Nonparametric Demand Systems, Instrumental Variables and a Heterogeneous Population

Stefan Hoderlein
Mannheim University

September 2004

Abstract

This paper is concerned with empirically modelling the demand behavior of a population with heterogeneous preferences under a weak conditional independence assumption. More specifically, we characterize the testable implications of negative semidefiniteness and symmetry of the Slutsky matrix across a heterogeneous population without assuming anything on the functional form of individual preferences. In the same spirit, implications of a linear budget set are being considered.

Since the conditional independence assumption is the only substantial restriction in this model, we analyze possible alternatives and solutions if this assumption is violated. In particular, we consider in detail the concept of instruments in this framework.

Besides being able to integrate econometric concepts, the same framework admits also economic extensions. As an example we consider welfare analysis.

Finally, we provide asymptotic distribution theory for the new test statistics that emerge out of this framework, and apply these to Canadian data.

1 Introduction

Economic theory yields strong implications for the actual behavior of individuals. This is particularly true for demand theory, where a couple of well-known restrictions arise. All restrictions imposed by rationality on demand behavior are qualitative in nature, which means that they do not predict a specific functional form for the demands of individuals. To test the implications of rational behavior, by and large two strands of literature have emerged. The first uses revealed preference theory, is nonparametric in nature and concentrates on violations of the Strong Axiom in observable data. Key contributions are Afriat (1967) and Varian (1982). More recently, a similar approach has been suggested by Blundell, Browning and Crawford (2002).

The second strand of literature tests a couple of restrictions on demand behavior, using fully specified parametric demand systems. This literature dates back to at least the fifties (Stone (1954)), but has really peaked with the advent of fully flexible functional form demand systems. More recent examples are the Translog, Jorgenson et al. (1982), the AIDS, Deaton and Muellbauer (1980), Blundell et al (1993), or the "exact QUAIDS", Banks et al.(1997), see also Lewbel (1999) for a comprehensive survey. Obviously, both approaches have their limitations: The first usually leads to tests of low power, as price movements are dwarfed by movements in income, and concentrates on one specific property only. The second suffers from the limitations that demands take a certain functional form and that the introduction of preference heterogeneity has not been solved convincingly (see, e.g. Brown and Walker (1989)).

We aim in this paper at laying the foundations for nonparametric demand systems, ideally combining the advantages of both approaches: Being nonparametric in nature, i.e. not specifying any functional form, and still able to judge the restrictions imposed by rationality robustly as well as comprehensively. Since we model individual heterogeneity by a nonseparable model, there are some similarities to the work of Imbens and Newey (2003) and Matzkin (2003). However, their general approach is different, as will become apparent in the following.

This paper makes several contributions: First, we provide a framework for nonparametric demand analysis in a heterogeneous population. Second, within this framework we establish testable implications of the key elements of demand theory. In so far, this approach has some parallels to those of Lewbel (1990, but especially 2001, Theorem 1) and Brown and Walker (1989). However, these approaches are significantly extended in that new and much more general testable consequences are derived. For instance, a very general test for Slutsky negative semidefiniteness is derived that has no remote similarity to anything in the literature. These two contributions are being considered in the second section.

Another important contribution is the generalization of the econometric concepts of endogeneity and instruments to this framework, which occupies the third section. It is a major result of this analysis that all of the general results regarding the identifiability of structures inheriting properties from economic theory continue to hold under this weaker set of assumptions.

As a fourth contribution we discuss welfare analysis, both on the micro- and on the aggregate level, within this framework of a heterogeneous population. This will be our concern in the fourth section.

All contributions made thus far have been confined to the identification of economically relevant structures. As a second set of contributions we provide the econometric theory for the tests that emerge out of the identification sections. Individual contributions include first the joint asymptotic distribution of estimators for levels and derivatives of the mean regression and the scedastic function in a systems locally linear model. One innovation here is the inclusion of both pre-estimated regressors and pre-estimated dependent variables. Building on this result, we analyze the asymptotic behavior of nonparametric local tests for symmetry and homogeneity, both in the baseline as well as in the endogeneity scenario. Moreover, we propose a test for endogeneity in this nonparametric regression framework. The fifth section covers this econometric part.

Finally, a brief application and an outlook conclude this paper.

2 The Demand Behavior of a Heterogeneous Population

2.1 Structure of the Model

Demand theory assumes that the demand of individual i is the result of a well behaved utility maximization problem, yielding a demand function

$$w_i = \phi(p, y_i, u_i), \tag{2.1}$$

where w_i, p and y_i are budget shares, log prices and log total expenditure, vectors of length L, L and 1, respectively. Furthermore, $u_i = u_i(\cdot)$ denotes the individual's utility function. Throughout, we restrict ourselves to continuously differentiable demand functions, which restricts preferences to be itself

continuous, strictly convex and locally nonsatiated, with utility function everywhere twice differentiable. Moreover, we follow the demand literature in assuming that preferences be additively separable over time, which justifies the use of total expenditure instead of income.
The existence of the $\phi(\cdot)$ functional (from now on called *theoretical microrelation)* can be derived from the argmax operator, i.e. a rule that relates these variables. The theoretical properties of this functional are as follows: For fixed u_i, say u_0, $\phi(\cdot, \cdot, u_0)$ behaves like a standard rational demand function, which obeys the usual conditions of rational behavior, e.g. the compensated price derivatives form the negative semidefinite and symmetric Slutsky matrix.
In order to avoid technical difficulties arising with the differentiation on function spaces, we shall assume henceforth that u_i may be completely described by a finite vector $v_i = (v_{1i}, \ldots, v_{Mi})$ of parameters[1]. Therefore we consider ϕ as a $[0,1]^L$ valued function defined on $\mathbb{R}^L \times \mathbb{R} \times \mathbb{R}^M$, continuously differentiable in p and y. This set of assumptions characterizing the space of admissable preferences will be maintained throughout the paper, and denoted by (P). We will also confine ourselves to observationally distinct preferences, i.e. if $v_j, v_k \in \mathbb{R}_+^M$ and $v_j \neq v_k$, then there exist some $p_0, y_0 \in \mathbb{R}^L \times \mathbb{R}$ such that $\phi(p_0, y_0, v_1) \neq \phi(p_0, y_0, v_2)$. We strengthen this assumption by requiring that the same inequality holds for the derivatives.

If we interpret each individual as a realization from an underlying population, we can give the equivalent formulation to (**??**) in terms of random variables. We assume that (W_i, Y_i, V_i) and all other random variables to appear below, denoted as random vector by G_i, are iid with $(W_i, Y_i, V_i, G_i) \backsim (W, Y, V, G)$, where the latter denote the population variables. Also, for simplicity of exposition, we consider p to be a positive nonrandom vector. This is immaterial for our argumentation as the same arguments go through if prices depend on time series randomness alone, while other variables exhibit cross-section variation. Also, the case of prices varying across the population can easily be accommodated within this framework, see Hoderlein (2002) on both issues. Summarizing, we have

Assumption 2.1 *Let all variables and functions be as defined above. Then,*

[1]This does not mean that the concepts can not be defined more generally, see Hoderlein (2002), who uses Frechet-derivatives (see Luenberger (1997)). Little is, however, gained in terms of Economic understanding.

demand is given by

$$W = \phi(p, Y, V) \tag{2.2}$$

As our aim is to establish the link between the theoretical microrelation and empirically estimable quantities, we consider the conditional average[2]. The conditioning here is on observables, where the set of observables obviously depends on the information at hand.
To relate preferences to observables, we assume that every preference depends on the individuals' *current* observable and unobservable attributes, denoted as random vectors by Z and A respectively. Here, Z denotes all observable household attributes (like age, household size, etc.). However, economic choice variables should be excluded, because otherwise preferences would become endogenous, i.e. would depend on other economic choices.

The variable A in turn is meant to capture individual specific unobservables. These could in principle be time-varying as well as infinite dimensional, however, for simplicity of exposition we desist from this greater generality and consider only the case of a finite (S-dimensional) and time invariant vector[3]. This leads to the following

Assumption 2.2 *Let all variables be as defined above. Then*

$$V = \vartheta(Z, A), \tag{2.3}$$

where ϑ *is a fixed Borel-measurable* $\mathbb{R}^M$*-valued function defined on the set* $\mathcal{Z} \times \mathcal{A}$ *of possible values of* (Z, A).

So far we have defined all main components of our framework. To state the next assumption, which ensures that interchanging differentiation and integration is well defined, we need the following notation: Let F_G be the cumulative distribution function of a random variable G, and denote by $F_{G|H}$ the conditional cdf of G given H.
Let $m(p, y, z) = \mathbb{E}[W|Y = y, Z = z] = \mathbb{E}[\phi(p, Y, V)|Y = y, Z = z]$ denote the empirical regression function. Moreover, let $\partial_x f$ denote the partial derivative of a vector valued function f with respect to a scalar x, and $D_x f$ denote the

[2]The above mentioned observational distinctness is now formulated as follows: There exist two sets $\Omega_0, \Omega_1 \subset \Omega$ with $\mathbb{P}[\Omega_l] > 0$, $l = 0, 1$. For all $\omega_j \in \Omega_0$ and $\omega_k \in \Omega_1$, we have $Y(\omega_j) = y_0 = Y(\omega_k)$, $V(\omega_j) = v_0 \neq v_1 = V(\omega_k)$ and $\phi(p_0, Y(\omega_j), V(\omega_j)) \neq \phi(p_0, Y(\omega_k), V(\omega_k))$ for some $p_0 \in \mathbb{R}^L$.

[3]Both complications can be handled by the methods below.

matrix of derivatives of a vector valued function f with respect to a vector x. Finally, whenever convenient we suppress the arguments of the respective functions.

Assumption 2.3: *There exists a function* g, *s. th.*

$$\left\| \left(\ \partial_y \phi(p,y,\vartheta(z,a))' \quad vec\left[D_p \phi(p,y,\vartheta(z,a)) \right]' \ \right) \right\| \leq g(a) \ ,$$

with $\int g(a) F_A(da) < \infty$, *uniformly in* (p,y,z).[4]

Note that this assumption implies that every element of the $M \times M$ matrix $\partial_y \left\{ \phi(p,y,\vartheta(z,a)) \phi(p,y,\vartheta(z,a))' \right\}$ is also uniformly bounded in absolute value[5].

Finally, the last assumption is

Assumption 2.4: $F_{A|Y,Z} = F_{A|Z}$.

Basically, this assumption states that - conditional on Z - income and unobserved heterogeneity are distributed independently. To give an example, take a subgroup of the population, e.g. catholic female students living in small university towns. Suppose there are only two income classes for this group, rich and poor, and two types of preferences, type 1 and 2. Then, for both rich and poor individuals within this subgroup, the proportion of type 1 and 2 must be identical.

This is obviously a substantial assumption, needed in this strength due to the generality of the other assumptions. Nevertheless, in every subgroup of the society there is certainly a tendency towards social cohesion, towards a relatively homogeneous preference structure. Hence, if the information set is very large and allows to identify these more homogeneous subgroups, this assumption may approximately hold. Conversely, if our information set is small, we are likely to mix up subgroups whose distribution may well be correlated with income. This would lead to a breakdown of $A2.4$.

Finally, as we will demonstrate in section 3 below, assumption $A2.4$ may be relaxed along several lines.

---

[4] Among the primitive economic conditions that ensure that this assumption holds are: strict convexitiy, local nonsatiation and continuous differentiability of the preferences, a linear budget constraint and $p >> 0$. Here, *vec* denotes the operator that stacks a matrix into a single vector columnwise.

[5] To see this, take j,k arbitrary and consider $\left| \partial_y \phi_j(p,y,\vartheta(z,a)) \phi_k(p,y,\vartheta(z,a)) \right|$. But this is smaller than $\left| \partial_y \phi_j(p,y,\vartheta(z,a)) \right| \left| \phi_k(p,y,\vartheta(z,a)) \right|$. Since the second term is smaller than one for all p,y,z,a, and the first term is uniformly bounded by $A2.3$, the statement follows.

2.2 Implications for Conditional Moments

Given these assumptions and notations, we concentrate first on the relation of theoretical quantities and the conditional moments. Specifically, we focus on the following questions:

1. How are the empirically obtained derivatives ($\partial_y m$, $D_p m$) with respect to prices and income related to the theoretical ones ($\partial_y \phi$, $D_p \phi$)?

2. How and under what kind of assumptions do elements of observable behavior allow inference on key elements of economic theory. Especially, what does observable behavior tell us about homogeneity, adding up as well as negative semidefiniteness and symmetry of the Slutsky-matrix

$$S(p,y,v) = D_p\phi(p,y,v) + \partial_y\phi(p,y,v)\phi(p,y,v)' + \phi(p,y,v)\phi(p,y,v)' - diag\left\{\phi(p,y,v)\right\},$$

where $diag\{m\}$ denote the matrix having the m_j, $j = 1,..,L$ on the diagonal and zero off the diagonal. These concepts are commonly known as "rationality" in this scenario[6], and shall be subject of Proposition 2.2.

Let us start with the very trivial Proposition 2.1 which establishes the relationship between the derivatives. In what follows all equalities are meant to hold almost surely.

Proposition 2.1

Let all the variables and functions be as defined above. Assume that (P) *and* $(A2.1)$ - $(A2.3)$ *hold. Then follows that* (i) $D_p m(p,Y,Z) = \mathbb{E}[D_p\phi(p,Y,V)|Y,Z]$. *If in addition (A2.4) holds, we have (ii)* $\partial_y m(p,Y,Z) = \mathbb{E}[\partial_y\phi(p,Y,V)|Y,Z]$. *Moreover, iff* V *is* Z*-measurable, then* (iii) $\partial_y m(p,Y,Z) = \partial_y\phi(p,Y,V)$ *as well as* $D_p m(p,Y,Z) = D_p\phi(p,Y,V)$.

Proof: Appendix.

Parts (i) and (ii) of this proposition state that each individual's empirically obtained marginal effect is the best approximation (in the sense of minimizing distance with respect to L_2-norm) to the individual's theoretical marginal effect. For price derivatives, this holds under virtually no conditions at all, for income derivatives we have to invoke the additional assumption $A2.4$, because the individually varying income effects are not to be confounded with

[6] We adopt this language. For other definitions of rationality, see Chiappori and Rochet (1987).

the individually varying preference heterogeneity that is correlated with income. In this general scenario, this is as close as current observables allow us to get to the true marginal effects. Usually, the empirical coefficients will still be an average across individuals with the same realization of Z, and the preference-induced heterogeneity will still be bigger than the observed heterogeneity. However, the third part of the proposition gives a condition on the information needed for both to coincide: all individual randomness that affects demand must be fully captured by current observables.
Note that $A2.4$ could be relaxed to a local independence condition $\partial_y F_{A|Y,Z}(a,y,z) = 0$, $y \in [y_0, y_1]$ for a certain fixed y_0 and y_1 if we were just interested in the marginal effects of a subgroup of the population defined by income. Regarding the average across a population or a subgroup, the following corollary holds:

Corollary 2.2

Let all the variables and functions be as defined above. Suppose that (P) *and* $(A2.1)$ - $(A2.4)$ *hold. Then follows* $\mathbb{E}[\partial_y m(p,Y,Z)|\mathcal{F}] = \mathbb{E}[\partial_y \phi(p,Y,V)|\mathcal{F}]$, *for any* $\mathcal{F} \subseteq \sigma\{Y,Z\}$. *In particular* $\mathbb{E}[\partial_y m(p,Y,Z)] = \mathbb{E}[\partial_y \phi(p,Y,V)]$. *A similar condition holds for* D_p *under* (P), $(A2.1)$ - $(A2.3)$.

Proof: Appendix.

Thus, the average of the empirical marginal effects over the whole population or over a subgroup coincides almost surely with the true average marginal effect across population or subpopulation. Hence, we could see the average marginal effect as a *treatment effect under exogeneity*.

Another straightforward corollary concerns the standard practice of inferring elasticities from the observed regression function. It is instructive to note that the average elasticities have to be calculated from the *log* budget share regressions, and can in general not be obtained by a transformation of the coefficients of the budget share regression (see Hoderlein (2004) for details).

We turn now to the question which economic properties in the heterogeneous population have testable counterparts. This problem bears some similarities with the literature on aggregation over agents in demand theory, because taking conditional expectations can be seen as an aggregation step, as long as the measurability condition of $P2.1$ (iii) is not met. We introduce the following notation: Let $\mathbb{V}[G, H|\mathcal{F}]$ denote the conditional covariance matrix between two random vectors G and H, conditional on some σ-algebra $\mathcal{F}$, and $\mathbb{V}[H|\mathcal{F}]$ be the conditional covariance matrix of a random vector H. Moreover, let $m_2(p, y, z) = \mathbb{E}[WW'|Y = y, Z = z]$. Now we are in the position to state the following

Proposition 2.3

Let all the variables and functions be as defined above, and suppose that (P), $(A2.1)$ - $(A2.3)$ *hold.*

(i) *If* ϕ *fulfills* $\iota'\phi = 1$ $(a.s.) \Rightarrow \iota' m = 1$ $(a.s.)$.

Let additionally $(A2.4)$ *hold as well. Then follows that*

(ii) *If* ϕ *fulfills* $\phi(p + \lambda, Y + \lambda, V) = \phi(p, Y, V)$ $(a.s.) \Rightarrow D_p m\iota + \partial_y m = 0$ *and* $m(p + \lambda, Y + \lambda, Z) = m(p, Y, Z)$ $(a.s.)$.

(iii) *If* S *is negative semidefinite (nsd) (a.s.)*
$\Rightarrow \overline{D_p m} + \partial_y m_2 + 2 (m_2 - diag\{m\})$ *is nsd (a.s.),*
where $\overline{D_p m} = D_p m + D_p m'$.

(iv) *If* S *and* $\mathbb{V}[\partial_y \phi, \phi|Y, Z]$ *are symmetric* $(a.s.)$
$\Rightarrow D_p m + \partial_y m m'$ *is symmetric (a.s.).*

(v) *Let* V *be* Z *measurable*
$\Leftrightarrow \{S$ *is symmetric and nsd iff*
$D_p m + \partial_y m m' + m_2 - diag\{m\}$ *is symmetric and nsd*$\}$.
Moreover, if V *is* Z *measurable, the converse holds in* (i) *and* (ii).

Proof: Appendix.

The importance of this proposition lies in the fact that it allows testing the key elements of rationality without having to specify the functional form of the individual demand function or their distribution in a heterogeneous population. Suppose we see any of these conditions rejected in the observable (generally nonparametric) regression at a position y, z, p. Recalling the interpretation of the conditional expectation as average (over a “neighborhood”) this proposition tells us that there exists a set of positive measure of the population (“some individuals in this neighborhood”) which does not conform with the postulates of rationality. This is the case regardless of how rich our information about heterogeneity is: If our information set is poorly, and we are nevertheless able to identify a local average for which one of the conditions is violated, then it must be a fortiori violated if our information set increases.

