Hence the same line of argumentation can be carried through.

## 7.5 Proof of Theorem 6

Start by noticing that the unrestricted estimator for the  $h$ -scaled slopes,  $\hat{\theta}_\beta(z)$ , is related to our two-step auxiliary estimator via

$$\begin{aligned} \check{\theta}_\beta(z) &= \left( I - Q_n^{-1} R(\hat{m}(z))' (R(\hat{m}(z)) Q_n^{-1} R(\hat{m}(z))')^{-1} R(\hat{m}(z)) \right) \hat{\theta}_\beta(z) \\ &= H_n(\hat{m}(z)) \hat{\theta}_\beta(z), \end{aligned}$$

where

$$Q_n = \begin{pmatrix} I_{M-1} \otimes (\tilde{Z}'_0 \tilde{Z}_0) & I_{M-1} \otimes (\tilde{Z}'_0 \tilde{Z}) \\ I_{M-1} \otimes (\tilde{Z}' \tilde{Z}_0) & I_{M-1} \otimes (\tilde{Z}' \tilde{Z}) \end{pmatrix},$$

$\tilde{Z} = (\tilde{Z}_1, \dots, \tilde{Z}_{M+1})$ , with  $\tilde{Z}_0 = (K_1^{1/2}, \dots, K_n^{1/2})'$  and  $\tilde{Z}_m = (K_1^{1/2}(Z_{m1} - z_m), \dots, K_n^{1/2}(Z_{mn} - z_m))'$  for all  $m = 1, \dots, M+1$ . In order to show the theorem, we need to determine the behavior of the leading bias and variance terms. As is common in this literature, we consider the leading term in  $\mathbf{cov}(\hat{m}(z), \check{\theta}_\beta(z) \mid Z_1, \dots, Z_n)$  and  $\mathbf{var}(\check{\theta}_\beta(z) \mid Z_1, \dots, Z_n)$  and show that this is free of the data. To this end, consider first

$$\begin{aligned} & \mathbf{var}(H_n(\hat{m}(z))\hat{\theta}_\beta(z) \mid Z_1, \dots, Z_n) \\ &= H_n(m(z)) \mathbf{var}(\hat{\theta}_\beta(z) \mid Z_1, \dots, Z_n) H(m(z))' \\ & \quad + D_\alpha H_n(m(z)) \mathbf{var}((\hat{m}(z) - m(z)) \otimes I_{\dim(\theta_\beta)} \hat{\theta}_\beta(z) \mid Z_1, \dots, Z_n) D_\alpha H_n(m(z))' + \xi \end{aligned}$$

where  $\xi$  contains second-order terms evaluated at intermediate positions. At this point note that the first term is  $O_p(n^{-1}h^{-(M+1)})$ , but the second is  $O_p(n^{-1}h^{2-(M+1)})$ , due to  $\theta_\beta(z) = h\beta(z)$ . Further, rewrite  $H_n(m(z)) = H(m(z)) + [H_n(m(z)) - H(m(z))]$ , and  $D_\alpha H_n(m(z)) = D_\alpha H(m(z)) + [D_\alpha H_n(m(z)) - D_\alpha H(m(z))]$ , and observe that the respective terms converge uniformly in probability. Therefore, to determine the behavior of  $\mathbf{var}(H_n(\hat{m}(z))\hat{\theta}_\beta(z) \mid Z_1, \dots, Z_n)$ , we can rewrite the lhs of as follows:

$$\begin{aligned} & H(m(z)) \mathbf{var}(\hat{\theta}_\beta(z) \mid Z_1, \dots, Z_n) H(m(z))' \\ & \quad + D_\alpha H(m(z)) (\mathbf{var}(\hat{m}(z) \mid Z_1, \dots, Z_n) \otimes \theta_\beta(z) \theta_\beta(z)') D_\alpha H(m(z))' \\ & \quad + D_\alpha H(m(z)) \mathbf{var}((\hat{m}(z) - m(z)) \otimes (\hat{\theta}_\beta(z) - \theta_\beta(z)) \mid Z_1, \dots, Z_n) D_\alpha H(m(z))' + \xi, \end{aligned}$$

