# How Many Consumers are Rational?

Stefan Hoderlein*

Mannheim University

July 27, 2007

**Abstract**

Rationality places strong restrictions on individual consumer behavior. This paper is concerned with assessing the validity of the integrability constraints imposed by standard utility maximization, arising in classical consumer demand analysis. 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, homogeneity of degree zero is being considered. Our approach employs nonseparable models and is centered around a conditional independence assumption, which is sufficiently general to allow for endogenous regressors. It is the only substantial assumption a researcher has to specify in this model. Most of the results follow from this assumption under regularity conditions. Finally, we apply all concepts to British household data, and show that rationality is an acceptable description for very large parts of the population.

**Keywords:** Nonparametric, Integrability, Testing Rationality, Nonseparable Models, Demand, Nonparametric IV.

## 1 Introduction

Economic theory yields strong implications for the actual behavior of individuals. In the standard utility maximization model for instance, economic theory places significant restrictions

*Mannheim University, Department of Economics, Institute for Econometrics, L7, 3-5, 68131 Mannheim, Germany, email: stefan_hoderlein@yahoo.com. We have received helpful comments from Richard Blundell, Andrew Chesher, Joel Horowitz, Arthur Lewbel, Simon Lee, Enno Mammen, Rosa Matzkin, Whitney Newey and by seminar participants at Berkeley, Brown, ESEM, Copenhagen, Madrid, Mannheim, Northwestern, LSE, Stanford, Tilburg, UCL and Yale. The usual disclaimer applies. I am also indebted to Sonya Mihaleva and David Hohlfeldt for excellent research assistance. Financial support by Landesstiftung Baden-Württemberg (Eliteförderungsprogramm) is gratefully acknowledged.

on individual responses to changes in prices and wealth, the so-called integrability constraints. However, when we want to evaluate the validity of the integrability conditions using real data, we face the following problem: In a typical consumer data set we observe individuals only under one or a very limited number of different price-wealth combinations, often only once. Consequently, the observations of "comparable" individuals have to be taken into consideration. But this conflicts with the important notion of unobserved heterogeneity, a notion that is supported by the widespread empirical finding that individuals with the same covariates vary dramatically in their actual consumption behavior. Thus, general and unrestrictive models for handling unobserved heterogeneity are essential for testing integrability.

Endogeneity is another major source of concern in applied work. In consumer demand, total expenditure is taken as income concept, which is justified by assuming intertemporal separability of preferences. Since the categories of goods considered are broad (e.g., food) and they frequently consitute a large part of total expenditure, the latter is believed to be endogenous. As instrument, the demand literature usually employs labor income whose determinants are thought to be exogenous to the unobserved preferences determining, say, food consumption. Other endogeneities that could arise are in particular related to prices. In empirical industrial organization for instance, prices for indivdiual goods are thought to be set by the firm targeting (partially unobserved) characteristics of individual consumers. The arising endogeneities could be tackled in exactly the same fashion as we propose in this paper. In our application, however, we consider broad aggregates of goods and we expect such effects to wash out. Moreover, following the recent Microeconometric demand literature we consider atomic individuals that act as price takers, and we control for time effects. Summarizing, we concentrate on total expenditure endogeneity, but add that this approach could easily be extended.

Testing integrability constraints dates back at least to the early work of Stone (1954), and has spurned the extensive research on (parametric) flexible functional forms demand systems (e.g., the Translog, Jorgenson et al. (1982), and the Almost Ideal, Deaton and Muellbauer (1980)). Nonparametric analysis of some derivative constraints was performed by Stoker (1989) and Härdle, Hildenbrand and Jerison (1991), but none of these has its focus on modeling unobserved heterogeneity. More closely related to our approach is Lewbel (2001) who analyzes integrability constraints in a purely exogenous setting. In comparison to his work, we make three contributions: First, we show how to handle endogeneity. Second, even restricted to the exogenous case, some of our results (e.g., on negative semidefiniteness) are new and more general. Third, we propose, discuss and implement nonparametric test statistics. An alternative method for checking some integrability constraints is revealed preference analysis, see Blundell, Browning and Crawford (2003), and references therein.

In this paper we extend the recent work on nonseparable models - in particular Imbens and

Newey (2003) and Altonji and Matzkin (2005) who treat the estimation of average structural derivatives when regressors are endogenous - to testing integrability conditions. Even though our application comes from traditional consumer demand, most of the analysis is by no means confined to this setup and could be applied to any standard utility maximization problem with linear budget constraints. Moreover, the spirit of the analysis could be extended to cover, e.g., decisions under uncertainty or nonlinear budget constraints.