If we believe the information to be complete - see case (v) - then we may directly identify these individuals, for then they are completely characterized by their observables. Moreover, the reverse implication is perhaps even more significant. Statements linking the observed model $D_p m + \partial_y m m'$ to individual behavior[7], namely the S, are *only* true if V is Z measurable, i.e. if all individual heterogeneity has been captured by observables. This is a fortiori true for the parametric literature: Appending "an additive error capturing unobserved heterogeneity" and proceeding as usual is not a solution to solve the problem of unobserved heterogeneity. Note that we may always append an additive error, since $m = \phi + (m - \phi) = \phi + \varepsilon$. The crux is now that the error is generally a function of y and p, as was already noted by Brown and Walker (1989). For instance, the potentially nonsymmetric part of the Slutsky matrix becomes

$$S = D_p m + \partial_y m m' + D_p \varepsilon + (\partial_y m)\, \varepsilon' + (\partial_y \varepsilon)\, m' + (\partial_y \varepsilon)\, \varepsilon' \ ,$$

and the last four terms will not vanish in general.

Returning to Proposition 2.2., one should note a key difference between negative semidefiniteness and symmetry. For the former we may provide an "if" characterization without any assumptions other than the basic ones (see (iii)). To obtain a similar result for symmetry, we have to invoke an additional assumption about the conditional covariance matrix. This matrix is unobservable - at least without any further identifying assumptions. Note that this assumption is (implicitly) implied in all of the demand literature, since symmetry is inherited by $D_p m + \partial_y m m'$ only under this assumption.

Conversely, if this additional assumptions does not hold, we are at most able to test the first three elements of rationality. It is well known that homogeneity, adding up and Slutsky negative semidefiniteness alone amount to demand behavior generated by complete, but not necessarily transitive preferences. Details of this "demand theory of the weak axiom" can be found in Kihlstrom, Mas-Colell and Sonnenschein (1976), Kim and Richter

[7]For instance: "All individuals display a negative semidefinite Slutsky matrix, as is evident from the empirical results".

(1986) or Quah (2000). Note further some parallels with the aggregation literature in economic theory: Only adding up and homogeneity carry immediately through to the mean regression. This result is similar in spirit to the Mantel-Sonnenschein theorem, where only these two properties are inherited by aggregate demand. Furthermore, it is also well known in this literature that the aggregation of negative semidefiniteness (usually shown for the Weak Axiom) is more straightforward than that of symmetry. Also, a matrix similar to $\mathbb{V}\left[\partial_y \phi, \phi\right]$ has been used in this literature (as "increasing dispersion", see Jerison (1984)).

As a final note, the assumption that $\mathbb{V}\left[\partial_y \phi, \phi|Y, Z\right]$ be symmetric may be relaxed in the control function spirit, see again Hoderlein (2004) for details.

3 Endogeneity and Exclusion Restrictions

At the center of our argumentation in the second section stood the conditional independence assumption $F_{A|Y,Z} = F_{A|Z}$. This assumption may well be violated, in particular if our information set is small, as argued above. In traditional econometrics, the breakdown of $A2.4$ is called an endogeneity. Possible solutions to this problem arise if we have additional variables, which were excluded from the original model. Throughout this section, we assume to possess only an additional random scalar , denoted by X. Recall in this respect that economic choice variables had to be excluded from Z, and remain therefore natural candidates. In particular, like in the traditional demand system literature we may use current labor income as an instrument for total expenditure. In a world of rational, but potentially heterogeneous agents, labor income is the result of maximizing behavior by individual agents (consumers and firms). Much as in the second section, we can model $X = \upsilon(s, Z, A_2)$,where υ is a fixed Borel-measurable scalar valued function defined on the set $\mathcal{S} \times \mathcal{Z} \times \mathcal{A}_2$. Here, s denotes macroeconomic variables like wage rates and interest rates, which are again for simplicity assumed to be nonrandom. A_2 is a set of unobservables, which may contain unobserved characteristics of the production sector, but more importantly contain unobservables that govern the decision of the individual's intertemporal optimal labor supply problem. Examples include the attitude towards risk, but also elements in the possibly idiosyncratic information sets.

With this additional variable, two possible solutions may be devised in this framework. They differ in the respective independence assumption, and each of them will occupy a subsection. The two solutions can be seen as generalizations of the concepts of instruments and proxies. Each of these two solutions will admit a number of subcases, depending on which kind of variables are employed for conditioning.

3.1 Conditioning on Instruments instead of Endogeneous Regressors

As in section 2, the structural model is given by assumptions $A2.1$ and $A2.2$. However, additionally we make the following

Assumption 3.1: *Let there exist a random scalar X, such that*

$$Y = \mu(X, Z) + \sigma(X, Z)U,$$

where μ and σ are fixed Borel-measurable scalar valued functions defined on the set $\mathcal{X} \times \mathcal{Z}$ of possible values of (X, Z) with $\mathbb{E}[U|X, Z] = 0$ and $\mathbb{V}[U|X, Z] = 1$.

Hence, μ and σ are the nonparametric mean regression and scedastic function respectively. Neglecting marginal modifications of the boundedness condition $A2.3$, which will be denoted as $A2.3'$, the second material modification concerns the dependence structure. Instead of $A2.4$ we assume

Assumption 3.2 $F_{A|X,U,Z} = F_{A|U,Z}$.

Note that this implies $F_{A|Y,U,Z} = F_{A|U,Z}$. The question we have to answer now is whether $A3.2$ plausible for the variables as defined above. In particular, is labor income defined as $X = \upsilon(s, Z, A_2)$ likely to fulfill $A3.2$?

Take first the extreme case that A and A_2 are independent conditional on Z, and we require not only that $\mathbb{E}[U|X, Z] = 0$ and $\mathbb{V}[U|X, Z] = 1$, but in addition also that U is not a function of X, i.e. $U = \varphi(A, Z)$. This property holds, if U is conditionally normal. Then, assumption 3.2 is quite plausible as in this case the condition $F_{A|X,U,Z} = F_{A|U,Z}$ is equivalent to

$F_{A|\upsilon(A_2,Z),\varphi(A,Z),Z} = F_{A|\varphi(A,Z),Z}$, and can be derived as follows:

$$\begin{aligned} F_{A|\upsilon(A_2,Z),\varphi(A,Z),Z} &= \frac{F_{A,\upsilon(A_2,Z),\varphi(A,Z)|Z}}{F_{\upsilon(A_2,Z),\varphi(A,Z)|Z}} \\ &= \frac{F_{A,\varphi(A,Z)|\upsilon(A_2,Z),Z}F_{\upsilon(A_2,Z)|Z}}{F_{\varphi(A,Z)|Z}F_{\upsilon(A_2,Z)|Z}} \\ &= \frac{F_{A,\varphi(A,Z)|Z}}{F_{\varphi(A,Z)|Z}} \\ &= F_{A|\varphi(A,Z),Z} \end{aligned}$$

where only the conditional independence between A and A_2 has been used[8].

At the other extreme, assume that $A = A_2$, i.e. the unobservable characteristics of the household that govern the two decisions are exactly the same. Then we have that both X and U are functions of A, although almost independent ones. Nevertheless, $F_{A|X,U,Z} = F_{A|U,Z}$ is still not completely impossible as U already reflects some of the influence of A. Hence, even in this case $A3.2$ is still more likely to be fulfilled than $A2.4$.

The reality is of course a mixture of both: For instance, labor income may depend on the attitude towards risk that does not affect the preferences for apples vs. bananas. Conversely, conservatism may be reflected in both work choice and goods chosen. As argued, the case for $A3.2$ gets stronger, the more the first scenario dominates.

Along similar lines as in the second section, we explore now the implications of this new concept for the derivatives before turning to economic structures. As mentioned above, there are three possibilities of conditioning, namely $\sigma\{\mu(X), Z\}$, $\sigma\{X, Z\}$ and $\sigma\{Y, Z, U\}$. Note that if $\mu(\cdot)$ is bijective then $\sigma\{\mu(X), Z\} \subseteq \sigma\{Y, Z, U\}$ and $\sigma\{X, Z\} \subseteq \sigma\{Y, Z, U\}$ holds, but not in general (in particular not in the multivariate X case). In what follows, we concentrate on $\sigma\{X, Z\}$ and $\sigma\{Y, Z, U\}$. The latter case will turn out to be the most convenient. Since it is a obvious generalization, we will call it the

[8] To see the third equality, consider

$$\begin{aligned} & \mathbb{P}[A \in I_1, \varphi(A,Z) \in I_2 | \upsilon(A_2,Z) \in I_3, Z \in I_4] \\ = & \mathbb{P}\left[A \in \left\{I_1 \cap \varphi^{-1}(Z, I_2)\right\} | \upsilon(A_2,Z) \in I_3, Z \in I_4\right] \\ = & \mathbb{P}\left[A \in \left\{I_1 \cap \varphi^{-1}(Z, I_2)\right\} | Z \in I_4\right], \end{aligned}$$

where φ^{-1} is the partial inverse of φ with respect to A. This is true for any Borel sets I_1, I_2, I_3 and I_4.

control function approach, and treat it first.

Proposition 3.1 *Let all the variables and functions be as defined above. Suppose* (P), $(A2.2), (A2.3)', (A3.1)$ *and* $(A3.2)$ *hold. Then,* $\mathbb{E}[\partial_y \phi|Y, Z, U] = \partial_y \mathbb{E}[\phi|Y, Z, U]$.

Proof: By similar arguments as $P2.1$.

In addition, all of the results of the second section continue to hold, simply conditioning on the augmented set of random variables. Moreover, in this scenario we may come up with a test for exogeneity. The following lemma is central to this:

Lemma 3.2 *Let all the variables and functions be as defined above. Suppose* (P), $(A2.2)$, $(A2.3)', (A3.1)$ *and* $(A3.2)$ hold. *Then follows that*

$$\partial_y \mathbb{E}[\phi|Y, Z] = \mathbb{E}[\partial_y \phi|Y, Z] + \mathbb{E}[\mathbb{E}[\phi|Y, Z, U]\partial_y \log f_{U|Y,Z}|Y, Z].$$

Proof: Appendix.

Recall from $P2.1$ that under exogeneity $\partial_y \mathbb{E}[\phi|Y, Z] = \mathbb{E}[\partial_y \phi|Y, Z]$. Hence, if we assume that $F_{A|Y,U,Z} = F_{A|U,Z}$ holds in any case, we may base a test for endogeneity on, for instance, whether $\mathbb{E}\left[\mathbb{E}[\mathbb{E}[\phi|Y, Z, U]\partial_y \log f_{U|Y,Z}|Y, Z]^2\right]$ or $\sup_{x,z} \left|\mathbb{E}[\mathbb{E}[\phi|Y, Z, U]\partial_y \log f_{U|Y,Z}|Y = y, Z = z]\right|$ is bigger than zero, or, in the local spirit of this paper, whether $\left|\mathbb{E}[\mathbb{E}[\phi|Y, Z, U]\partial_y \log f_{U|Y,Z}|Y = y, Z = z]\right|$ is bigger than zero for any y, z.

As mentioned, if $\sigma\{X, Z\} \nsubseteq \sigma\{Y, Z, U\}$ the conditional independence assumption generates an additional set of implications[9]. Again we start with the relationship among the derivatives. Let $m(p, x, z)$ and $m_2(p, x, z)$ denote $\mathbb{E}[W|X = x, Z = z]$ and $\mathbb{E}[WW'|X = x, Z = z]$. Then,

Proposition 3.3 *Let all the variables and functions be as defined above.*

[9] If $\sigma\{X, Z\} \subset \sigma\{Y, Z, U\}$ holds, then $\mathbb{E}[\partial_y \phi|X, Z] = \mathbb{E}\left[\partial_y \mathbb{E}[\phi|Y, Z, U]|X, Z\right]$, by iterated expectations. Hence, no really new implications arise, as these are only nonparametric regressions on $\sigma\{X, Z\}$.

Suppose $(P),(A2.2)$, $(A2.3)'$, $(A3.1)$ *and* $(A3.2)$ *hold Then follows that*

$$\begin{aligned}
\mathbb{E}[\partial_y\phi|X,Z] &= \partial_x m(X,Z)\partial_x\mu(X,Z)^{-1} \\
&\quad -\mathbb{E}\left\{\xi(X,Z,U)|X,Z\right\}\partial_x\sigma(X,Z)\partial_x\mu(X,Z)^{-1} \\
&\quad -\mathbb{E}\left\{W\partial_x\log f_{U|X,Z}(U,X,Z)|X,Z\right\}\partial_x\mu(X,Z)^{-1} \ (a.s).,\\
\textit{where } \xi(X,Z,U) &\equiv \partial_x\mathbb{E}[W|X,Z,U]\left[\partial_x\mu(X,Z)+\partial_x\sigma(X,Z)U\right]^{-1}U.
\end{aligned}$$

Proof: Appendix.

This proposition provides again a relationship between the average theoretical derivative and observable quantities, albeit a more complicated one. The leading term can be seen as the most important element in this formula, since it contains the empirical derivative with respect to the instrument, $\partial_x m$. Note here in particular the weighting by $\partial_x\mu(x,z)=\partial_x\mathbb{E}[Y|X=x,Z=z]$. If the instrument is weak, then $\partial_x\mu(x,z)$ will be very small. Hence, in any application, where we replace population quantities by sample analogs, any of the terms on the rhs will be very imprecise. The second and third term present correction expressions: The second is a correction for heteroscedasticity in the regression of Y on X and Z. The last captures higher order effects and vanishes if U and X are conditionally independent.

Let us now focus on the economic implications, in particular on the Slutsky properties. The following proposition summarizes the results:

Proposition 3.4 *Let all the variables and functions be as defined above, and let* $(P),(A2.2)$, $(A2.3)'$, $(A3.1)$ *and* $(A3.2)$ *hold.*

(i) If ϕ *fulfills* $\phi(p+\lambda,Y+\lambda,V)=\phi(p,Y,V)$ *(a.s.)* $\Rightarrow$

$$D_p m\iota + \left[\partial_x m - \mathbb{V}\left\{\xi,U|X,Z\right\}\partial_x\sigma - \mathbb{E}\left\{W\partial_x\log f_{U|X,Z}|X,Z\right\}\right]\partial_x\mu^{-1} = 0 \ (a.s.).$$

(ii) If S *is negative semidefinite (nsd) (a.s.)* $\Rightarrow$

$$\begin{aligned}
&\overline{D_p m} + \partial_x m_2\partial_x\mu^{-1} - \mathbb{E}\left\{\pi(X,Z,U)|X,Z\right\}\partial_x\sigma\partial_x\mu^{-1}\\
&-\mathbb{E}[WW'\partial_x\log f_{U|X}|X,Z]\partial_x\mu^{-1} + 2\left(m_2 - diag\left\{m\right\}\right)
\end{aligned}$$

is nsd (a.s.), where $\overline{D_p m} = D_p m + D_p m'$
and $\pi(X,Z,U)=\partial_x\mathbb{E}[WW'|X,Z,U]\left\{\partial_x\mu(X,Z)+\partial_x\sigma(X,Z)U\right\}^{-1}U$

(iii) If S and $\mathbb{V}[\partial_y \phi, \phi|X,Z]$ are symmetric (a.s.) $\Rightarrow$

$$D_p m + \left[\partial_x m - \mathbb{V}\{\xi, U|X,Z\}\partial_x \sigma - \mathbb{E}\left\{W \partial_x \log f_{U|X,Z}|X,Z\right\}\right]\partial_x \mu^{-1} m',$$

is symmetric (a.s.). Here, all variables are as defined in Proposition 3.3.

Proof: Appendix.

The essence of this proposition is that the results of the second section continue to hold, with derivatives with respect to labor income replacing derivatives with respect to (endogenous) total expenditure. All remarks after $P2.3$ remain in place, in particular it is still possible to test for negative semidefiniteness without further assumptions.

Are there reasons to prefer one projection over the other? As mentioned, if μ is bijective, $\sigma\{X,Z\} \subseteq \sigma\{Y,Z,U\}$. Hence, we should use the larger sigma algebra, as further averaging only deprives us of possibilities to test. If $\sigma\{X,Z\} \nsubseteq \sigma\{Y,Z,U\}$, the problem of weak instruments may be crucial for the first, as $\partial_x \mu$ may be close to zero. However, it is also problematic for the second, as then - in terms of the influence of A - U and Y may be quite similar, meaning that in any application some type of comovement (”collinearity”) may make estimation problematic[10]. The upshot is that in the $\sigma\{X,Z\} \nsubseteq \sigma\{Y,Z,U\}$ scenario we should use projections in all spaces, not just in one.

3.2 Conditioning on Endogenous Regressors and Excluded Variables - a ”Proxy” Solution

Thus far exploited the exclusion restriction with a method that worked particularly well if, loosely speaking, the endogenous variable and the excluded variable have little in common (at least conditionally). In this section we will consider a method that works better for the reverse case. This second scenario employs a different independence assumption, namely

Assumption 3.3 $F_{A|Y,X,Z} = F_{A|X,Z}$.

If Y and X are generated by the same preference parameters, then this assumption may likely hold. This is the case, because then, loosely speaking,

[10]Perhaps the strongest argument in favor of $\sigma\{X,Z\}$ is tradition.

the additional variable already contains a lot of the influence of A. However, the similarity in this respect may have a negative effects on the estimation, much as in the above "control function IV" approach. It is almost needless to mention that this assumption is likely to break down if A and A_2 are independent or nearly so. The consequences for testing economic theory are straightforward: Most of the statements of section 2 continue to hold, with the extended σ-algebra $\sigma\{Y, X, Z\}$ in place of $\sigma\{Y, Z\}$.

A projection that has not been used in the literature would involve $\sigma\{U, Z\}$. However, this is under our assumption always a subalgebra of $\sigma\{Y, X, Z\}$. Hence, no new testable implications may be derived.

4 Implications for Welfare Analysis

to be completed.

5 Econometric Specification

In this section we examine the asymptotic properties of the empirical nonparametric demand systems, which is going to involve a system of local polynomial regression, and the associated test statistics. More importantly, we propose local test statistics for the symmetry and the homogeneity hypothesis. Recall that - as a consequence of our model - it are the local properties of the empirical demand function that had an interpretation as conditional averages. Nevertheless, global statistics may be of interest as well since one aim of the analysis might be to determine whether homogeneity is rejected for the population as whole. Global tests are also going to be implemented in the empirical part. However, their econometric theory involves functionals and is very different. It is being derived in a companion paper (Hoderlein, Haag and Pendakur (2004)). Finally, for testing negative semidefiniteness a bootstrap test will be applied, to which no theory will be presented in this paper.

In the following subsection, we will analyze the exogenous, as well as the endogenous scenario with IV. To treat these scenarios under the same format, we consider the regression on W on P, Y and Z, where, Y may now denote either total expenditure (under exogeneity), which was already denoted as Y, or labor income, which was previously denoted as X under exogeneity. Z in turn remains unchanged. However, recall that the control function approach involved the additional regressor U.

In principle, this means that our set Z of nuisance regressors increases by one. But this extension is far from trivial, as this regressor has to be replaced by a pre-estimated version, which involves, for instance, a Nadaraya Watson pre-estimator $\hat{\mu}$ for μ. We will treat the issue of a pre-estimated regressor in general fashion in the appendix.