To evaluate the orders of the second and third expression, examine without loss of generality the first element of  $\mathbf{var}((\hat{m}(z) - m(z)) \otimes (\hat{\theta}_\beta(z) - \theta_\beta(z)) \mid Z_1, \dots, Z_n)$ ,

$$\begin{aligned} & \mathbf{var}((\hat{\alpha}_1(z) - \alpha_1(z))(\hat{\theta}_{\beta,1}(z) - \theta_{\beta,1}(z)) \mid Z_1, \dots, Z_n) \\ & \leq (\mathbb{E}((\hat{\alpha}_1(z) - \alpha_1(z))^4 \mid Z_1, \dots, Z_n))^{1/2} (\mathbb{E}((\hat{\theta}_{\beta,1}(z) - \theta_{\beta,1}(z))^4 \mid Z_1, \dots, Z_n))^{1/2} \\ & \quad + \mathbb{E}((\hat{\alpha}_1(z) - \alpha_1(z))^2 \mid Z_1, \dots, Z_n) \mathbb{E}((\hat{\theta}_{\beta,1}(z) - \theta_{\beta,1}(z))^2 \mid Z_1, \dots, Z_n). \end{aligned}$$

Here, the leading term on the right-hand side is  $O_p(h^{4r} + n^{-2}h^{2-(M+1)})$  which is under general conditions  $O_p(n^{-1}h^{2-(M+1)})$ . Higher order terms converge even faster, provided the second derivatives of  $H(m(z))$  are uniformly bounded. Hence,

$$nh^{M+1} \mathbf{var}(H(\hat{m}(z))\hat{\theta}_\beta(z) \mid Z_1, \dots, Z_n) = nh^{M+1} H(m(z)) \mathbf{var}(\hat{\theta}_\beta(z) \mid Z_1, \dots, Z_n) H(m(z))' + o_p(1).$$

By similar arguments

$$\begin{aligned} nh^{M+1} \mathbf{cov}(\hat{m}(z), H(\hat{m}(z))\hat{\theta}_\beta(z) \mid Z_1, \dots, Z_n) \\ = nh^{M+1} \mathbf{cov}(\hat{m}(z), \hat{\theta}_\beta(z) \mid Z_1, \dots, Z_n) H(m(z))' + o_p(1), \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}(H(\hat{m}(z))\hat{\theta}_\beta(z) \mid Z_1, \dots, Z_n) - H(m(z))\theta_\beta(z) \\ = h^2 H(m(z)) \text{bias}_{\hat{\theta}_\beta}(z) + h^2 D_\alpha H(m(z)) (\text{bias}_{\hat{\alpha}}(z) \otimes I_{\dim(\theta_\beta)}) \theta_\beta(z) + o_p(h^2), \end{aligned}$$

where  $\text{bias}_{\hat{\alpha}}(z)$  and  $\text{bias}_{\hat{\theta}_\beta}(z)$  denote the bias terms belonging to the levels and slopes in a locally linear models respectively. However, note that due to  $\theta_\beta(z) = h\beta(z)$ , the leading bias term is the term of lowest order. Finally, the theorem follows upon combining these results with theorem 4, yielding

$$\begin{aligned} \sqrt{nh^{M+1}} \begin{pmatrix} \hat{m}(z) - m(z) - h^2 \text{bias}_{\hat{\alpha}}(z) \\ \check{\theta}_\beta(z) - \theta_\beta(z) - h^2 H(m(z)) \text{bias}_{\hat{\theta}_\beta}(z) \end{pmatrix} \\ \xrightarrow{\mathcal{D}} \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, f(z)^{-1} \begin{pmatrix} \Sigma(z) \kappa_0^{M+1} & (\Sigma(z) \otimes b') \Xi' \\ \Xi (\Sigma(z) \otimes b) & \Xi (\Sigma(z) \otimes A) \Xi' \end{pmatrix} \right). \end{aligned}$$

Using the result of theorem 5 on the asymptotic equivalence of the two step and the nonlinearly restricted Lagrangian estimator, the theorem follows.

## 7.6 Simulation

Simulation methods extend on the Smooth Conditional Moment bootstrap of Gozalo (1997). In particular, we match the conditional variance of error terms across equations, and the conditional third moment of each equation's error term. We refer to this simulation procedure as the Multi-Variate Smooth Conditional Moment (MVSCM) bootstrap. The procedure requires estimation of conditional moments, and then simulation of error terms satisfying these conditional moment restrictions. Estimation of conditional moments is via local quadratic estimation of squared and cubed error terms on  $z$ , with  $\hat{h}^{\text{Moments}} = 3.2$ , in order to reduce bias in moment estimation.

### 7.6.1 Multivariate Smooth Conditional Moment Bootstrap

To generate bootstrap residuals the first moments of the conditional distribution of  $W_i$  given  $Z_i$  has to be replicated. In the classical wild bootstrap by Härdle and Mammen (1993) only one observation ( $\hat{\varepsilon}_i$ ) is used to estimate these conditional moments. In our context however, it seems

reasonable to assume these moments to change in a smooth fashion. For this case, Gozalo (1997) proposes to use smoothed estimators of the conditional variance and the conditional third moment of  $W_i$  given  $Z_i$ . This procedure is called Smooth Conditional Moment (SCM) bootstrap and has been developed for one-dimensional residuals only. To apply this to our model, we have to extend it to multivariate SCM bootstrap:

For each  $i$  the bootstrap residual  $\varepsilon_i^*$  is drawn from an arbitrary distribution  $\hat{F}_i$  such that

$$\begin{aligned}\mathbb{E}^{\hat{F}_i} \varepsilon_i^* &= 0 \\ \mathbb{E}^{\hat{F}_i} \varepsilon_i^* (\varepsilon_i^*)' &= \hat{\Sigma}_h(Z_i)\end{aligned}$$

and in addition we restrict some of the third moments

$$\mathbb{E}^{\hat{F}_i} (\varepsilon_i^{j*})^3 = \hat{\mu}_h^j(Z_i) \quad j = 1, \dots, M-1$$

where  $\hat{\Sigma}_h(z) = (\hat{\sigma}_h^{ij}(z))_{i,j}$  is the matrix of estimated (co)variances and  $\hat{\mu}_h(z) = (\hat{\mu}_h^1(z), \dots, \hat{\mu}_h^{M-1}(z))'$  is the vector of kernel estimators of the third conditional moments  $E((W^j - m^j(Z))^3 \mid Z = z)$ . Define  $\hat{S}_h(z) = (\hat{s}_{ij}(z))_{i,j}$  by  $\hat{S}_h(z) \hat{S}_h(z)' = \hat{\Sigma}_h(z)$  using an eigenvector decomposition of  $\hat{\Sigma}_h(z)$ .

To find distributions  $\hat{F}_i$  we construct uncorrelated residuals  $\eta_i^* = (\eta_i^{1*}, \dots, \eta_i^{M-1,*})'$  from a distribution  $\hat{F}_i^\eta$  with

$$\begin{aligned}\mathbb{E}^{\hat{F}_i^\eta} \eta_i^* &= 0 \\ \mathbb{E}^{\hat{F}_i^\eta} \eta_i^* (\eta_i^*)' &= \boldsymbol{\iota} \\ \mathbb{E}^{\hat{F}_i^\eta} (\eta_i^{j*})^3 &= \hat{\lambda}^j(Z_i)\end{aligned}$$

where

$$\hat{\lambda}(z) = (\hat{\lambda}^1(Z_i), \dots, \hat{\lambda}^{M-1}(Z_i))' = \hat{\Gamma}(Z_i)^{-1} \hat{\mu}(Z_i)$$

with  $\hat{\Gamma}(z) = ((\hat{s}_{ij}(z))^3)_{i,j}$ . Then

$$\varepsilon_i^* = \hat{S}_h(Z_i) \eta_i^*$$

gives the bootstrap residuals. Simple matrix algebra shows that  $\varepsilon_i^*$  has the desired moments.  $\eta_i^*$  can be drawn from classical two-point distributions or a continuous mixture of normals which we use in our application.

### 7.6.2 Pointwise Confidence Intervals

Pointwise confidence intervals are estimated by a wild bootstrap procedure as described in Härdle (1990), Theorem 4.2.2. The distribution of  $\hat{m}_h(x) - m(x)$  is approximated by the distribution of  $\hat{m}_h^*(x) - \hat{m}_{h^*}(x)$  where  $\hat{m}_h^*(x)$  is estimated from the bootstrap sample  $(W_i^*, Z_i^*)_{i=1, \dots, n}$  with  $W_i^* = \hat{m}_{h^*}(x) + \varepsilon_i^*$ . Here  $\varepsilon_i^*$  is drawn from the MVSCM-bootstrap as described above (using residuals  $\hat{\varepsilon}_i = W_i - \hat{m}_h(Z_i)$ ) and  $h^* = 1.25 \times h$ .