5.1 Local Linear Systems of Equations

Again we model the dependence of the shares of total expenditure commanded by each of L goods on the log-prices of all goods and on the log of total household expenditure. Denote by $P_i = [P_i^1, ..., P_i^L]'$ the log-price vector, by Y_i log-expenditure, by $Z_i = [Z_i^1, ..., Z_i^K]'$ the vector of individual characteristics and $W_i = [W_i^1, ..., W_i^L]'$ the expenditure share vector of the i'th expenditure unit - individuals in this application - where $i = 1, ..., N$. Let $X_i = [P_i', Y_i, Z_i']'$and $d = K + L + 1$. This notation is not be confounded with the notation X to denote instruments in the third section . It is introduced to handle the exogenous as well as the endogenous case under the same format: In the exogenous case X_i is simply the vector of all regressors, in the endogenous case it is the vector of exogenous regressors and instruments. In what follows, i indexes individuals and j indexes goods (equations). For most parts of this paper we shall assume to have an iid sample $\{(W_i, X_i)\}_{i=1}^n$, such that $(W_i, X_i) \sim (W, X)$, but see the note in the appendix for extensions to models that fulfill a mixing condition.

The task of estimating an empirical object, which has some economically interpretable structure, involves now in both scenarios the model

$$W_i = m(X_i) + \Sigma_1(X_i)\eta_i,$$

where $m(\cdot)$ and $\Sigma_1(\cdot)$ are now assumed to be Borel-measurable, smooth, $L-1$ vector valued and $L-1 \times L-1$ matrix valued functions, respectively. Moreover, they admit a second order Taylor expansion[11]. We assume that $m(x_0) = \mathbb{E}(W_i|X_i = x_0)$ at any fixed vector $x_0 = [p_0^1, ..., p_0^L, y_0, z_0^1, ..., z_0^K]'$, $\Sigma_1(x_0)\Sigma_1(x_0)' = Var(W_i|X_i = x_0)$ with finite elements for any x_0, and that the η_i are mutually independent and identically distributed zero mean random vectors with an identity covariance matrix, independent of the X_i's. A

[11]We impose the adding up constraint that expenditure shares add up to 1.

complete list of all assumptions involved, including regularity conditions, can be found in the appendix. Since W is assumed to be bounded, all higher moments - and thus also conditional moments - exist, making this assumptions unrestrictive.

Now, in any of the two scenarios, we consider additional dependent variables, denoted as R_{ni}. Obviously, we should not accept the exogeneity assumption without having a tool to test for this property within an encompassing set of assumptions. This tool was provided by Lemma 3.2, where the regression $\mathbb{E}\left[\mathbb{E}\left[W_i|Y_i, Z_i, U_i\right]\partial_y \log f_{U|Y,Z}(U_i; Y_i, Z_i)|Y_i, Z_i\right]$ played a central role. Therefore we have $R_{ni} = \widehat{\mathbb{E}}\left[W_i|Y_i, Z_i, U_i\right]\partial_y \log \hat{f}_{U|Y,Z}(U_i; Y_i, Z_i)$, which is again a $L-1$ vector, and involves pre-estimated quantities, $\widehat{\mathbb{E}}\left[W_i|Y_i, Z_i, U_i\right]$ and $\hat{f}_{U|Y,Z}$ (hence the subscript n). However, in the remainder of this paper we focus on the endogenous scenario with instruments. Here R_{ni} consists of two components: First we have the endogenous regressor (in demand: total expenditure denoted by D_i).on instruments regression, i.e. $D_i = \mu(X_i, Z_i) + \sigma_2(X_i, Z_i)\varepsilon_i$. Here, ε_i is a mean zero, unit variance random scalar, jointly (with η_i) independent of X_i, Z_i. For compactness of notation, we denote μ as m_L, i.e. the L-th element of m. Second, recall from Lemma 3.1 that we need to estimate both the levels and derivatives of the functions μ and σ_2, and to determine the joint distribution of estimators for m, μ and σ_2 and derivatives thereof. Hence, R_{ni} contains total expenditure as well as the estimated squared residuals from the regression of total expenditure on instruments, formally $U_{ni} = D_i - \hat{m}_L(X_i)$, where $\hat{m}_L$ denotes a Nadaraya Watson pre-estimator for the conditional expectation.

Let $S_i = [W_i', D_i, U_{ni}^2]' = [W_i', R_{ni}']'$, denote the vector of dependent variables, which is, as a consequence, of length $\overline{L} = L + 1$. The estimation of the function m their derivatives at a particular point x_0 as well as the conditional scedastic function σ_2 is based on local polynomial modelling. Although the asymptotics of local polynomial estimators (LPEs, for short) for single equation models have been derived, among others, by Härdle and Tsybakov (1997), we give a full proof of asymptotic normality for the estimators in our model. This is done for the following reasons: First, we extend the result to systems of equations, analogous to SURE in the linear case. Second and more importantly, we derive the joint distribution of estimators of the mean regression m and the scedastic function σ_2, as well as their derivatives. In particular, on our way we have to tackle problems involving the uniform convergence of the Nadaraya Watson pre-estimator. Third, we will treat the problem of pre-estimation of some of the regressors, as is for instance needed when the residuals of the regression of endogenous regressors on instruments are employed. And finally, since the focus in this paper is on pointwise testing, we need these results in a slightly different formulation than usually considered, in order to emphasize the differences in speed of convergence.

To this end, consider the multivariate local linear (LL) model where we solve the following weighted least squares minimization problem

$$\min_{\alpha^j(x_0),\beta_l^j(x_0),\beta_y^j(x_0),\beta_l^j(x_0)} n^{-1}\sum_{i=1}^{n} K_{H_n}\left(X_i - x_0\right)\xi_i'\xi_i$$

where for all $j = 1,..,\overline{L}$,

$$\xi_i^j = S_i^j - \alpha^j(x_0) - \sum_{l=1}^{L}\beta_l^j(x_0)\left(P_i^l - p^l\right) - \beta_y^j(x_0)\left(Y_i - y\right) - \sum_{k=1}^{K}\beta_k^j(x_0)\left(Z_i^k - z^k\right),$$

$\xi_i = [\xi_i^1, ..., \xi_i^{\overline{L}}]'$, K is an L-variate kernel such that $\int K(\psi)d\psi = 1$, $K_{H_n}(\psi) = |H_n|^{-1/2}K(H_n^{-1/2}\psi)$ and H_n is an $L\times L$ symmetric positive definite bandwidth matrix depending on n. Here, $K_{H_n}\left(X_i - z\right)$ is a weight which penalizes the distance of the observation X_i from x_0 so that observations near x_0 get more weight than those distant from x_0. The kernel function depends on the bandwidth matrix H_n which puts scale on the distances of the various components the independent variable vector X_i. For simplicity of exposition, we shall use a product Kernel and a diagonal bandwidth matrix, with $H_n = h_n^2 I_L$. Moreover, to keep track of the difference in speed of convergence, we consider

$$\min_{\theta(z)} n^{-1}\sum_{i=1}^{n} K_{H_n}\left(X_i - x_0\right)\xi_i'\xi_i, \tag{5.1}$$

where, for all $j = 1,..,\overline{L}$,

$$\xi_i^j = S_i^j - \alpha^j(x_0) - \sum_{l=1}^{L} h\beta_l^j(x_0)\frac{P_i^l - p^l}{h} - h\beta_y^j(x_0)\frac{Y_i - y}{h} - \sum_{k=1}^{K} h\beta_k^j(x_0)\left(\frac{Z_i^k - z^k}{h}\right),$$

and we denote the list of all parameters as $\theta^j(x_0) = \{\alpha^j(x_0), h\beta_l^j(x_0), h\beta_y^j(x_0), h\beta_k^j(x_0),\}$, $j = 1,..,\overline{L}$, $l = 1,..,L$, $k = 1,..,K$, and denote the parameters which minimize (5.1) as $\hat{\theta}^j(x_0) = \{\widehat{\alpha}^j(x_0),\ \widehat{h\beta}_l^j(x_0),\ \widehat{h\beta}_y^j(x_0),\ \widehat{h\beta}_k^j(x_0),\}$, $j = 1,..,\overline{L}$,

$l = 1,..,L$, $k = 1,..,K$. Finally, let $\theta(x_0) = \{\theta^1(x_0), ..., \theta^{\overline{L}}(x_0))'$ and $\hat{\theta}(x_0) = (\hat{\theta}^1(x_0), ..., \hat{\theta}^{\overline{L}}(x_0))'$.

These are our parameters of interest. As shall become clear from the proof in the appendix, they have the following nice properties: Estimators for the levels as well as derivatives of the budget share regression as well as the instrument regression are, for $j = 1,..,L : \widehat{m}^j(x_0) = \widehat{\alpha}^j(x_0)$, $\widehat{\partial m^j}(x_0)/\partial p^l = \widehat{\beta}_l^j(x_0)$, $l = 1,..,L$, $\widehat{\partial m^j}(x_0)/\partial y = \widehat{\beta}_y^j(x_0)$, $\widehat{\partial m^j}(x_0)/\partial z^k = \widehat{\beta}_k^j(x_0)$, $k = 1,..,K$. In turn, estimators for the covariance are given by $\widehat{\sigma_2^2}(x_0) = \widehat{\alpha}^{\overline{L}}(x_0)$, and for their derivatives by $\partial\widehat{\sigma_2^2}(x_0)/\partial p^l = \widehat{\beta}_l^{\overline{L}}(x_0)$, $l = 1,..,L$, $\partial\widehat{\sigma_2^2}(x_0)/\partial y = \widehat{\beta}_y^{\overline{L}}(x_0)$, and $\partial\widehat{\sigma_2^2}(x_0)/\partial z^k = \widehat{\beta}_k^{\overline{L}}(x_0)$, $k = 1,..,K$.

It is a comparably trivial exercise to the show that - analogously to the SUR literature - the parameters $\theta(x_0)$ can be estimated by locally weighted equation-by-equation OLS of expenditure shares, the endogenous regressors, and the squared empirical residuals on a constant, log-prices divided by h, log-expenditure divided by h and individual specific characteristics divided by h, if the regressors are the same across equations. The formal result, which is established in the appendix, is as follows:

Propostion 5.1: *Let the model be as defined above, and let A1-A8 given in the Appendix hold. Then follows that*

$$\sqrt{nh^d}\left(\hat{\theta}(x_0) - \theta(x_0) - h^2 bias(x_0)\right) \xrightarrow{d} \mathcal{N}(0, \Xi(x_0) \otimes A),$$

where d denotes the number of regressors excluding the constant, A is a fixed $(d+1)\times(d+1)$ matrix given by $A = f_X(x_0)^{-1}B^{-1}CB^{-1}$, $f_X(x_0) > 0$ denotes the joint distribution of all regressors, and the fixed matrices B and *C are defined as*

$$\underset{(d+1)\times(d+1)}{B} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & \mu_2 & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \mu_2 \end{bmatrix} \quad \text{and} \quad \underset{(d+1)\times(d+1)}{C} = \begin{bmatrix} \kappa_0^{d+1} & \kappa_0^d\kappa_1 & \cdots & \kappa_0^d\kappa_1 \\ \kappa_0^d\kappa_1 & \kappa_0^d\kappa_2 & \cdots & \kappa_0^d\kappa_2 \\ \vdots & \vdots & \ddots & \vdots \\ \kappa_0^d\kappa_1 & \kappa_0^d\kappa_2 & \cdots & \kappa_0^d\kappa_2 \end{bmatrix},$$

where $\mu_2 = \int \psi^2 K(\psi)d\psi$, and $\kappa_l = \int \psi^l K^2(\psi)d\psi$, $l = 0, 1, 2$. Finally

$$\Xi(x_0) = \begin{bmatrix} \Sigma_1(x_0)\Sigma_1(x_0)' & \sigma_2(x_0)\Sigma_1(x_0)\mu_{\eta\varepsilon} & \sigma_2^2(x_0)\Sigma_1(x_0)\mu_{\eta\varepsilon^2} \\ \sigma_2(x_0)\mu'_{\eta\varepsilon}\Sigma_1(x_0)' & \sigma_2^2(x_0) & \mu_{\varepsilon^3}\sigma_2^3(x_0) \\ \sigma_2^2(x_0)\mu'_{\eta\varepsilon^2}\Sigma_1(x_0)' & \mu_{\varepsilon^3}\sigma_2^3(x_0) & \mu_{\varepsilon^4}\sigma_2^4(x_0) \end{bmatrix},$$

with $\Sigma(x_0)$ *and* $\sigma_2(x_0)$ *as defined above, and* $\mu_{\eta^k \varepsilon^l} = \mathbb{E}\left[\eta_i^k \varepsilon_i^l\right]$, $k = 0, 1$, $l = 1, 2, 3$.

Proof: Appendix.

Remarks: 1. As is well known, but crucial for our analysis, the speed of convergence is not the same for the parameters of interest. Only the premultiplied derivative coefficients converge at the same speed as the $\hat{\alpha}^j(z)$ $j = 1, .., \overline{L}$. As a consequence, some test statistic may have perhaps not immediately expected properties. As an example consider the nonlinear test for symmetry under exogeneity. It will turn out that this test simplifies dramatically in this scenario.

2. The proof given in appendix $A1$ extends immediately to the case of all random variables being α-mixing stochastic processes. This class covers most commonly used stationary stochastic processes, e.g. of the AR(p)-type. Similarly, but not formally shown, a deterministic time trend may be included. However, like both the parametric demand system literature (exception: Lewbel and Ng (2003)) and the nonparametric literature, with nonparametric methods we can, as of yet, not handle nonstationary regressors.

3. If the pre-estimated residuals are included as additional regressors, we need for this result to hold in addition that μ be four times differentiable and that we employ a fourth order Kernel. Details are available from the author upon request.

5.2 Local Homogeneity

Recall the testable implications of the assumptions that homogeneity holds in a heterogeneous population, as given in $P2.3$ and $P3.3$. Under exogeneity, as stated in $P2.3$ (ii),

$$D_p m \iota + \partial_y m = 0 \ (a.s.),$$

while under endogeneity, $P3.3$ (i) stipulates that

$$D_p m \iota + \left[\partial_y m - \mathbb{E}\left\{\xi | Y, Z\right\} \partial_y \sigma - \mathbb{E}\left\{W \partial_y \log f_{U|Y,Z} | Y, Z\right\}\right] \partial_y \mu^{-1} = 0 \ (a.s.),$$

using the change notation from x to y compared to section 3. The first hypothesis is straightforwardly testable given $P5.1$. However, we will treat the second hypothesis only in a simplified version, in give some indications below how the test of the not simplified hypothesis may behave asymptotically. First, in the estimation part we already assumed that U is conditionally independent of X_i. Hence, $\mathbb{E}\left\{W\partial_x \log f_{U|X,Z}|X,Z\right\} = 0$, by assumption. Hence $D_p m\iota + \partial_x m \partial_x \mu^{-1} - \mathbb{E}\{\xi|X,Z\}\partial_x\sigma\partial_x\mu^{-1} = 0$. As we shall se below, the asymptotic behavior of multiplicative statistics is determined by the speed of convergence of the slowest part. Hence, the asymptotic distribution of an estimator for $\mathbb{E}\{\xi|X,Z\}$ does not matter. As a consequence, the fact that ξ has to be pre-estimated is also irrelevant. This is because even if it were to impact the asymptotic distribution of an estimator of $\mathbb{E}\{\xi|X,Z\}$ this would not matter for the behavior of the test statistic (unless it does not cause divergence).

From the asymptotic theory of the preceding subsection, the exogenous case is easily covered. Let R denote the $L-1 \times d(L-1)$ matrix, $R = I_{L-1} \otimes \begin{bmatrix} 0 & \iota_{L+1} & \mathbf{0}_K \end{bmatrix}$, and consider the test statistic

$$\hat{\tau} = \left[R\left[\hat{\theta} - h^2\widehat{bias}\right]\right]' \left[R'\left[\widehat{\Sigma_1\Sigma_1'}(x) \otimes \hat{A}\right]R\right]^{-1} R\left[\hat{\theta} - h^2\widehat{bias}\right],$$

where $\hat{A} = \hat{f}_X(x)^{-1}B^{-1}CB^{-1}$ and $\hat{\sigma}^2_{kl}(x)$, $k,l = 1,..,L-1$ is the k,l-th element of $\widehat{\Sigma_1\Sigma_1'}(x)$, given by $\hat{\sigma}^2_{kl}(x) = \sum_i K\left(\frac{X_i - x}{h}\right)\hat{U}_{ij}\hat{U}_{il}/\sum_i K\left(\frac{X_i-x}{h}\right)$, where $\hat{U}_{ij}$ are the fitted residuals in the j-th regression. Moreover, $\widehat{bias}$ is a pre-estimator for the bias. Since the bias contains largely second derivatives, we may use a local quadratic or cubic estimator for the second derivative, with a substantial amount of undersmoothing.

Then, by a trivial corollary to proposition 5.1, $\hat{\tau} \xrightarrow{d} \chi^2_{L-1}$.

The issue becomes more involved, if we want to test for homogeneity in the exogenous case. First, note that $\partial_y\sigma_2 = \partial_y\sigma_2^2\left[2\left(\sigma_2^2\right)^{-1/2}\right]^{-1}$. Then, $G(\psi) = (G_1(\psi), .., G_{L-1}(\psi))' = 0$, where for all $j = 1,..,L-1$,
$G_j(\psi) = \sum_l \beta^j_l(x_0) + \beta^j_y(x_0)\beta^L_y(x_0)^{-1} + 0.5\alpha^{\xi,j}(x_0)\beta^{\overline{L}}_y(x_0)\beta^L_y(x_0)^{-1}\alpha^{\overline{L}}(x_0)^{-1/2}$, where $\alpha^{\xi,j}(x_0)$ is the j-th element of $\alpha^{\xi}(x_0) = \mathbb{E}\{\xi|X = x_0\}$. This leads to the test statistic,

$$\tilde{\tau} = \left[G\left(\hat{\psi}\right)\right]'\left[D_\psi G\left(\hat{\psi}\right)'\Sigma(\hat{\psi})D_\psi G\left(\hat{\psi}\right)\right]^{-1} G\left(\hat{\psi}\right),$$

where $\hat{\psi} = (\hat{\theta}', \hat{\alpha}^{\xi\prime})'$, $\hat{\alpha}^{\xi}$ is any $\sqrt{nh^d}$ consistent estimator of $\alpha^{\xi}(x_0)$, and

$\Sigma(\hat{\psi})$ is the asymptotic covariance matrix. Then due to

$$\begin{aligned} G_j\left(\hat{\psi}\right) &= \sum_l \hat{\beta}_l^j(x_0) + \hat{\beta}_y^j(x_0)\hat{\beta}_y^L(x_0)^{-1} + \hat{\alpha}^\xi \hat{\beta}_y^{\overline{L}}(x_0)\hat{\beta}_y^L(x_0)^{-1}\hat{\alpha}^{\overline{L}}(x_0)^{-1/2} \\ &= \sum_l \hat{\beta}_l^j(x_0) + \hat{\beta}_y^j(x_0)\hat{\beta}_y^L(x_0)^{-1} + \alpha^\xi \hat{\beta}_y^{\overline{L}}(x_0)\hat{\beta}_y^L(x_0)^{-1}\hat{\alpha}^{\overline{L}}(x_0)^{-1/2} \\ &\quad + \left(\hat{\alpha}^\xi - \alpha^\xi\right)\hat{\beta}_y^{\overline{L}}(x_0)\hat{\beta}_y^L(x_0)^{-1}\hat{\alpha}^{\overline{L}}(x_0)^{-1/2} \\ &= \sum_l \frac{\hat{\theta}_l^j(x_0)}{h} + \hat{\theta}_y^j(x_0)\hat{\theta}_y^L(x_0)^{-1} + \alpha^\xi \hat{\theta}_y^{\overline{L}}(x_0)\hat{\theta}_y^L(x_0)^{-1}\hat{\theta}^{\overline{L}}(x_0)^{-1/2} \\ &\quad + \left(\hat{\alpha}^\xi - \alpha^\xi\right)\hat{\theta}_y^{\overline{L}}(x_0)\hat{\theta}_y^L(x_0)^{-1}\hat{\theta}^{\overline{L}}(x_0)^{-1/2}, \end{aligned}$$

we know that $\sqrt{nh^d}G\left(\hat{\psi}\right) = \sqrt{nh^d}G\left(\hat{\theta},\alpha^\xi\right) + o_p\left(\sqrt{nh^{d+4}}\right) = \sqrt{nh^d}G\left(\hat{\theta},\alpha^\xi\right) + o_p(1)$. Moreover, $\frac{1}{h}\sqrt{nh^d}\sum_l \hat{\theta}_l^j(x_0)$ diverges, and hence we have to pre-multiply $G\left(\hat{\psi}\right)$ by h. This implies that this first linear combination asymptotically dominates the test statistic, as it's variance is $O(h^{-2})$ bigger than the variance of the second and third expression. Nevertheless, we should keep the variance of the last two terms, as in any finite sample they can be expected to be of some importance. Employing the same bias correction as above,

$$\sqrt{nh^d}G\left(\tilde{\psi}\right) \xrightarrow{d} \mathcal{N}\left(0, D_\theta G\left(\theta,\alpha^\xi\right)' \left[\Xi(x_0)\otimes A\right] D_\theta G\left(\theta,\alpha^\xi\right)\right),$$

where $\tilde{\psi} = \hat{\psi} - h^2\widehat{bias}$, and

$$\tilde{\tau} = \left[G\left(\tilde{\psi}\right)\right]'\left[D_\theta G\left(\tilde{\psi}\right)'\left[\Xi(x_0)\otimes A\right] D_\theta G\left(\tilde{\psi}\right)\right]^{-1} G\left(\tilde{\psi}\right) \xrightarrow{d} \chi^2_{L-1}.$$

5.3 Local Symmetry

Now the testable implications of symmetry in a heterogeneous population, as given in $P2.3$ and $P3.3$, are being scrutinized. Under exogeneity, we know that the matrix $D_p m + \partial_y m m'$ is almost surely symmetric. while under endogeneity, $P3.3$ (iii) establishes that the matrix

$$D_p m + \left[\partial_y m - \mathbb{E}\left\{\xi | Y, Z\right\}\partial_y \sigma\right]\partial_y \mu^{-1} m',$$

is symmetric under the assumptions of section 5. Using again the change of notation from x to y compared to section 3, we obtain that under exogeneity we have the $1/2L(L-1)$ restrictions

$$\beta_k^l + \beta_y^l \alpha^k - \beta_l^k - \beta_y^k \alpha^l = 0, \quad k, l = 1, .., L-1, k > l.$$

Although this is again a nonlinear test statistic, the fact the we multiply estimators with different speeds of convergence makes life easier. Indeed, $G_{s,x}(\hat{\theta})$ given by this system of restrictions, behaves asymptotically as if we had only a linear restrictions among the derivatives, i.e. the variance of the estimators for α does not enter the test statistic. Here, the subscript s, x stands for "*s*ymmetry under e*x*ogeneity". Hence,

$$\hat{\tau}_{s,x} = \left[G_{s,x}(\tilde{\theta})\right]' \left[D_\beta G_{s,x}(\tilde{\theta})' \left[\Xi(x_0) \otimes A\right] D_\beta G_{s,x}(\tilde{\theta})\right]^{-1} G_{s,x}(\tilde{\theta}) \xrightarrow{d} \chi^2_{1/2L(L-1)},$$

where D_β denotes only the vector having derivatives only with respect to some $\beta' s$ and zeros in place of the derivatives with respect to α, and $\tilde{\theta} = \hat{\theta} - h^2 \widehat{bias}$. Finally, under endogeneity life becomes even simpler asymptotically. By similar remarks as above, it is really only the variance of estimators for the price derivatives β_k^l and β_l^k that matter in each restriction asymptotically. Nevertheless, as in the homogeneity test, we keep the h multiplied ratio of income derivatives. Hence, with the subsript s, d denoting for "*s*ymmetry under en*d*ogeneity"

$$\hat{\tau}_{s,d} = \left[G_{s,d}(\tilde{\theta})\right]' \left[D_\beta G_{s,d}(\tilde{\theta})' \left[\Xi(x_0) \otimes A\right] D_\beta G_{s,d}(\tilde{\theta})\right]^{-1} G_{s,d}(\tilde{\theta}) \xrightarrow{d} \chi^2_{1/2L(L-1)}.$$

This almost summarizes the behavior of the test statistics, whose asymptotic distribution may straightforwardly be derived from $P5.1$. It's implementation, as well as the above mentioned bootstrap based test for Slutsky negative semidefiniteness, is discussed in the following section.

6 Empirical Implementation

to be completed

7 Summary

Unifying the treatment of preference heterogeneity in applied and theoretical work has been a long unresolved issue. This paper tries to fill this gap by introducing a nonparametric framework under which a great many issues can be modelled. It is demonstrated in this paper that very general tests for the various elements of rational demand behavior exist, which may be based upon nonparametric demand systems. Moreover, it is shown that such diverse concepts as welfare and endogeneity of regressors can also be treated under the same formate. Aiming at applications of this approach, we introduced asymptotic theory for system local polynomials including pre-estimated regressors and dependent variables and the test statistic originating from them.

This is by no means an exhaustive list of what a projection based approach to demand analysis may achieve. A great variety of other issues may also be studied: Examples include measurement error and functional form restrictions (Hoderlein, 2004), aggregation and the Law of Demand (Hoderlein and Quah (2004)), or reducing the dimensionality of demand problems and dispensing with the assumption of separability (Hoderlein and Lewbel (2004)). Finally, combined with nonparametric techniques this approach can be applied to real data, as is sketched in the last section and detailed in Hoderlein, Haag and Pendakur (2004), where the asymptotic behavior of global tests of the same hypothesis is also analyzed.

8 Appendix 1

Proof of Proposition 2.1:

Ad $(i), (ii)$ First recall that, by definition, $0 \leq W \leq 1$. Thus, the expectation exists and $\mathbb{E}[|W|] \leq k < \infty$ (the same holds for the second moment). From this follows that all conditional expectations exist as well, and are bounded. Let now p, y, z be fixed, but arbitrary. Then, inserting $A2.1$

$$\partial_y m(p, y, z) = \partial_y \mathbb{E}[W|Y = y, Z = z] = \partial_y \int_{\mathcal{A}} \phi(p, y, \vartheta(z, a)) F_{A|Y,Z}(da, y, z)$$

Under $A2.4$, the rhs equals $\partial_y \int_{\mathcal{A}} \phi(p, y, \vartheta(z, a)) F_{A|Z}(da, z)$, and using the dominated convergence assumption $A2.3$, we obtain

$$\int_{\mathcal{A}} \partial_y \phi(p, y, \vartheta(z, a)) F_{A|Z}(da, z) \tag{A.1}$$

But due to $A2.4$ this is a version of $\mathbb{E}[\partial_y \phi | Y = y, Z = z]$. Upon inserting random variables for the fixed z, y the statement follows. The proof is identical for D_p, save for the fact that we do not need $A2.4$.

To see the "if" part in (iii) of the proposition, simply note that if V is Z-measurable

$$\mathbb{E}[\partial_y \phi(p, Y, \vartheta(Z, A)) | Y = y, Z] = \partial_y \phi(p, y, \theta(Z))$$

for any y and some function θ.
To see the "only if" part, assume that V is not Z-measurable. Then, there exist two sets $\Omega_0, \Omega_1 \subset \Omega$ such that $\mathbb{P}[\Omega_l] > 0$, $l = 0, 1$ and for all $\omega_j \in \Omega_0, \omega_k \in \Omega_1$, $k \neq j$, $Z(\omega_k) = Z(\omega_j)$, $Y(\omega_k) = Y(\omega_j)$, but $V(\omega_k) \neq V(\omega_j)$ since otherwise V would be Y, Z-measurable on $\Omega_0 \cup \Omega_1$. By the identification condition it follows that $\phi(p, Y(\omega_k), V(\omega_k)) \neq \phi(p, Y(\omega_j), V(\omega_j))$, $\partial_y \phi(\omega_k) \neq \partial_y \phi(\omega_j)$ and $D_p \phi(\omega_k) \neq D_p \phi(\omega_j)$, for all $\omega_j \in \Omega_0, \omega_k \in \Omega_1$, $k \neq j$, although
$\mathbb{E}[\partial_y \phi(p, Y, V) | Y, Z, \{\omega \in \Omega_0\}] = \mathbb{E}[\partial_y \phi(p, Y, V) | Y, Z, \{\omega \in \Omega_1\}]$ □

Proof of Corollary 2.2:

By iterated expectations and $P2.1$,

$$\begin{aligned} \mathbb{E}[\partial_y \phi(p, Y, V) | \mathcal{F}] &= \mathbb{E}[\mathbb{E}[\partial_y \phi(p, Y, V) | Y, Z | \mathcal{F}] \\ &= \mathbb{E}[\partial_y \mathbb{E}[\phi(p, Y, V) | Y, Z | \mathcal{F}] \\ &= \mathbb{E}[\partial_y m(p, Y, Z) | \mathcal{F}] \end{aligned}$$

for any $\mathcal{F} \subseteq \sigma\{Y, Z\}$. The same holds of course for the trivial sigma algebra $\{\emptyset, \Omega\}$. □

Proof of Proposition 2.3:

Ad (i) Assume adding up $\iota' \phi = 1$ $(a.s)$. Taking conditional expectations produces
$\iota' m = \mathbb{E}[\iota' \phi | Y, Z] = 1$ $(a.s.)$, by which $\iota' m = 1$ $(a.s.)$ is obvious.

Ad (ii) Assume homogeneity holds across the population, i.e.

$\phi(p+\lambda, y+\lambda, V) = \phi(p, y, V)$ $(a.s.)$ for all p, y. Thus

$$\begin{aligned} m(p,y,z) &= \int\limits_{\mathcal{A}} \phi(p,y,\vartheta(a,z)) F_{A|Y,Z}(da,y,z) \\ &= \int\limits_{\mathcal{A}} \phi(p+\lambda, y+\lambda, \vartheta(a,z)) F_{A|Y,Z}(da,y,z) \end{aligned}$$

But since $F_{A|Y,Z} = F_{A|Z}$, we have that $F_{A|Y,Z}(da,y,z) = F_{A|Z}(da,z) = F_{A|Y,Z}(da,y+\lambda,z)$. Thus,

$$\begin{aligned} & \int\limits_{\mathcal{A}} \phi(p+\lambda, y+\lambda, \vartheta(a,z)) F_{A|Y,Z}(da,y,z) \\ = & \int\limits_{\mathcal{A}} \phi(p+\lambda, y+\lambda, v) F_{A|Y,Z}(da,y+\lambda,z) \\ = & \; m(p+\lambda, y+\lambda, z) \end{aligned}$$

Ad (iii), Note that for any random matrix $A(\omega)$ we have if $p'A(\omega)p \leq 0$ for all ω, it follows that upon taking expectations w.r.t. an arbitrary probability measure F,[12]

$$\int p'A(\omega)pF(d\omega) \leq 0 \Leftrightarrow p' \int A(\omega)F(d\omega)p \leq 0, \text{ for all } p \in \mathbb{R}^L.$$

From this S *nsd* $(a.s.) \Rightarrow \mathbb{E}[S|Y,Z]$ *nsd* $(a.s.)$ is immediate. Let $\mathbb{E}[S|Y,Z] = B$, and note that since the definition of negative semidefiniteness of a square matrix B of dim L involves the quadratic form, $p'Bp \leq 0$, we see that if we put $\bar{B} = B + B'$, we have

$$p'\bar{B}p = p'Bp \text{ for all } p \in \mathbb{R}^L,$$

and $\bar{B}$ symmetric, implying that B is negative semidefinite if and only if $\bar{B}$ is negative semidefinite. From

$$\begin{aligned} B &= \mathbb{E}[S|Y,Z] \\ &= \mathbb{E}[D_p\phi|Y,Z] + \mathbb{E}[\partial_y\phi\phi'|Y,Z] + \mathbb{E}[\phi\phi'|Y,Z] - \mathbb{E}[diag(\phi)|Y,Z] \\ &= B_1 + B_2 + B_3 + B_4 \end{aligned}$$

[12]For conditional probability measures this works similarly in the spaces under consideration.

follows that $\bar{B} = B + B' = B_1 + B_2 + B_3 + B_4 + B_1' + B_2' + B_3' + B_4' = \bar{B}_1 + \bar{B}_2 + 2(B_3 + B_4)$, since B_3 and B_4 are symmetric. Thus we have that

$$S \; nsd \; (a.s.) \Rightarrow \bar{B}_1 + \bar{B}_2 + 2(B_3 + B_4) \; nsd \; (a.s.)$$

From $P2.1$ it is apparent that $\bar{B}_1 = D_p m + D_p m'$. To see that $\bar{B}_2 = \partial_y m_2(p, y, z)$, first note that due to the boundedness of W the second moments and conditional moments exist, so that

$$\partial_y m_2(p, y, z) = \partial_y \mathbb{E}[WW'|Y = y, Z = z] =$$

$$\partial_y \int_{\mathcal{A}} \phi(p, y, \vartheta(z, a))\phi'(p, y, \vartheta(z, a)) F_{A|Y,Z}(da; y, z)$$

Finally, by a modification of $A2.3$, we have

$$\partial_y \int_{\mathcal{A}} \phi(p, y, \vartheta(z, a))\phi'(p, y, \vartheta(z, a)) F_{A|Y,Z}(da; y, z) = \mathbb{E}[\partial_y(\phi\phi')|Y = y, Z = z],$$

but the rhs equals $\mathbb{E}[\partial_y \phi\phi' + \phi\partial_y \phi'|Y = y, Z = z]$ which is $\bar{B}_2$. B_3 and B_4 are trivial. Upon inserting random variables, the statement follows.

Ad (iv) First note that S symmetric implies that $K = D_p\phi + \partial_y \phi\phi'$ is symmetric, which implies that $\mathbb{E}[K|Y, Z]$ is symmetric since

$$A_{ij} = \mathbb{E}[K_{ij}|Y, Z] = \mathbb{E}[K_{ji}|Y, Z] = A_{ji}.$$

This implies in turn that

$$\begin{aligned} \mathbb{E}[K|Y, Z] &= \mathbb{E}[D_p\phi|Y, Z] + \mathbb{E}[\partial_y \phi\phi'|Y, Z] \\ &= \mathbb{E}[D_p\phi|Y, Z] + \mathbb{E}[\partial_y \phi|Y, Z]\,\mathbb{E}[\phi'|Y, Z] + \mathbb{V}[\partial_y \phi, \phi'|Y, Z] \end{aligned}$$

is symmetric, from which $\mathbb{E}[D_p\phi|Y, Z] + \mathbb{E}[\partial_y \phi|Y, Z]\,\mathbb{E}[\phi'|Y, Z]$ is symmetric since $\mathbb{V}[\partial_y \phi, \phi'|Y, Z]$ is assumed to be symmetric.
By Proposition 2.1. this equals $D_p m + \partial_y m m'$.

Ad (v) Consider first the case the implication of V is Z measurable:
The 'if' part follows from (iv) and the observation that under measurable V, $\mathbb{V}[\partial_y \phi, \phi'|Y, Z] = 0$, and thus symmetric, by which $K = D_p m + \partial_y m m'$.

For the 'only if' we argue by contradiction: Assume $K_{ij} \neq K_{ji}$. We have to show now that $A_{ij} \neq A_{ji}$. But $A = K$ under measurable V, so that the result is obvious. This shows also why the converse does not hold under $\mathbb{V}\left[\partial_y \phi, \phi' | Y, Z\right] = 0$ alone,
because then $K_{ij} \neq K_{ji}$ does not necessarily imply $\mathbb{E}\left[A_{ij}|Y, Z\right] \neq \mathbb{E}\left[A_{ji}|Y, Z\right]$.

Consider now the reverse case, i.e. that
$\{S$ is symmetric and nsd iff $D_p m + \partial_y m m'$ is symmetric and nsd$\}$ implies V is Z measurable:
This is equivalent to:
If V is not Z measurable $\Rightarrow$
either
$\{S$ is symmetric nsd does not imply $D_p m + \partial_y m m'$ is symmetric nsd$\}$
or
$\{D_p m + \partial_y m m'$ is nsd symmetric and does not imply S is nsd symmetric$\}$.
The first statement can be true which is implied by $P2.4$ (iv) for $F = \mathbb{V}\left[\partial_y \phi, \phi' | Y, Z\right]$ not symmetric. Also the second may be true. To give an example where under non-measurability of V $D_p m + \partial_y m m'$ is symmetric but S is not, consider a two goods example, with two possible realizations, where the superscript $l = 1, 2$ denote these two realizations. Assume that $\pi^1 = \pi^2 = \frac{1}{2}$. Assume further that $\partial_y \phi = 0$ and that

$$S^1 = \begin{pmatrix} -1 & 1 \\ 2 & -1 \end{pmatrix},\; S^2 = \begin{pmatrix} -1 & 2 \\ 1 & -1 \end{pmatrix}$$

Note that $\mathbb{E}\left[S|Y, Z\right] = \begin{pmatrix} -1 & 1.5 \\ 1.5 & -1 \end{pmatrix}$ which is symmetric although the "individual" Slutsky matrices have not been so. □

Proof of Lemma 3.2: Note that by iterated expectations

$$\begin{aligned} \partial_y \mathbb{E}[\phi|Y, Z] &= \partial_y \mathbb{E}\left[\mathbb{E}[\phi|Y, Z, U]|Y, Z\right] \\ &= \mathbb{E}\left[\partial_y \mathbb{E}[\phi|Y, Z, U]|Y, Z\right] + \mathbb{E}\left[\mathbb{E}[\phi|Y, Z, U]\partial_y \log f_{U|Y,Z}(Y, Z)|Y, Z\right], \end{aligned}$$

by the same argument as above. Due to $F_{A|Y,Z,U} = F_{A|Z,U}$ we have $\partial_y \mathbb{E}[\phi|Y, Z, U] = \mathbb{E}[\partial_y \phi|Y, Z, U]$, by the same arguments as in $P2.1$. Again from iterated expectations

$$\mathbb{E}\left[\mathbb{E}[\partial_y \phi|Y, Z, U]|Y, Z\right] = \mathbb{E}[\partial_y \phi|Y, Z],$$

and hence the result follows.