## 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] Banks, Blundell, R. and A. Lewbel. 1997. Quadratic Engel curves and consumer demand, *Review of Economic studies* 79, 527–539.
- [3] Bierens, H. J. 1982. Consistent model specification tests, *Journal of Econometrics* 20, 105–134.
- [4] Bierens, H. J. and W. Ploberger. 1997. Asymptotic theory of integrated conditional moment tests, *Econometrica* 65, 1129–1151.
- [5] Blundell, R. and J.-M. Robin. 1999. Estimation in large and disaggregated demand systems: an estimator for conditionally linear systems *Journal of Applied Econometrics* 14, 209–232
- [6] Blundell, R., A. Duncan and K. Pendakur. 1998. Semiparametric estimation of consumer demand, *Journal of Applied Econometrics* 13, 435–461.
- [7] Brown, R. and M.B. Walker, (1989); The random utility hypothesis and inference in demand systems, *Econometrica*, 57, 815–829.
- [8] Deaton, A. and A. Tarozzi. 2000. Prices and Poverty in India, Research Program in Development Studies Working Paper, Princeton University.
- [9] Epstein, L. and A. Yatchew. 1985, Nonparametric Hypothesis Testing and Applications to Demand Analysis, *Journal of Econometrics* 30, 149–169.
- [10] Fan, J. and I. Gijbels. 1996. *Local Polynomial Modelling and Its Applications*, Chapman and Hall, London.
- [11] Fan, Y. and Q. Li. 1999. Central limit theorem for degenerate U-Statistics of absolutely regular processes with applications to model specification testing, *Journal of Nonparametric Statistics* 10, 245–271.
- [12] Fan, Y. and Q. Li. 2000. Consistent model specification tests: Kernel based tests versus Bierens’ ICM tests, *Econometric Theory* 16, 1016–1041.
- [13] Gallant, A.R. and G. Souza. 1991. On the asymptotic normality of Fourier flexible form estimates, *Journal of Econometrics* 50 329–353.

- [14] Gozalo, P. 1997. Nonparametric bootstrap analysis with applications to demographic effects in demand functions, *Journal of Econometrics* 81, 357–393.
- [15] Härdle, W. 1990. *Applied Nonparametric Regression*, Cambridge University Press, Cambridge UK.
- [16] Härdle, W. and E. Mammen. 1993. Comparing nonparametric vs. parametric regression fits, *Annals of Statistics* 21, 1926–1947.
- [17] Härdle, W. and A. Tsybakov. 1997. Local polynomial estimators of the volatility function in nonparametric regression, *Journal of Econometrics* 81, 223–242.
- [18] Hjellvik, V. and D. Tjøstheim. 1995. Nonparametric tests of linearity for time series. *Biometrika* 82, 351–368.
- [19] Hjellvik, V., Yao, Q. and D. Tjøstheim. 1998. Linearity testing using local polynomial approximation, *Journal of Statistical Planning and Inference* 68, 295–321
- [20] Haag, B. R., Hoderlein, S. and K. Pendakur. 2004. Testing and imposing economic restrictions in nonparametric demand systems, Working Paper, University of Mannheim.
- [21] Hoderlein, S. 2002. *Econometric Modelling of Heterogeneous Consumer Behavior - Theory, Empirical Evidence and Aggregate Implications*, PhD Thesis, LSE.
- [22] Hoderlein, S. 2007. How many consumers are rational?, Working Paper, University of Mannheim.
- [23] de Jong, P. 1987. A central limit theorem for generalized quadratic forms, *Probability Theory and Related Fields* 75, 261–275.
- [24] Kim, W. and G. Tripathi. 2003. Nonparametric estimation of homogeneous functions. *Econometric Theory*, 19, 640–663.
- [25] Kreiss, J.-P., M. H. Neumann and Q. Yao (2002). Bootstrap tests for simple structures in nonparametric time series regression, Working Paper, London School of Economics.
- [26] Lewbel, A. 1995. Consistent nonparametric hypothesis tests with an application to Slutsky symmetry, *Journal of Econometrics* 67, 379–401.
- [27] Lewbel, A., (2001); Demand Systems With and Without Errors, *American Economic Review*, 611–18.

- [28] Mas-Colell, A., M. D. Whinston and J. R. Green. 1995. *Microeconomic Theory*. Oxford: Oxford University Press.
- [29] Masry, E. 1996. Multivariate local polynomial regression for time series: uniform strong consistency, *Journal of Time Series Analysis* 17, 571–599.
- [30] Pendakur, K. 1999. Nonparametric estimation and testing of base-independent equivalence scales, *Journal of Econometrics* 88, 1-40.
- [31] Tripathi, G. 1998. Nonparametric estimation and testing of homogenous functional forms. Working paper, University of Wisconsin.