Proof of Proposition 3.3:

Start by noting that - again by assumption - all conditional expectations exist. Let
$\psi(x,z,u) = \mu(x,z) + \sigma(x,z)\,u$. Then,

$$\begin{aligned}
\partial_x m(x,z) &= \partial_x \int\limits_{\mathcal{A}\times\mathcal{U}} \phi(\psi(x,z,u),z,a) F_{A,U|X,Z}(da,du,x,z) \\
&= \int\limits_{\mathcal{A}\times\mathcal{U}} \partial_y \phi(\psi(x,z,u),z,a)\left\{\partial_x \mu(x,z) + \partial_x \sigma(x,z)\,u\right\} F_{A,U|X,Z}(da,du,z) \\
&\quad + \int\limits_{\mathcal{U}} \int\limits_{\mathcal{A}} \phi(\psi(x,z,u),z,a) F_{A|U,Z}(da,u,z) \partial_x F_{U|X,Z}(du,x,z) \\
&= \mathbb{E}[\partial_y \phi | X=x, Z=z] \partial_x \mu(x,z) \\
&\quad + \mathbb{E}\left[\partial_y \phi U | X=x, Z=z\right] \partial_x \sigma(x,z) \\
&\quad + \mathbb{E}\left[W \partial_x \log f_{U|X,Z}(U,X,Z) | X=x, Z=z\right] \\
&= \mathbb{E}[\partial_y \phi | X=x, Z=z] \partial_x \mu(x,z) \\
&\quad + \mathbb{V}\left[\partial_y \phi, U | X=x, Z=z\right] \partial_x \sigma(x,z) \\
&\quad + \mathbb{E}\left[W \partial_x \log f_{U|X,Z}(U,X,Z) | X=x, Z=z\right],
\end{aligned}$$

where the last equality is due to $\mathbb{E}[U|X=x,Z=z]=0$. Since

$$\begin{aligned}
&\mathbb{V}\left[\partial_y \phi, U | X=x, Z=z\right] \\
=\ &\mathbb{V}\left[\mathbb{E}[\partial_y \phi | X,Z,U], U | X=x, Z=z\right] \\
&+ \mathbb{E}\left[\mathbb{V}\left[\partial_y \phi, U | X,Z,U\right] | X=x, Z=z\right],
\end{aligned}$$

and, by similar arguments as above,

$$\mathbb{E}[\partial_y \phi | X,Z,U] = \partial_x \mathbb{E}[W|X,Z,U] \partial_x \psi(X,Z,U)^{-1},$$

we obtain

$$\begin{aligned}
&\mathbb{V}\left[\partial_y \phi, U | X=x, Z=z\right] \partial_x \sigma(x,z) \\
=\ &\mathbb{V}\left[\partial_x \mathbb{E}[W|X,Z,U] \partial_x \psi(X,Z,U)^{-1}, U | X=x, Z=z\right] \partial_x \sigma(x,z) \\
=\ &\mathbb{E}\left[\partial_x \mathbb{E}[W|X,Z,U] \partial_x \psi(X,Z,U)^{-1} U | X=x, Z=z\right] \partial_x \sigma(x,z).
\end{aligned}$$

The result follows upon rearranging $\square$

Discussion: Suppose now that we have $X_j, j = 1, .., J$ instruments. Than the above derivations remain valid, with x replaced by x_j and **for all** j we have

$$\begin{aligned}\mathbb{E}[\partial_y \phi | X, Z] &= \partial_{x_j} m(X,Z) \partial_{x_j} \mu(X,Z)^{-1} \\ &\quad - \mathbb{V}\left\{\xi_j(X,Z,U), U | X, Z\right\} \partial_{x_j} \sigma(X,Z) \partial_{x_j} \mu(X,Z)^{-1} \\ &\quad - \mathbb{E}\left\{W \partial_{x_j} \log f_{U|X,Z}(X,Z) | X, Z\right\} \partial_{x_j} \mu(X,Z)^{-1} \ (a.s.),\end{aligned}$$

where $\xi_j(X,Z,U) \equiv \partial_{x_j} \mathbb{E}[W|X,Z,U] \left[\partial_{x_j} \mu(X,Z) + \partial_{x_j} \sigma(X,Z) U\right]^{-1}$. Thus, any linear combination of the rhs j yields $\mathbb{E}[\partial_y \phi | X, Z]$ as well. $\square$

Proof of Proposition 3.4: Follows straightforwardly by a combination of $P3.3$, with the arguments in $P2.4$.

Proof of Corollary 4.1:

$$\begin{aligned}&\mathbb{E}\left[AV(p_1^1, p_1^0, Y, V) | Y = y, Z = z\right] \\ =& \ \mathbb{E}\left[\lim_{n\to\infty} \sum_{j=1,..,n} \phi_1(p_1^{(j)}, \overline{p_{-1}}, y, v)\left(p_1^{(j+1)} - p_1^{(j)}\right) | Y = y, Z = z\right] \\ =& \lim_{n\to\infty} \sum_{j=1,..,n} \mathbb{E}\left[\phi_1(p_1^{(j)}, \overline{p_{-1}}, y, v)\left(p_1^{(j+1)} - p_1^{(j)}\right) | Y = y, Z = z\right] \\ =& \int_{p_1^0}^{p_1^1} m_1(\pi, \overline{p_{-1}}, y, z) d\pi,\end{aligned}$$

where the second equality follows by the boundedness of ϕ_1, for any partition $\left(p_1^{(j)}\right)_{j=1,..,n}$, $p_1^{(1)} = p_1^0$, $p_1^{(n)} = p_1^1$ and $\lim_{n\to\infty} \sup_j \left|p_1^{(j+1)} - p_1^{(j)}\right| = 0$. The second statement is trivial using iterated expectations. $\square$

Assumptions for Section 5: Let us state the assumptions to be made in the following. Without further mentioning, we shall always assume that Σ is positive definite, as well as $K \geq 0$, $\int K(u)du = 1$ and $\int K^4(u)du < \infty$.

(A1) The η_i are zero mean and unity diagonal variance random vectors s.t. η_i is independent of $X_1, \cdots, X_i, \eta_1, \cdots, \eta_{i-1}, , \varepsilon_1, \cdots, \varepsilon_i$ for each $i \geq 1$ and every element of $\eta_i\eta_i'$ is uniformly integrable. The ε_i are zero mean and unit variance random vectors s.t. ε_i is independent of $X_1, \cdots, X_i, \eta_1, \cdots, \eta_i, , \varepsilon_1, \cdots, \varepsilon_{i-1}$ for each $i \geq 1$. Moreover, the ε_i have finite fourth moments and $\mathbb{E}\left[\varepsilon_i^l \eta_i\right]$, $l = 1, 2, 3$ are finite is uniformly integrable. Finally, ε_i and $\varepsilon_i^l \eta_i$, $l = 1, 2, 3$ are uniformly integrable.

(A2) The X_i are iid with common density f_X.

(A3) K is twice continuously differentiable with bounded second derivatives.

(A4) f_X is bounded, as well as $f_X(x) > 0 \ \forall x$.

(A5) Every element of the Hessian of m_j, $j = 1, .., L$ and of σ_2^2 is bounded.

(A6) Every element of Ξ is bounded.

(A7) $h_n \to 0, nh^d \to \infty$.

(A8) $nh^{d+4} \to 0$.

Note that the boundedness restrictions (A5)and (A6) are not as restrictive in this scenario, as the dependent variable only takes values in $[0, 1]$.

Propostion 5.1: *Let the model be as defined above, and let A1-A8 hold. Then follows that*

$$\sqrt{nh^d}\left(\hat{\theta}(x_0) - \theta(x_0) - h^2 bias(x_0)\right) \xrightarrow{d} \mathcal{N}(0, \Xi(x_0) \otimes A),$$

where all quantities are defined in the text.

Intuition of Proposition 5.1:

This proposition is best illustrated in the one equation case. In this example, we will also illustrate the issue of pre-estimation. Assume that the model is given by $W_i = m(X_i, V_i) + \sigma(X_i, V_i)\varepsilon_i$, where ε_i is a mean zero, unit variance scalar random variable, independent of the random scalars X_i, and V_i. Moreover, $V_i = X_i - \mu(Q_i)$ and $\hat{V}_i = X_i - \hat{\mu}(Q_i)$, where the hat denotes a Nadaraya Watson pre-estimator. Hence, $V_i = \hat{V}_i + (\hat{\mu}_i - \mu_i)$ in an obvious notation. Let

$$\hat{U}_i^2 = (W_i - \hat{m}(X_i))^2 = U_i^2 + 2U_i\left(\hat{m}(X_i) - m(X_i)\right) + \left(\hat{m}(X_i) - m(X_i)\right)^2,$$

where $U_i = W_i - m(X_i)$. Then

$$\begin{bmatrix} W_i \\ \hat{U}_i^2 \end{bmatrix} = \begin{bmatrix} 1 & (X_i - x_0) & (V_i - v_0) & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & (X_i - x_0) & (V_i - v_0) \end{bmatrix} \begin{bmatrix} m(x_0, v_0) \\ \partial_x m(x_0, v_0) \\ \partial_v m(x_0, v_0) \\ \sigma^2(x_0, v_0) \\ \partial_x \sigma^2(x_0, v_0) \\ \partial_v \sigma^2(x_0, v_0) \end{bmatrix}$$
$$+ \begin{bmatrix} 0 \\ 2U_i \left(\hat{m}(X_i) - m(X_i)\right) + \left(\hat{m}(X_i) - m(X_i)\right)^2 \end{bmatrix}$$
$$+ \begin{bmatrix} \frac{1}{2}\partial_x^2 m(X_r, V_r) \left(X_i - x_0\right)^2 \\ \frac{1}{2}\partial_x^2 \sigma^2(X_r, V_r) \left(X_i - x_0\right)^2 \end{bmatrix} + \begin{bmatrix} \partial_x \partial_v m(X_r, V_r) \left(X_i - x_0\right) \left(V_i - v_0\right) \\ \partial_x \partial_v \sigma^2(X_r, V_r) \left(X_i - x_0\right) \left(V_i - v_0\right) \end{bmatrix}$$
$$+ \begin{bmatrix} \frac{1}{2}\partial_v^2 m(X_r, V_r) \left(V_i - v_0\right)^2 \\ \frac{1}{2}\partial_v^2 \sigma^2(X_r, V_r) \left(V_i - v_0\right)^2 \end{bmatrix} + \begin{bmatrix} \sigma(X_i)\varepsilon_i \\ \sigma^2(X_i) \left(\varepsilon_i^2 - 1\right) \end{bmatrix}. \tag{A.2}$$

This resembles the standard type of expansion save for the second summand, which constitutes another bias term, and will therefore be called $Bias_2$. The usual locally linear estimator for
$\beta(x_0) = (m(x_0, v_0), h\partial_x m(x_0, v_0), h\partial_v m(x_0, v_0), \sigma^2(x_0, v_0), h\partial_x \sigma^2(x_0, v_0), h\partial_v \sigma^2(x_0, v_0))'$
is then given by

$$\hat{\beta}(x_0) = \left[I_2 \otimes [\mathbb{X}'\mathbb{X}]\right]^{-1} \left(I_2 \otimes \mathbb{X}'\right) \mathbb{S}$$

where,

$$\mathbb{X}'\mathbb{X} = \begin{bmatrix} \sum_{i=1}^n K_{ni} & . & . \\ \sum_{i=1}^n K_{ni} \frac{X_i - x_0}{h} & \sum_{i=1}^n K_{ni} \left(\frac{X_i - x_0}{h}\right)^2 & . \\ \sum_{i=1}^n K_{ni} \frac{\hat{V}_i - v_0}{h} & \sum_{i=1}^n K_{ni} \frac{X_i - x_0}{h} \frac{\hat{V}_i - v_0}{h} & \sum_{i=1}^n K_{ni} \left(\frac{\hat{V}_i - v_0}{h}\right)^2 \end{bmatrix},$$

where $K_{ni} = K_{ni}(x_0, v_0) = K\left(\frac{X_i - x_0}{h}\right) K\left(\frac{\hat{V}_i - v_0}{h}\right)$,

and
$$\left(I_2 \otimes \mathbb{X}'\right) \mathbb{S} = \begin{bmatrix} \sum_{i=1}^n K_{ni}(x_0, v_0) W_i \\ \sum_{i=1}^n K_{ni}(x_0, v_0) W_i \left(X_i - x_0\right)/h \\ \sum_{i=1}^n K_{ni}(x_0, v_0) W_i \left(\hat{V}_i - v_0\right)/h \\ \sum_{i=1}^n K_{ni}(x_0, v_0) \hat{U}_i^2 \\ \sum_{i=1}^n K_{ni}(x_0, v_0) \hat{U}_i^2 \left(X_i - x_0\right)/h \\ \sum_{i=1}^n K_{ni}(x_0, v_0) \hat{U}_i^2 \left(\hat{V}_i - v_0\right)/h \end{bmatrix}.$$

The first thing to note is that

$$
\begin{aligned}
K_{ni}(x_0, v_0) &= K_i\left(\frac{X_i - x_0}{h}\right) K_i\left(\frac{\hat{V}_i - v_0}{h}\right) \\
&= K_i\left(\frac{X_i - x_0}{h}\right) K_i\left(\frac{V_i - v_0}{h}\right) \\
&+ K_i\left(\frac{X_i - x_0}{h}\right) K_i'\left(\frac{V_i - v_0}{h}\right) \frac{\mu_i - \hat{\mu}_i}{h} \\
&+ \frac{1}{2} K_i\left(\frac{X_i - x_0}{h}\right) K_i''\left(\frac{V_i - v_0}{h} + \lambda \frac{\mu_i - \hat{\mu}_i}{h}\right) \left(\frac{\mu_i - \hat{\mu}_i}{h}\right)^2,
\end{aligned}
\tag{A.3}
$$

and $\left(\hat{V}_i - v_0\right)/h = \left(V_i - v_0\right)/h + \left(\mu_i - \hat{\mu}_i\right)/h$. Define

$$
K_{ni}^*(x_0, v_0) = K_i\left(\frac{X_i - x_0}{h}\right) K_i\left(\frac{V_i - v_0}{h}\right).
$$

Inserting the expression (A.2) into the estimator and rearranging produces

$$
\begin{aligned}
\hat{\beta}(x_0) - \beta(x_0) &= \left[I_2 \otimes [\mathbb{X}'\mathbb{X}]\right]^{-1} \left(I_2 \otimes \mathbb{X}'\right) \left(\mathbb{B}_1 + \mathbb{B}_2\right) \\
&\quad + \left[I_2 \otimes [\mathbb{X}'\mathbb{X}]\right]^{-1} \left(I_2 \otimes \mathbb{X}'\right) \mathbb{U},
\end{aligned}
$$

where typical bias expressions in $\mathbb{B}_1$ are

$$
\frac{1}{2}\begin{bmatrix}
\sum_{i=1}^n K_{ni}^*(x_0, v_0)\partial_x^2 m(X_r, V_r)\left(X_i - x_0\right)^2 \\
\sum_{i=1}^n K_{ni}^*(x_0, v_0)\partial_x^2 m(X_r, V_r)\left(X_i - x_0\right)^3/h \\
\sum_{i=1}^n K_{ni}^*(x_0, v_0)\partial_x^2 m(X_r, V_r)\left(X_i - x_0\right)^2\left(V_i - v_0\right)/h \\
\sum_{i=1}^n K_{ni}^*(x_0, v_0)\partial_x^2 \sigma^2(X_r, V_r)\left(X_i - x_0\right)^2 \\
\sum_{i=1}^n K_{ni}^*(x_0, v_0)\partial_x^2 \sigma^2(X_r, V_r)\left(X_i - x_0\right)^3/h \\
\sum_{i=1}^n K_{ni}^*(x_0, v_0)\partial_x^2 \sigma^2(X_r, V_r)\left(X_i - x_0\right)^2\left(V_i - v_0\right)/h
\end{bmatrix},
$$

which may be treated by standard methods, and

$$
\frac{1}{2}\begin{bmatrix}
\sum_{i=1}^n K_i\left(\frac{X_i - x_0}{h}\right) K_i'\left(\frac{V_i - v_0}{h}\right)\partial_x^2 m(X_r, V_r)\left(X_i - x_0\right)^2\left(\mu_i - \hat{\mu}_i\right)/h \\
\sum_{i=1}^n K_i\left(\frac{X_i - x_0}{h}\right) K_i'\left(\frac{V_i - v_0}{h}\right)\partial_x^2 m(X_r, V_r)\left(X_i - x_0\right)^3\left(\mu_i - \hat{\mu}_i\right)/h^2 \\
\sum_{i=1}^n K_i\left(\frac{X_i - x_0}{h}\right) K_i'\left(\frac{V_i - v_0}{h}\right)\partial_x^2 m(X_r, V_r)\left(X_i - x_0\right)^2\left(V_i - v_0\right)\left(\mu_i - \hat{\mu}_i\right)/h^2 \\
\sum_{i=1}^n K_i\left(\frac{X_i - x_0}{h}\right) K_i'\left(\frac{V_i - v_0}{h}\right)\partial_x^2 \sigma^2(X_r, V_r)\left(X_i - x_0\right)^2\left(\mu_i - \hat{\mu}_i\right)/h \\
\sum_{i=1}^n K_i\left(\frac{X_i - x_0}{h}\right) K_i'\left(\frac{V_i - v_0}{h}\right)\partial_x^2 \sigma^2(X_r, V_r)\left(X_i - x_0\right)^3\left(\mu_i - \hat{\mu}_i\right)/h^2 \\
\sum_{i=1}^n K_i\left(\frac{X_i - x_0}{h}\right) K_i'\left(\frac{V_i - v_0}{h}\right)\partial_x^2 \sigma^2(X_r, V_r)\left(X_i - x_0\right)^2\left(V_i - v_0\right)\left(\mu_i - \hat{\mu}_i\right)/h^2
\end{bmatrix},
$$

which involves the additional expression $(\mu_i - \hat{\mu}_i)/h$. The same is true for the second bias term

$$(I_2 \otimes \mathbb{X}')\,\mathbb{B}_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \sum_{i=1}^n K_{ni}\left(U_i\left(\hat{m}(X_i) - m(X_i)\right) + \frac{1}{2}\left(\hat{m}(X_i) - m(X_i)\right)^2\right) \\ \sum_{i=1}^n K_{ni}\left(U_i\left(\hat{m}(X_i) - m(X_i)\right) + \frac{1}{2}\left(\hat{m}(X_i) - m(X_i)\right)^2\right)(X_i - x_0)/h \\ \sum_{i=1}^n K_{ni}\left(U_i\left(\hat{m}(X_i) - m(X_i)\right) + \frac{1}{2}\left(\hat{m}(X_i) - m(X_i)\right)^2\right)\left(\hat{V}_i - v_0\right)/h \end{bmatrix},$$

which may split up in a similar fashion. Finally, the last expression

$$(I_2 \otimes \mathbb{X}')\,\mathbb{U} = \begin{bmatrix} \sum_{i=1}^n K_{ni}(x_0,v_0)\sigma(X_i)\varepsilon_i \\ \sum_{i=1}^n K_{ni}(x_0,v_0)\sigma(X_i)\varepsilon_i\,(X_i - x_0)/h \\ \sum_{i=1}^n K_{ni}(x_0,v_0)\sigma(X_i)\varepsilon_i\left(\hat{V}_i - v_0\right)/h \\ \sum_{i=1}^n K_{ni}(x_0,v_0)\sigma^2(X_i)\left(\varepsilon_i^2 - 1\right) \\ \sum_{i=1}^n K_{ni}(x_0,v_0)\sigma^2(X_i)\left(\varepsilon_i^2 - 1\right)(X_i - x_0)/h \\ \sum_{i=1}^n K_{ni}(x_0,v_0)\sigma^2(X_i)\left(\varepsilon_i^2 - 1\right)\left(\hat{V}_i - v_0\right)/h \end{bmatrix},$$

has to be split into

$$[(I_2 \otimes \mathbb{X}')\,\mathbb{U}]^* = \begin{bmatrix} \sum_{i=1}^n K_{ni}^*(x_0,v_0)\sigma(X_i)\varepsilon_i \\ \sum_{i=1}^n K_{ni}^*(x_0,v_0)\sigma(X_i)\varepsilon_i\,(X_i - x_0)/h \\ \sum_{i=1}^n K_{ni}^*(x_0,v_0)\sigma(X_i)\varepsilon_i\,(V_i - v_0)/h \\ \sum_{i=1}^n K_{ni}^*(x_0,v_0)\sigma^2(X_i)\left(\varepsilon_i^2 - 1\right) \\ \sum_{i=1}^n K_{ni}^*(x_0,v_0)\sigma^2(X_i)\left(\varepsilon_i^2 - 1\right)(X_i - x_0)/h \\ \sum_{i=1}^n K_{ni}^*(x_0,v_0)\sigma^2(X_i)\left(\varepsilon_i^2 - 1\right)(V_i - v_0)/h \end{bmatrix},$$

and $\mathbb{B}_3 = (I_2 \otimes \mathbb{X}')\,\mathbb{U} - [(I_2 \otimes \mathbb{X}')\,\mathbb{U}]^*$.

Now the procedure is simple: We will show that $\frac{1}{nh^2}\left[I_2 \otimes [\mathbb{X}'\mathbb{X}]\right]^{-1}$ is $O_p(1)$, $\frac{1}{\sqrt{nh^2}}\frac{1}{h^2}\left[(I_2 \otimes \mathbb{X}')(\mathbb{B}_1 + \mathbb{B}_2) + \mathbb{B}_3\right]$ is also $O_p(1)$, i.e. $\frac{1}{\sqrt{nh^2}}(I_2 \otimes \mathbb{X}')(\mathbb{B}_1 + \mathbb{B}_2 + \mathbb{B}_3)$ is $o_p(1)$. Then we will apply a central limit theorem to $\frac{1}{\sqrt{nh^2}}\left[(I_2 \otimes \mathbb{X}')\,\mathbb{U}\right]^*$ and show that this converges to a nondegenerate, mean zero random vector. The major difficulty in this argumentation comes from the fact that all expression contain pre-estimated quantities, namely $\hat{m}_i - m_i$, and $\hat{\mu}_i - \mu_i$, and therefore some type of uniform convergence result has to be applied. To show this, we employ a functional expansions as in Ait-Sahalia et al. (2002). Details of the proof are available from the author upon request.

9 Appendix 2

In this appendix we will exclusively treat the asymptotics for the systems local polynomial estimator, including with pre-estimated dependent variable and regressors. The structure will be as follows: In the first subsection we will give a proof of the systems local polynomial including pre-estimated squared residuals, but excluding pre-estimated regressors. In the second subsection we will establish how the results change, if pre-estimated regressors are included.

9.1 Proof of Proposition 5.1:

The estimator given by the first order conditions is

$$\begin{aligned}\hat{\theta}(x_0) &= \left(I_{\bar{L}} \otimes [\mathbb{X}'\mathbb{X}]\right)^{-1} \left(I_{\bar{L}} \otimes \mathbb{X}'\right) \mathbb{S} \\ &= \theta(x_0) + \left(I_{\bar{L}} \otimes [\mathbb{X}'\mathbb{X}]\right)^{-1} \left(I_{\bar{L}} \otimes \mathbb{X}'\right) \mathbb{B}_1 \\ &+ \left(I_{\bar{L}} \otimes [\mathbb{X}'\mathbb{X}]\right)^{-1} \left(I_{\bar{L}} \otimes \mathbb{X}'\right) \mathbb{B}_2 \\ &+ \left(I_{\bar{L}} \otimes [\mathbb{X}'\mathbb{X}]\right)^{-1} \left(I_{\bar{L}} \otimes \mathbb{X}'\right) \mathbb{U}\end{aligned} \tag{AA.1}$$

Here, $\mathbb{S}$ is the $\bar{L}n \times 1$ vector given by $\mathbb{S} = (\tilde{S}_1', .., \tilde{S}_{M-1}')'$ with $\tilde{S}_j = (K_1^{\frac{1}{2}} S_{j1}, .., K_n^{\frac{1}{2}} S_{jn})'$ $\forall j = 1, .., \bar{L}$,
and recall that $\bar{L} = (L-1) + 1 + 1 = L + 1$,
$I_{\bar{L}} \otimes \mathbb{X}$ is the $\bar{L}n \times \bar{L}\,(d+1)$ with $\mathbb{X} = (X^0, .., X^d)$ and
$X^m = (K_1^{\frac{1}{2}} \left(X_1^m - x_0^m\right)/h, .., K_n^{\frac{1}{2}} \left(X_n^m - x_0^m\right)/h)'$ $\forall m = 1, .., d$ as well as
$X^0 = (K_1^{\frac{1}{2}}, .., K_n^{\frac{1}{2}})'$. Moreover, we need the notation
$X_i - x_0 = (1, \left(X_i^1 - x_0^1\right)/h, .., \left(X_i^d - x_0^d\right)/h))'$
Regarding the bias terms, the first, $\mathbb{B}_1$, is the $\bar{L}n \times 1$ vector given by $\mathbb{B}_1 = (\tilde{B}_1', .., \tilde{B}_{\bar{L}}')'$
with $\tilde{B}_j = (K_1^{\frac{1}{2}} \frac{h^2}{2} \iota' \mathcal{H}_m(X_{r1}) \iota, .., K_n^{\frac{1}{2}} \frac{h^2}{2} \iota' \mathcal{H}_m(X_{rn}) \iota)'$ $\forall j = 1, .., L$, where
$\mathcal{H}_m(X_{r1})$ is the Hessian of m at an intermediate position, $X_{r1} = x_0 + \lambda(X_i)'(X_i - x_0)$
and $\tilde{B}_{\bar{L}} = (K_1^{\frac{1}{2}} \frac{h^2}{2} \mathcal{H}\sigma_2^2(X_{r1})' \iota, .., K_n^{\frac{1}{2}} \frac{h^2}{2} \mathcal{H}\sigma_2^2(X_{rn})' \iota)'$,
where $\mathcal{H}\sigma_2^2$ denotes the vector of second derivatives including cross derivatives.
The second bias term, $\mathbb{B}_2$, is the $\bar{L}n \times 1$ vector given by $\mathbb{B}_2 = (0, ..0, \breve{B}_{\bar{L}}')'$,
where $\breve{B}_{\bar{L}} = (K_1^{\frac{1}{2}} G_1, .., K_n^{\frac{1}{2}} G_n)'$ and

$G_i = 2\sigma_2(X_i)U_i(\hat{\mu}(X_i) - \mu(X_i)) + (\hat{\mu}(X_i) - \mu(X_i))^2$
Finally, $\mathbb{U}$ is the $\overline{L}n \times 1$ vector given by $\mathbb{U} = (\tilde{U}_1', .., \tilde{U}_{\bar{L}}')'$ with
$\tilde{U}_j = (K_1^{\frac{1}{2}}\Sigma_{1,j}(X_1)\eta_1, .., K_n^{\frac{1}{2}}\Sigma_{1,j}(X_n)\eta_n)' \; \forall j = 1, .., L-1,$
$\tilde{U}_L = (K_1^{\frac{1}{2}}\sigma_2(X_1)\varepsilon_1, .., K_n^{\frac{1}{2}}\sigma_2(X_n)\varepsilon_n)'$ and
$\tilde{U}_{\overline{L}} = (K_1^{\frac{1}{2}}\sigma_2^2(X_1)\left(\varepsilon_1^2 - 1\right), .., K_n^{\frac{1}{2}}\sigma_2^2(X_n)\left(\varepsilon_n^2 - 1\right))'.$

The proof proceeds via establishing the validity of the following four lemmata, the first of which is concerned with the asymptotic behavior of the squared regressor matrix in A.1

Lemma A.1

$$\frac{1}{nh^d}\left(I_{\bar{L}} \otimes [\mathbb{X}'\mathbb{X}]\right)^{-1} \xrightarrow{p} f_X(x_0)^{-1}\left[I_{\bar{L}} \otimes B^{-1}\right]$$

The second lemma treats the first bias expression:

Lemma A.2

$$\text{plim}_{n\to\infty}\frac{1}{h^2}\frac{1}{\sqrt{nh^d}}\left(I_{\bar{L}} \otimes \mathbb{X}'\right)\mathbb{B}_1 = bias(x_0)$$

The third lemma establishes that the second bias expression vanishes even faster

Lemma A.3

$$\text{plim}_{n\to\infty}\frac{1}{h^2}\frac{1}{\sqrt{nh^d}}\left(I_{\bar{L}} \otimes \mathbb{X}'\right)\mathbb{B}_2 = 0.$$

Finally, the last Lemma establishes asymptotic normality for the rhs vector in (A.1).

Lemma A.4

$$\frac{1}{\sqrt{nh^d}}\left(I_{\bar{L}} \otimes \mathbb{X}'\right)\mathbb{U} \xrightarrow{d} \mathcal{N}\left(0, f_X(x_0)\left[\Xi(x_0) \otimes C\right]\right).$$

From this Lemmata the result is obvious.

Proof of Lemma A.1. The proof is well-known. In particular, L_2 convergence follows by standard arguments. ■

Proof of Lemma A.2. The bias term in (A.1) consists of two types of components. One comes from differentiating twice w.r.t. the same arguments, one comes from cross differentiating. Consider the first one first. A typical expression - involving the j-th term, hence the superscript - is of the form

$$w_{jn} = \frac{1}{\sqrt{nh^d}} \frac{h^2}{2} \sum_{i=1}^{n} \left(Q_i^j\right)^2 K(X_i - x_0) \partial_{x^j}^2 m \left(x_0 + h\eta(x_0)' Q_i\right),$$

where $Q_i^j = (X_i^j - x_0)/h$ and $Q_i = \left(Q_i^1, .., Q_i^d\right)$. As an example, take the first element of $(X_i - x_0)$, which is unity. Writing $\bar{S}_{ni}^j$ for the general term under the sum, at each joint continuity point x of $\partial_x^2 m$ and f,

$$\begin{aligned}
& \mathbb{E}\left[\bar{S}_{ni}^{M+1}\right] \\
= & \int_X \left(\frac{y^j - x^j}{h}\right)^2 K\left(\frac{y-x}{h}\right) * \\
& \partial_{x^j}^2 m \left(x_0^{-j} + \eta^{-j}(x_0) h \left(y^{-j} - x^{-j}\right), x_0^j + \eta^j(x_0) h \left(y^j - x^j\right)\right) f_X(y) dy \\
= & h^d \partial_{x^j}^2 m(x_0) \mu_2 + o(1).
\end{aligned}$$

Thus,

$$\mathbb{E}\left[w_{jn}\right] = \frac{1}{\sqrt{nh^d}} n h^{d+2} O(1) = \sqrt{nh^{d+4}} O(1),$$

A similar argument as above can be used in connection with the *iid* assumption to show that $\mathbb{V}\left[w_{jn}\right] = o(1)$. Hence

$$w_{jn} \xrightarrow{P} 0, \text{ and } \sqrt{nh^d} h^{-2} w_{jn} \xrightarrow{P} \frac{1}{2} \partial_{x^j}^2 m(x) \mu_2$$

The same argument holds for any other derivative involving twice differencing w.r.t. to the same variable. The terms involving cross derivatives vanish even faster with $O_p(nh^{d+8})$. Collecting terms, the statement follows. ■

Proof of Lemma A.4
First, recall the notation $X_i - x_0 = (1, \left(X_i^1 - x_0^1\right)/h, .., \left(X_i^d - x_0^d\right)/h))'$.
We show: $(\phi_{ni}, \mathcal{F}_{ni})$, $i = 1, \ldots, n$, $n \geq 1$, with $\phi_{ni} = (\phi_{ni1}', .., \phi_{ni\bar{L}}')'$ and
$\phi_{nij} = \frac{1}{\sqrt{nh^d}} K\left(\left(X_i - x_0\right)/h\right)\left(X_i - x_0\right) \Sigma_{1,j}(X_i) \eta_{ni}$, $\forall j = 1, .., L-1$,
$\phi_{niLs} = \frac{1}{\sqrt{nh^d}} K\left(\left(X_i - x_0\right)/h\right)\left(X_i - x_0\right) \sigma_2(X_i) \varepsilon_{ni}$,
$\phi_{ni\bar{L}s} = \frac{1}{\sqrt{nh^d}} K\left(\left(X_i - x_0\right)/h\right)\left(\left(X_i - x_0\right)/h\right) \sigma_2^2(X_i)\left(\varepsilon_{ni}^2 - 1\right)$,

$\mathcal{F}_{ni} = \mathcal{F}_i$,
is a martingale difference array such that

(i) $\text{plim}_{n\to\infty} \sum_{i=1}^n \mathbb{E}\left[\phi_{ni}\phi_{ni}'|\mathcal{F}_{i-1}\right] = \Xi(x_0) \otimes C$,
(ii) $\text{plim}_{n\to\infty} \sum_{i=1}^n \mathbb{E}\left[\|\phi_{ni}\|^2 \mathbf{1}_{\{\|\phi_{ni}\|>\delta\}}|\mathcal{F}_{i-1}\right] = 0$ for every $\delta > 0$.

The assertion will then follow from a standard central limit theorem for martingale difference arrays (e.g., Pollard (1984)). Of course, in the present scenario of independent row entries, any other central limit theorem for such arrays will also do. But the above version easily lends itself for extension to certain dependence structures, e.g. mixing processes. The martingale difference is obvious, as to (i), note that

$$\begin{aligned}
\mathbb{E}\left[\phi_{nij}\phi_{nik}'|\mathcal{F}_{i-1}\right] &= \frac{1}{nh^d}\mathbb{E}\left[K_i^2\left(X_i - x_0\right)\Sigma_{1,j}(X_i)\eta_{ni}\eta_{ni}'\Sigma_{1,k}'(X_i)\left(X_i - x_0\right)'|\mathcal{F}_{i-1}\right] \\
&= \frac{1}{nh^d}K_i^2\left[\Sigma_{jk}(X_i)\left(\left(X_i - x_0\right)\left(X_i - x_0\right)'\right)\right],
\end{aligned}$$

where Σ_{jk} denotes the jk-th element of Σ_1. The same holds for terms involving σ_2, σ_2^2. Therefore, by similar reasoning as above,

$$\begin{aligned}
\sum_{i=1}^n \mathbb{E}\left\{\phi_{ni}\phi_{ni}'|\mathcal{F}_{i-1}\right\} &= \frac{1}{nh^d}\sum_{i=1}^n K_i^2\left[\Xi(X_i) \otimes \left(\left(X_i - x_0\right)\left(X_i - x_0\right)'\right)\right] \\
&\xrightarrow{p} f_Z(x_0)\left(\Xi(X_i) \otimes C\right).
\end{aligned}$$

As to (ii), note first that

$$\begin{aligned}
&\sum_{i=1}^n \mathbb{E}\left\{\|\phi_{ni}\|^2 \mathbf{1}_{\{\|\phi_{ni}\|>\delta\}}|\mathcal{F}_{i-1}\right\} \\
\leq &\sum_{i=1}^n \mathbb{E}\left\{c \max_{j=1,..,\overline{L}} \left\|\phi_{nij}\right\|^2 \mathbf{1}_{\{C \max_{j=1,..,\overline{L}}\|\phi_{ni}\|>\delta\}}|\mathcal{F}_{i-1}\right\},
\end{aligned}$$

with a suitably defined finite positive constant c. Without loss of generality, assume that the max is attained for $j = 1$. Now, with $\gamma_n = \frac{1}{\sqrt{nh^d}}$ and $b_{ni} =$

$K_i(X_i - x_0)$, we have

$$\begin{aligned}
&\sum_{i=1}^{n} \mathbb{E}\left[\|\phi_{ni1}\|^2 \mathbf{1}_{\{\max_{j=1,..,L} \|\phi_{n1}\|>\delta/C\}} |\mathcal{F}_{i-1}\right] \\
&= \gamma_n^2 \sum_{i=1}^{n} \mathbb{E}\left[||b_{ni}||^2 \Sigma_{1,11}(X_i)\eta_i^2 \mathbf{1}_{\{||b_{ni}||\Sigma_{1,11}(X_i)|\eta_i|>\delta/C\gamma_n\}}|\mathcal{F}_{i-1}\right] \\
&\leq \gamma_n^2 \sum_{i=1}^{n} \mathbb{E}\left[||b_{ni}||^2 \Sigma_{1,11}(X_i)\eta_i^2 \mathbf{1}_{\{||b_{ni}||^2\Sigma_{1,11}(X_i)>\delta/C\gamma_n\}}|\mathcal{F}_{i-1}\right] \\
&\quad + \gamma_n^2 \sum_{i=1}^{n} \mathbb{E}\left[||b_{ni}||^2 \Sigma_{1,11}(X_i)\eta_i^2 \mathbf{1}_{\{\eta_i^2>\delta/C\gamma_n\}}|\mathcal{F}_{i-1}\right] \\
&= \gamma_n^2 \sum_{i=1}^{n} ||b_{ni}||^2 \Sigma_{1,11}(X_i) \mathbf{1}_{\{||b_{ni}||^2\Sigma_{1,11}(X_i)>\delta/C\gamma_n\}} \\
&\quad + \gamma_n^2 \sum_{i=1}^{n} ||b_{ni}||^2 \Sigma_{1,11}(X_i)\mathbb{E}[\eta_i^2 \mathbf{1}_{\{\eta_i^2>\delta/C\gamma_n\}}].
\end{aligned}$$

Here, for the inequality, we have used the simple fact that $|ab| > \epsilon$ implies $a^2 > \epsilon$ or $b^2 > \epsilon$. Under (A3) the $||b_{ni}||^2$ are zero if X_i is outside a h-neighborhood of x_0. Therefore the $||b_{ni}||^2\Sigma_{1,11}(X_i)$ are uniformly bounded by a constant for h small enough, if the realizations of X_i are continuity points of $\Sigma_{1,11}$, and hence the first term eventually becomes zero with probability one. For the second term, note that by uniform integrability of the η_i^2, since $\gamma_n \to 0$,

$$\lim_{n\to\infty} \sup_i \mathbb{E}[\eta_i^2 \mathbf{1}_{\{\eta_i^2>\delta/C\gamma_n\}}] = 0,$$

and since

$$\gamma_n^2 \sum_{i=1}^{n} b_{ni} b_{ni}' \Sigma_{1,11}(X_i) \xrightarrow{p} V,$$

where V is a nonrandom matrix, the last term tends to zero in probability as well. ■

Proof of Lemma A.3. Returning to the second bias term, this is actually a $d+1$ vector, i.e.

$$\begin{aligned}
& \frac{1}{\sqrt{nh^d}} \sum_{i=1}^{n} K_i \left(X_i - x_0 \right) \left[2U_i \left(\hat{\mu}(X_i) - \mu(X_i) \right) + \left(\hat{\mu}(X_i) - \mu(X_i) \right)^2 \right] \\
= & \frac{2}{\sqrt{nh^d}} \sum_{i=1}^{n} K_i \left(X_i - x_0 \right) U_i \left(\hat{\mu}(X_i) - \mu(X_i) \right) \\
& + \frac{1}{\sqrt{nh^d}} \sum_{i=1}^{n} K_i \left(X_i - x_0 \right) \left(\hat{\mu}(X_i) - \mu(X_i) \right)^2 .
\end{aligned}$$

Consider the second expression first. Take again a typical element, which is

$$\sqrt{\frac{n}{h^d}} \frac{1}{n} \sum_{i=1}^{n} K_i \left(\frac{X_i^j - x_0^j}{h} \right)^s \left(\hat{\mu}(X_i) - \mu(X_i) \right)^2 , \ j = 0, 1., , .d, \ s = 0, 1.$$

and let $\hat{D}_n = \frac{1}{n} \sum_{i=1}^{n} K_i \left[\left(X_i^j - x_0^j \right) / h \right]^2 \left(\hat{\mu}(X_i) - \mu(X_i) \right)^2 .$
Introduce the notation, $X_i^j = X_i$, for the j-th component, and $X_i^{-j} = Z_i$ for the others. Then, $\hat{D}_n$

$$\hat{D}_n = \int \int \left(\frac{x - x_0}{h} \right)^s K \left(\frac{x - x_0}{h}, \frac{z - z_0}{h} \right) \left[\widehat{\mu}(x, z) - \mu(x, z) \right]^2 d\hat{F}_{XZ}, \ s = 0, 1.$$

where $\hat{F}_{XZ}$ is the empirical *c.d.f.* of X_i and Z_i. We follow the proof in Ait-Sahalia, Bickel and Stoker (2002, ASBS for short). The strategy will be to establish the behavior of the statistic $\hat{D}_n$ which is a functional $\Gamma \left(\hat{\mu}, \hat{F}_{XZ} \right)$, by studying first the behavior of $D_n = \Gamma \left(\hat{\mu}, F_{XZ} \right)$, and show then that the difference is asymptotically negligible. As in ASBS we analyze $\Gamma \left(\hat{\mu}, F_{XZ} \right)$ using a functional expansion around $\Gamma \left(\mu, F_{XZ} \right)$. Introduce the following notation. Let $f_{WXZ}(w, x, z)$ denote the joint density of (W_i, X_i, Z_i), and let $f_{XZ}(x, z)$ denote the joint density of (X_i, Z_i). Let

$$\widehat{f_{YXZ}}(y, x, z) = \frac{1}{nh^{d+1}} \sum_{i=1}^{n} K \left(\frac{Y_i - y}{h}, \frac{X_i - x}{h}, \frac{Z_i - z}{h} \right)$$

denote a Kernel based estimator. Similarly, let

$$\widehat{f_{XZ}}(x, z) = \frac{1}{nh^d} \sum_{i=1}^{n} K \left(\frac{X_i - x}{h}, \frac{Z_i - z}{h} \right)$$

and

$$\widehat{\mu}(x,z) = \frac{\int y\widehat{f_{YXZ}}(y,x,z)dy}{\widehat{f_{XZ}}(x,z)}.$$

Let

$$\begin{aligned}
\varphi(t,x,z) &= \frac{\int yf_{YXZ}(y,x,z)dy + t\int yg(y,x,z)dy}{f_{XZ}(x,z)+tk(x,z)} - \frac{\int yf_{YXZ}(y,x,z)dy}{f_{XZ}(x,z)} \\
&= \frac{t\left(\left[\int yg(y,x,z)dy\right]f_{XZ}(x,z) - k(x,z)\int yf_{YXZ}(y,x,z)dy\right)}{f_{XZ}(x,z)\left[f_{XZ}(x,z)+tk(x,z)\right]},
\end{aligned}$$

for $t \in [0,1]$ and appropriately defined functions

$$g(x,z) = \widehat{f_{YXZ}}(y,x,z) - f_{YXZ}(y,x,z)$$

and

$$k(x,z) = \widehat{f_{XZ}}(x,z) - f_{XZ}(x,z),$$

Obviously, $\varphi(0,x,z)$. Moreover,

$$\begin{aligned}
\frac{\partial\varphi(t,x,z)}{\partial t} &= \frac{\int yg(y,x,z)dy\,(f_{XZ}(x,z)+tk(x,z))}{(f_{XZ}(x,z)+tk(x,z))^2} \\
&\quad - \frac{k(x,z)\int yf_{YXZ}(y,x,z)dy + t\int yg(y,x,z)dy}{(f_{XZ}(x,z)+tk(x,z))^2} \\
&= \frac{f_{XZ}(x,z)\int yg(y,x,z)dy - k(x,z)\int yf_{YXZ}(y,x,z)dy}{(f_{XZ}(x,z)+tk(x,z))^2},
\end{aligned}$$

$$\frac{\partial^2\varphi(t,x,z)}{\partial t^2} = -2\frac{\left\{f_{XZ}(x,z)\int yg(y,x,z)dy - k(x,z)\int yf_{YXZ}(y,x,z)dy\right\}k(x,z)}{(f_{XZ}(x,z)+tk(x,z))^3}.$$

Next, define

$$\Psi(t) = \int\int\left(\frac{x-x_0}{h}\right)^s K\left(\frac{x-x_0}{h},\frac{z-z_0}{h}\right)\varphi(t,x,z)^2 dF_{XZ},\ s=0,1.$$

where F_{XZ} is the joint density of X_i and Z_i. This implies that

$$\Psi'(t) = \int\int\left(\frac{x-x_0}{h}\right)^s K\left(\frac{x-x_0}{h},\frac{z-z_0}{h}\right)2\varphi(t,x,z)\frac{\partial\varphi(t,x,z)}{\partial t}dF_{XZ},\ s=0,1,$$

and

$$\Psi''(t) = 2\int\int\left(\frac{x-x_0}{h}\right)^s K\left(\frac{x-x_0}{h},\frac{z-z_0}{h}\right) \times\left[\varphi(t,x,z)\frac{\partial^2\varphi(t,x,z)}{\partial t^2}+\left(\frac{\partial\varphi(t,x,z)}{\partial t}\right)^2\right]dF_{XZ},$$

for $s = 0, 1$. Note that $\Psi(0) = \Psi'(0) = 0$ due to $\varphi(0, x, z) = 0$. Obviously,

$$D_n = \Psi(1).$$

Then, by a Taylor-approximation of Ψ around $t = 0$, we have

$$\Psi(t) = \Psi(0) + \Psi'(0)t + \frac{1}{2}\Psi''(\vartheta(t))t^2,$$

where $0 \le \vartheta(t) \le t$. Hence,

$$D_n = \int\int\left(\frac{x-x_0}{h}\right)^s K\left(\frac{x-x_0}{h},\frac{z-z_0}{h}\right)\varphi(\vartheta(t),x,z)\frac{\partial^2\varphi(\vartheta(t),x,z)}{\partial t^2}dF_{XZ} + \int\int\left(\frac{x-x_0}{h}\right)^s K\left(\frac{x-x_0}{h},\frac{z-z_0}{h}\right)\left(\frac{\partial\varphi(\vartheta(t),x,z)}{\partial t}\right)^2 dF_{XZ},$$

for $s = 0, 1$. Consider the behavior of the second term first

$$\begin{aligned}
&\int\int\left(\frac{x-x_0}{h}\right)^s K\left(\frac{x-x_0}{h},\frac{z-z_0}{h}\right)\left(\frac{\partial\varphi(\vartheta(t),x,z)}{\partial t}\right)^2 dF_{XZ}\\
=&\int\int\left(\frac{x-x_0}{h}\right)^s K\left(\frac{x-x_0}{h},\frac{z-z_0}{h}\right)\\
&\times\frac{\left(f_{XZ}(x,z)\int yg(y,x,z)dy - k(x,z)\int yf_{YXZ}(y,x,z)dy\right)^2}{\left[f_{XZ}(x,z)+\vartheta(t)k(x,z)\right]^4}dF_{XZ}\\
=&\int\int\left(\frac{x-x_0}{h}\right)^s K\left(\frac{x-x_0}{h},\frac{z-z_0}{h}\right)\frac{\left(f_{XZ}(x,z)\int yg(y,x,z)dy\right)^2}{\left[f_{XZ}(x,z)+\vartheta(t)k(x,z)\right]^4}dF_{XZ}\\
&-2\int\int\left(\frac{x-x_0}{h}\right)^s K\left(\frac{x-x_0}{h},\frac{z-z_0}{h}\right)\\
&\times\frac{\left(f_{XZ}(x,z)\int yg(y,x,z)dy\right)k(x,z)\int yf_{YXZ}(y,x,z)dy}{\left[f_{XZ}(x,z)+\vartheta(t)k(x,z)\right]^4}dF_{XZ}\\
&+\int\int\left(\frac{x-x_0}{h}\right)^s K\left(\frac{x-x_0}{h},\frac{z-z_0}{h}\right)\frac{\left(k(x,z)\int yf_{YXZ}(y,x,z)dy\right)^2}{\left[f_{XZ}(x,z)+\vartheta(t)k(x,z)\right]^4}dF_{XZ}.
\end{aligned}$$

All of these three terms are of the same structure. We take the first one as example, the others follow by similar arguments. Turning, to the first term,

$$\int\int\left(\frac{x-x_0}{h}\right)^s K\left(\frac{x-x_0}{h},\frac{z-z_0}{h}\right)\frac{\left(f_{XZ}(x,z)\int yg(y,x,z)dy\right)^2}{\left[f_{XZ}(x,z)+\vartheta(t)k(x,z)\right]^4}dF_{XZ},$$

and bounding first the denominator

$$\frac{1}{|f_{XZ}(x_0,z)+\vartheta(t)k(x_0,z)|}\leq\frac{1}{|f_{XZ}(x_0,z)|-|k(x_0,z)|}\leq\frac{2}{b},$$

since $\vartheta(t)\in[0,1]$, $|f_{XZ}(x,z)|\geq b$, since we assume continuously distributed RV with compact support. Moreover, $|k(x,z)|\leq b/2$ with probability approaching one, if $\hat{f}_{XZ}(x,z)$ consistent. Hence, for $s=0,1$,

$$\begin{aligned}
&\int\int\left(\frac{x-x_0}{h}\right)^s K\left(\frac{x-x_0}{h},\frac{z-z_0}{h}\right)\frac{\left[\int yg(y,x,z)dy\right]^2}{\left[f_{XZ}(x,z)+\vartheta(t)k(x,z)\right]^4}\left(f_{XZ}(x,z)\right)^3dxdz\\
\leq\ & c_1\int\int\left|\left(\frac{x-x_0}{h}\right)^s\right|K\left(\frac{x-x_0}{h}\right)K\left(\frac{z-z_0}{h}\right)\left[\int yg(y,x,z)dy\right]^2 f_{XZ}(x,z)^3dxdz\\
=\ & h^d c_1\int\int|\psi|\,K(\psi)\,K(\zeta)\left[\int yg(y,\psi h+x_0,\zeta h+z_0)dy\right]^2\\
&\times\left(f_{XZ}(\psi h+x_0,\zeta h+z_0)\right)^3d\psi d\zeta,
\end{aligned}$$

where c_1, is a constant, and the last equality is by change of variables. Taking the supremum over K and f_{XZ}, the last rhs can be bounded by

$$\begin{aligned}
&h^d c_2\left[\int yg(y,x_0,z_0)dy\right]^2\int|\psi|\,d\psi\\
&+h^{d+2}c_3\int|\psi|\,\psi\partial_x\int yg(y,x_r,z_r)dy\int yg(y,x_r,z_r)dyd\psi\\
&+h^{d+2}c_4\int|\psi|\,\zeta\left(D_z\int yg(y,x_r,z_r)dy\right)'\iota\int yg(y,x_r,z_r)dyd\psi.
\end{aligned}$$

Since we assumed compact support for X, $\int|\psi|\,d\psi=c_5<\infty$. Defining the seminorm $\|g\|$ as

$$\max\left\{\sup_{x,z}\left|\int yg(x,y,z)dy\right|,\sup_{x,z}\left|\partial_x\int yg(x,y,z)dy\right|,\sup_{x,z}\left|\partial_{z_j}\int yg(x,y,z)dy\right|,\ \forall j\right\},$$

we obtain that, for $s = 0, 1$,

$$\int\int \left(\frac{x - x_0}{h}\right)^s K\left(\frac{x - x_0}{h}, \frac{z - z_0}{h}\right) \frac{\left(f_{XZ}(x, z) \int yg(y, x, z)dy\right)^2}{\left[f_{XZ}(x, z) + \vartheta(t)k(x, z)\right]^4} dF_{XZ} = h^d O_p\left(\|g\|^2\right).$$

By closer inspection it becomes obvious that every term in (2) has a squared error element, and hence exhibits the same behavior. Thus, for $s = 0, 1$,

$$\begin{aligned} & \int\int \left(\frac{x - x_0}{h}\right)^s K\left(\frac{x - x_0}{h}, \frac{z - z_0}{h}\right) \left(\frac{\partial\varphi(\vartheta(t), x, z)}{\partial t}\right)^2 dF_{XZ} \\ = \; & h^d O_p\left(\|g\|^2\right), \end{aligned}$$

meaning that the behavior of the second element of D_n is clarified. The first and third terms in (1) are, by similar arguments, actually $h^d O_p\left(\|g\|^3\right)$, so that we conclude that $\sqrt{\frac{n}{h^d}} D_n = \sqrt{nh^d} O_p\left(\|g\|^2\right)$. Using standard results, e.g. in Haerdle (1990), we obtain that $\|g\| = O_p(H_1^r + n^{-1/2} H_1^{-d/2} \ln(n))$, where H_1 is a first step bandwidth, the statement follows if we set $H_1 = O(n^{-1/(d+2r)})$ and $h = O(n^{-1/(d+4)})$.

We will show now that $\hat{D}_n - D_n = h^d o\left(\|g\|^2\right)$, and without loss of generality we consider only the case of a scalar z. Note that since Γ is linear in F,

$$\begin{aligned} \hat{D}_n - D_n &= \Gamma\left(\hat{\mu}, \hat{F}\right) - \Gamma\left(\hat{\mu}, F\right) \\ &= \Gamma\left(\hat{\mu}, \hat{F} - F\right). \end{aligned}$$

Therefore, the same expansions as above may be used, with $\hat{F} - F$ in place of F. In particular, $\Psi(0) = \Psi'(0) = 0$, and hence in the remainder term we are left with

$$\begin{aligned} \hat{D}_n - D_n = & \int\int \left(\frac{x - x_0}{h}\right)^s K\left(\frac{x - x_0}{h}, \frac{z - z_0}{h}\right) \\ & \times\varphi(\vartheta(t), x, z) \frac{\partial^2\varphi(\vartheta(t), x_0, z)}{\partial t^2} d\left(\hat{F}_{XZ} - F_{XZ}\right) \\ & + \int\int \left(\frac{x - x_0}{h}\right)^s K\left(\frac{x - x_0}{h}, \frac{z - z_0}{h}\right) \\ & \times \left(\frac{\partial\varphi(\vartheta(t), x, z)}{\partial t}\right)^2 d\left(\hat{F}_{XZ} - F_{XZ}\right), \end{aligned}$$

we are just left with another error in our expression. As a next step, pick again a typical element. Being more explicit about the boundaries (recall that we have compact support)

$$\int_{\underline{z}}^{\overline{z}} \int_{\underline{x}}^{\overline{x}} \left(\frac{x-x_0}{h}\right)^s K\left(\frac{x-x_0}{h}, \frac{z-z_0}{h}\right) \times \frac{\left(f_{XZ}(x,z)\int yg(y,x,z)dy\right)^2}{\left[f_{XZ}(x,z)+\vartheta(t)k(x,z)\right]^4} d\left(\hat{F}_{XZ}(x,z)-F_{XZ}(x,z)\right).$$

Let

$$b_n(z,x,t) = \left(\frac{x-x_0}{h}\right)^s K\left(\frac{x-x_0}{h}, \frac{z-z_0}{h}\right) \frac{\left(f_{XZ}(x,z)\int yg(y,x,z)dy\right)^2}{\left[f_{XZ}(x,z)+\vartheta(t)k(x,z)\right]^4}.$$

Integration by parts yields,

$$\left[b_n(z,x,t)\left(\hat{F}_{XZ}(x,z)-F_{XZ}(x,z)\right)\right]_{x=\underline{x},z=\underline{z}}^{x=\overline{x},z=\overline{z}} - \int_{\underline{z}}^{\overline{z}} \int_{\underline{x}}^{\overline{x}} \left(\hat{F}_{XZ}(x,z)-F_{XZ}(x,z)\right) \partial_x\partial_z b_n(z,x,t)dxdz.$$

Turning to the first term, this equals

$$\begin{aligned}
& b_n(\overline{z},\overline{x},t)\left(\hat{F}_{XZ}(\overline{x},\overline{z})-F_{XZ}(\overline{x},\overline{z})\right) \\
& -b_n(\underline{z},\overline{x},t)\left(\hat{F}_{XZ}(\overline{x},\underline{z})-F_{XZ}(\overline{x},\underline{z})\right) \\
& -b_n(\overline{z},\underline{x},t)\left(\hat{F}_{XZ}(\underline{x},\overline{z})-F_{XZ}(\underline{x},\overline{z})\right) \\
& +b_n(\underline{z},\underline{x},t)\left(\hat{F}_{XZ}(\underline{x},\underline{z})-F_{XZ}(\underline{x},\underline{z})\right).
\end{aligned}$$

Each of these four expressions has the same structure. Since

$$\left(\hat{F}_{XZ}(\underline{x},\overline{z})-F_{XZ}(\underline{x},\overline{z})\right) = O_p\left(n^{-1/2}\right)$$

by Glivenko-Cantelli, we have

$$b_n(\overline{z},\overline{x},t)\left(\hat{F}_{XZ}(\underline{x},\overline{z})-F_{XZ}(\underline{x},\overline{z})\right) = O_p\left(\left\|g^2\right\|\right) O_p\left(n^{-1/2}\right),$$

and the same is true for all other terms. Hence

$$\left[b_n(z,x,t)\left(\hat{F}_{XZ}(x,z)-F_{XZ}(x,z)\right)\right]_{x=\underline{x},z=\underline{z}}^{x=\overline{x},z=\overline{z}}=O_p\left(\left\|g^2\right\|\right)O_p\left(n^{-1/2}\right).$$

Now turn to

$$\begin{aligned}
&\int_{\underline{z}}^{\overline{z}}\int_{\underline{x}}^{\overline{x}}\left(\hat{F}_{XZ}(x,z)-F_{XZ}(x,z)\right)\partial_x\partial_z b_n(z,x,t)dxdz\\
&=\int_{\underline{z}}^{\overline{z}}\int_{\underline{x}}^{\overline{x}}K\left(\frac{x-x_0}{h},\frac{z-z_0}{h}\right)\frac{\left(f_{XZ}(x,z)\int yg(y,x,z)dy\right)^2}{\left[f_{XZ}(x,z)+\vartheta(t)k(x,z)\right]^4}\\
&\times\left(\hat{F}_{XZ}(x,z)-F_{XZ}(x,z)\right)dxdz\\
&+\int_{\underline{z}}^{\overline{z}}\int_{\underline{x}}^{\overline{x}}\partial_x\partial_z K\left(\frac{x-x_0}{h},\frac{z-z_0}{h}\right)\left(\frac{x-x_0}{h}\right)^s\frac{\left(f_{XZ}(x,z)\int yg(y,x,z)dy\right)^2}{\left[f_{XZ}(x,z)+\vartheta(t)k(x,z)\right]^4}\\
&\times\left(\hat{F}_{XZ}(x,z)-F_{XZ}(x,z)\right)dxdz\\
&+\int_{\underline{z}}^{\overline{z}}\int_{\underline{x}}^{\overline{x}}K\left(\frac{x-x_0}{h},\frac{z-z_0}{h}\right)\left(\frac{x-x_0}{h}\right)^s\partial_x\partial_z\frac{\left(f_{XZ}(x,z)\int yg(y,x,z)dy\right)^2}{\left[f_{XZ}(x,z)+\vartheta(t)k(x,z)\right]^4}\\
&\times\left(\hat{F}_{XZ}(x,z)-F_{XZ}(x,z)\right)dxdz
\end{aligned}\tag{AA.2}$$

It is tedious but straightforward to show that more or less the same arguments that were used in bounding D_n may be used again. the only major modification concerns the last term, where we need the altered seminorm

$$\|g\|_*=\max\left\{\sup_{x,z}\left|\partial_x\partial_z\int yg(x,y,z)dy\right|,\sup_{x,z}\left|\int yg(x,y,z)dy\right|\right\}.$$

Then, all expressions in (AA.2) are $O_p\left(\|g\|_*^2\right)O_p\left(n^{-1/2}\right)$.

This establishes that the second part of the second bias term,

$$\frac{1}{\sqrt{nh^d}}\sum_{i=1}^n K_i\left(X_i-x_0\right)\left(\hat{\mu}(X_i)-\mu(X_i)\right)^2$$

converges to zero much more rapid then the first part of this bias term. We give now only a sketch why the same is true for the first expression of the second term, i.e.

$$\frac{2}{\sqrt{nh^d}}\sum_{i=1}^{n} K_i U_i \left(X_i - x_0\right)\left(\hat{\mu}(X_i) - \mu(X_i)\right). \qquad (1)$$

Note here that, for $s = 0, 1$,

$$D_n = \int\int\int \left(\frac{x-x_0}{h}\right)^s K\left(\frac{x-x_0}{h}, \frac{z-z_0}{h}\right) u\left[\widehat{\mu}(x,z) - \mu(x,z)\right] dF_{UXZ},$$

is zero as $\int u dF_{U|XZ} = 0$, provided we use a leave one out estimator, i.e. $\widehat{\mu}$ is not a function of u. Hence, we may directly proceed to $\hat{D}_n - D_n$. Since

$$D_n = \int\int\int \left(\frac{x-x_0}{h}\right)^s K\left(\frac{x-x_0}{h}, \frac{z-z_0}{h}\right) u\left[\widehat{\mu}(x,z) - \mu(x,z)\right] d\left(\hat{F}_{UXZ} - F_{UXZ}\right)$$

Hence we have by similar arguments that $D_n = h^d O_p\left(\|g\|_*\right) O_p\left(n^{-1/2}\right)$, and hence $\sqrt{\frac{n}{h^d}} D_n = \sqrt{h^d} O_p(H_1^r + n^{-1/2} H_1^{-(d+4)/2} \ln(n))$, where H_1 is a first step bandwidth, and if we set $H_1 = O(n^{-1/(d+2r+4)})$ and $h = O(n^{-1/(d+4)})$, $\sqrt{\frac{n}{h^d}} D_n = o_p(h^2)$ for $r \geq 2, d \geq 3$.

9.2 Changes to Proof if Pre-Estimated Regressors are included

In this section we will analyze what happens to the proof above, if pre-estimated regressors are being used. In our scenario, only the pre-estimated residuals of the L-th equation (the regression of endogenous variables on instruments) matter, namely $U_i = X_i - \mu(Q_i)$ are replaced by $U_{ni} = X_i - \hat{\mu}(Q_i)$, where $\hat{\mu}$ denotes a Nadaraya Watson pre-estimator. Hence, $U_i = U_{ni} + (\hat{\mu}_i - \mu_i)$ in an obvious notation.

Throughout this subsection, we will employ the following assumption:

Assumption A.1 In the estimation of $\hat{\mu}_i$ we use a fourth order Kernel. Also, assume that μ be four times continuously differentiable.

Finally, recall the expansion

$$
\begin{aligned}
K_{ni}(x_0,u_0) &= K\left(\frac{X_i-x_0}{h}\right)K\left(\frac{U_{ni}-u_0}{h}\right)\\
&= K\left(\frac{X_i-x_0}{h}\right)K\left(\frac{U_i-u_0}{h}\right)\\
&\quad +K\left(\frac{X_i-x_0}{h}\right)K'\left(\frac{U_i-u_0}{h}\right)\frac{\mu_i-\hat{\mu}_i}{h}\\
&\quad +\frac{1}{2}K_i\left(\frac{X_i-x_0}{h}\right)K_i''\left(\frac{U_i-u_0}{h}+\lambda\left[\frac{\mu_i-\hat{\mu}_i}{h}+\frac{U_i-u_0}{h}\right]\right)\left(\frac{\mu_i-\hat{\mu}_i}{h}\right)^2.
\end{aligned}
$$

Assumption A.2 $K\left(\frac{U_{ni}-u_0}{h}\right)$ has bounded first and second derivatives.

Changes to Lemma A.1: Let

$$
\mathbb{X}'\mathbb{X}=\begin{bmatrix}\sum_{i=1}^n K_{ni}\left(X_i-x_0\right)\left(X_i-x_0\right)'/h^2 & \sum_{i=1}^n K_{ni}\left(X_i-x_0\right)\left(U_{ni}-u_0\right)/h^2\\ \sum_{i=1}^n K_{ni}\left(U_{ni}-u_0\right)\left(X_i-x_0\right)'/h^2 & \sum_{i=1}^n K_{ni}\left(U_{ni}-u_0\right)^2/h^2\end{bmatrix},
$$

where $K_{ni}=K_{ni}(x_0,u_0)=K\left(\frac{X_i-x_0}{h}\right)K\left(\frac{U_i-u_0}{h}\right)$. As noted in the heuristic in appendix A1, this leads to each element in this matrix being the sum of two types of expressions. For instance,

$$
\begin{aligned}
&\sum_{i=1}^n K_{ni}(x_0,u_0)\left(U_{ni}-u_0\right)/h\\
=&\sum_{i=1}^n K\left(\frac{X_i-x_0}{h}\right)K\left(\frac{U_i-u_0}{h}\right)\left(U_i-u_0\right)/h\\
&+\sum_{i=1}^n K\left(\frac{X_i-x_0}{h}\right)K\left(\frac{U_i-u_0}{h}\right)\frac{U_i-u_0}{h}\frac{\mu_i-\hat{\mu}_i}{h}\\
&+\sum_{i=1}^n K\left(\frac{X_i-x_0}{h}\right)K'\left(\frac{U_i-u_0}{h}\right)\frac{U_i-u_0}{h}\frac{\mu_i-\hat{\mu}_i}{h}+...
\end{aligned}
$$

The leading term in this sum was already treated. Any other term in this sum, and indeed in $\mathbb{X}'\mathbb{X}$, is of the form

$$
\hat{M}_{n,g}=\sum_{i=1}^n K\left(\frac{X_i-x_0}{h}\right)K^{(a)}\left(\frac{U_i-u_0}{h}\right)\left(\frac{U_i-u_0}{h}\right)^b\left(\frac{X_i-x_0}{h}\right)^c\left(\frac{\mu_i-\hat{\mu}_i}{h}\right)^g,
$$

where $a,b,c\in\{0,1,2\}$ and $g\in\{1,2,3,4\}$. To treat this expression, take first the terms involving differences between $\mu_i-\hat{\mu}_i$ of quadratic order. Then it is obvious that $\left(nh^d\right)^{-1}\hat{M}_{n,g} = h^{-d}h^{-2}\hat{D}_n$, where $\hat{D}_n$ is as in $L.A.3$ above, and exactly the same arguments apply. Terms of higher order in $\left((\mu_i - \hat{\mu}_i)/h\right)$ are actually more benign, because $\|\mu_i - \hat{\mu}_i\| = o_p(h)$. In contrast, more problematic are terms that are linear in $\left((\mu_i - \hat{\mu}_i)/h\right)$.

Nevertheless, we can treat this object by the same arguments as in $L.A.3$. Then we obtain that

$$\begin{aligned}\left(nh^d\right)^{-1}\hat{M}_{n,1} &= h^{-d}h^{d-2}O_p\left(\|g\|\right)\\ &= h^{-2}O_p(H_1^r + n^{-1/2}H_1^{-d/2}\ln(n))\\ &= o_p(1),\end{aligned}$$

if $H_1 = O(n^{-1/(d+2r)})$, $h = O(n^{-1/(d+4)})$ and $r > 2$. Hence, **Lemma A.1** continues to hold with, say, a fourth order Kernel.

Changes to Lemma A.2: Consider again the typical expression, and the worst case, i.e. $\left((\mu_i - \hat{\mu}_i)/h\right)$ enters linearly. The question becomes, how

$$\begin{aligned}\hat{C}_{jn} &= \frac{1}{\sqrt{nh^d}}\frac{h^2}{2}\sum_{i=1}^{n}\left(Q_i^j\right)^2 K_{Xi}K'_{Ui}(X_i - x_0)\partial^2_{x^j}m\left(x_0 + \eta(x_0)hQ_i\right)\left((\mu_i - \hat{\mu}_i)/h\right)\\ &= \frac{1}{\sqrt{nh^d}}\frac{h}{2}\sum_{i=1}^{n}\left(Q_i^j\right)^2 K_{Xi}K'_{Ui}(X_i - x_0)\partial^2_{x^j}m\left(x_0 + \eta(x_0)hQ_i\right)\left(\mu_i - \hat{\mu}_i\right),\end{aligned}$$

where $K_{Xi}K'_{Ui} = K\left(\frac{X_i - x_0}{h}\right)K'\left(\frac{U_i - u_0}{h}\right)$. Taking again the first element of $(X_i - x_0)$, i.e. the constant, we obtain by now familiar arguments that $\hat{C}_{jn}$ behaves like $\sqrt{nh^{d+2}}O_p(H_1^r + n^{-1/2}H_1^{-d/2}\ln(n)) = o_p(1)$ if $H_1 = O(n^{-1/(d+2r)})$, $h = O(n^{-1/(d+4)})$, and $r \geq 2$.

Changes to Lemma A.3: The results continue to hold, because a typical expression under the sum is multiplied by powers of $\left((\mu_i - \hat{\mu}_i)/h\right)$, which are all $o_p(1)$.

Changes to Lemma A.4: Finally, we have to determine, under what conditions

$$\left(\sqrt{nh^d}\right)^{-1}\mathbb{B}_3 = \left(\sqrt{nh^d}\right)^{-1}\left[\left(I_2 \otimes \mathbb{X}'\right)\mathbb{U} - \left[\left(I_2 \otimes \mathbb{X}'\right)\mathbb{U}\right]^*\right],$$

as defined in appendix 1, will tend to zero in probability. A typiacl element in this expression is

$$\hat{J}_{n,g} = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{X_i - x_0}{h}\right) K^{(a)}\left(\frac{U_i - u_0}{h}\right)\sigma^l(X_i)\zeta(\varepsilon_i) \times \left(\frac{U_i - u_0}{h}\right)^b \left(\frac{X_i - x_0}{h}\right)^c \left(\frac{\mu_i - \hat{\mu}_i}{h}\right)^g$$

$a, b, c \in \{0, 1, 2\}$, $l \in \{1, 2\}$ $g \in \{1, 2, 3, 4\}$, $\zeta(x) = x$ or $x^2 - 1$, and $\mathbb{E}[\zeta(\varepsilon_i)|X_i, U_i] = 0$. Consider again the linear in $((\mu_i - \hat{\mu}_i)/h)$, and $a = 1, b = c = 0$. Using $\int \varepsilon dF_{\varepsilon|X,U} = 0$, we have by similar arguments as were used for the second term in $L.A.3$, that $\hat{J}_{n,1} = h^{d-1}O_p\left(n^{-1/2}\right)O_p\left(\|g\|_*\right)$. Then,

$$\sqrt{\frac{n}{h^d}}\hat{J}_{n,g} = \sqrt{h^{d-2}}O_p(H_1^r + n^{-1/2}H_1^{-(d+4)/2}\ln(n)).$$

Choosing again optimal bandwidths, i.e. $H_1 = O(n^{-1/(d+2r+4)})$, $h = O(n^{-1/(d+4)})$, for $r = 4$, and $d \geq 3$ we have $\sqrt{\frac{n}{h^d}}\hat{J}_{n,g} = o_p(h^2)$. If $d = 2$, we have that $\sqrt{\frac{n}{h^d}}\hat{J}_{n,g} = o_p(h)$, for $r = 4$, but this produces a new leading bias term.

References

[1] AIT-SAHALIA, Y., BICKEL, P. and T.M.STOKER,(2001); Goodness-of-Fit Test for Kernel Regression with an Application to Option Implied Volatility, *Journal of Econometrics*, 105, 363-412.

[2] AFRIAT, S., (1967); The Construction of Utility Functions from Expenditure Data, *International Economic Review*, 8, 67-77.

[3] BANKS, J., BLUNDELL, R. and A. LEWBEL, (1997); Quadratic Engel Curves and Consumer Demand, *Review of Economics and Statistics,* 527-539.

[4] BROWN, R. and M.B. WALKER, (1989); The Random Utility Hypothesis and Inference in Demand Systems, *Econometrica*, 57, 815-829.

[5] BLUNDELL, R., BROWNING, M. and I. CRAWFORD, Nonparametric Engel Curves and Revealed Preferences, IFS working paper, 2002.

[6] BLUNDELL, R., PASHARDES, P. and G. WEBER, (1993); What Do We Learn About Consumer Demand Patterns from Micro Data?, *American economic Review,* 83, 570-597.

[7] CHIAPPORI, P.A. and J.C. ROCHET, (1987), Revealed Preferences and Differentiable Demand: Notes and Comments, *Econometrica*, 55, 687-691.

[8] CHRISTOPEIT, N. and S.G.N. HODERLEIN (2002); Estimation of Models with Additive Structure via Local Quasi-differencing, Working Paper, Mannheim University Working Paper.

[9] DEATON, A., and J. MUELLBAUER, (1980); An almost ideal Demand System, *American economic Review,* 70, 312-26.

[10] HODERLEIN, S., (2002), Econometric Modelling of Heterogeneous Consumer Behavior - Theory, Empirical Evidence and Aggregate Implications, *PhD Thesis*, *LSE*.

[11] HAUSMAN, J. (1981); Exact Consumer Surplus and Deadweight Loss, American Econimc Review, 71, 662-76.

[12] HAUSMAN, J. and W. NEWEY (1995); Nonparametric Measurement of Consumer's Exact Surplus and Deadweight Loss, *Econometrica*, 63, 1145-76.

[13] HODERLEIN, S., (2004); Nonparametric Demand and Endogeneity, unpublished manuscript, Mannheim University.

[14] HODERLEIN, S., HAAG, B. and K. PENDAKUR (2004); Testing and Imposing Economic Restrictions in Nonparametric Demand Systems, Working Paper, Mannheim University.

[15] IMBENS, G. and W. NEWEY, (2003), Identification and Estimation of Triangular Simultaneous Equations

Models without Additivity, MIT Working Paper.

[16] JERISON, M., (1984); Aggregation and Pairwise Aggregation of Demand when the Distribution of Income is fixed, *Journal of economic Theory*, 33, 1-31.

[17] JORGENSEN, D., LAU, L. and T. STOKER, (1982); The Transcedental Logarithmic Model of Individual Behavior. In: BASMAN, R. and G. RHODES (Eds.), Advances in Econometrics, Vol 1. JAI Press.

[18] KIHLSTROM, R.; A. MAS-COLELL and H. SONNENSCHEIN (1976) The Demand Theory of the Weak Axiom of Revealed Preference. *Econometrica*, 44, 971-978.

[19] KIM and RICHTER, (1986); Nontransitive-Nontotal Consumer Theory, *Journal of economic Theory*, 38, 324-363,

[20] LEWBEL, A., (1990); Income Distribution Movements and Aggregate Money Illusion, *Journal of Econometrics,* 43, 35-42.

[21] LEWBEL, A., (1999); Consumer Demand Systems and Household Expenditure, in PESARAN, H. and M. WICKENS (Eds.), Handbook of Applied Econometrics, Blackwell Handbooks in economics.

[22] LEWBEL, A., (2001); Demand Systems With and Without Errors, *American Economic Review*, 611-18.

[23] LUENBERGER, D.G., (1997); Optimization by Vector Space Methods, Wiley&Son, New York.

[24] MATZKIN, R. (2003); Nonparametric Estimation of Nonadditive Random Functions, *Econometrica*, 1339-1376

[25] QUAH, J. (2000), Weak Axiomatic Demand Theory. Working Paper, Nuffield College, Oxford.

[26] STONE, J.R.N., (1954); Linear Expenditure Systems and Demand Analysis, *economic Journal*, 64, 511-24.

[27] VARIAN, H., (1982); The Nonparametric Approach to Demand Analysis; *Econometrica*, 50, 945-974.