Central to this paper is a conditional independence assumption. This assumption will be the only significant restriction we place on preference heterogeneity, and it explicitly allows for endogenous regressors. The main contribution of this paper is the formal clarification of the implications of this conditional independence assumption for testing integrability constraints with data. Much of the second section will be specifically devoted to this issue: we devise very general tests for Slutsky negative semidefiniteness and symmetry, as well as for homogeneity. In the third section we apply these concepts to british FES data. The results are affirmative as far as the validity of the integrability conditions are concerned and demonstrate the advantages of our framework. A summary and an outlook conclude this paper, while the appendix contains proofs, graphs and summary statistics. An overview of the econometric methods may be obtained from a supplement that is available on the author's webpage.

# 2 The Demand Behavior of a Heterogeneous Population

## 2.1 Structure of the Model and Assumptions

Our model of consumer demand in a heterogeneous population consists of several building blocks. As it is common in consumer demand, we assume that - for a fixed preference ordering - there is a causal relationship between budget shares, a $[0,1]$ valued random $L$-vector denoted by $W$, and regressors of economic importance, namely log prices $P$ and log total expenditure $Y$, real valued random vectors of length $L$ and 1, respectively. Let $X = (P', Y)' \in \mathbb{R}^{L+1}$. To capture the notion that preferences vary across the population, we assume that there is a random variable $V \in \mathcal{V}$, where $\mathcal{V}$ is a Borel space[1], which denotes preferences (or more generally, decision rules). We assume that heterogeneity in preferences is partially explained by observable differences in individuals' attributes (e.g., age), which we denote by the real valued random $G$-vector $Q$. Hence, we let $V = \vartheta(Q, A)$, where $\vartheta$ is a fixed $\mathcal{V}$-valued mapping defined on the sets $\mathcal{Q} \times \mathcal{A}$ of possible values of $(Q, A)$, and where the random variable $A$ (taking again values in a Borel space $\mathcal{A}$) covers residual unobserved heterogeneity in a general fashion.

[1] Technically: $\mathcal{V}$ is a set that is homeomorphic to the Borel subset of the unit interval endowed with the Borel $\sigma$-algebra. This includes the case when $V$ is an element of a polish space, e.g., the space of random piecewise continuous utility functions.

As an example for a heterogeneous population, we choose a linear structure (in consumer demand, this corresponds approximately to an Almost Ideal demand system, with the standard shortcut of setting the denominator terms in the income effect equal to a price index). Neglect the dependence on $Q$ and assume for the moment that the outcome equation would be given by a random coefficients model,

$$W = X'A_1, \tag{2.1}$$

where $A_1$ denote random coefficients that vary across the population. In Hoderlein, Klemelä and Mammen (2007), we show that the distribution of preference parameters $f_{A_1}$ is nonparametrically identified. However, here we do not want to assume a linear heterogeneous population from the outset, and allow for more general forms of heterogeneity. For instance,a heterogeneous population consisiting of two types of individuals, one linear in coefficients like model (2.1), one with a nonlinear, but parametric, both with (finite) parameters that vary across the population, may be formalized as

$$W = \mathbf{1}\left\{A_0 > 0\right\} X'A_1 + \mathbf{1}\left\{A_0 \leq 0\right\} g(X, A_2), \tag{2.2}$$

for a known function $g$ and random vector $A = (A_0, A_1', A_2')'$ of parameters that vary across the population. In this model, $f_A$ is already not identified. More generally, there may be infinitely many types, and the parameters may be infinitely dimensional, and hence we formalize the heterogeneous population as $W = \phi(X, A)$, for a general mapping $\phi$. Still, for any fixed value of $A$, say $a_0$, we obtain a demand function having standard properties, and the hope is that when averaging over the unobserved heterogeneity $A$, rationality properties of individual demand may still be preserved by some structure.

As mentioned, a contribution of this paper is that we allow for dependence between unobserved heterogeneity $A$ and all regressors of economic interest. To this end, we introduce the real valued random $K$-vector of instruments $S$. Note that $S$ may contain exogenous elements in $X$, which serve as their own instruments.

Having defined all elements of our model, we are now in the position to state it formally, including observable covariates $Q$:

**Assumption 2.1** *Let all variables be defined as above. The formal model of consumer demand in a heterogeneous population is given by*

$$\begin{aligned} W &= \phi(X, \vartheta(Q, A)) \qquad (2.3)\\ X &= \mu(S, Q, U) \qquad (2.4) \end{aligned}$$

*where $\phi$ is a fixed $\mathbb{R}^L$-valued Borel mapping defined on the sets $\mathrm{X} \times \mathcal{V}$ of possible values of $(X, V)$. Analogously, $\mu$ is a fixed $\mathbb{R}^{L+1}$-valued Borel mapping defined on the sets $\mathcal{S} \times \mathcal{Q} \times \mathcal{U}$ of* *possible values of* $(S, Q, U)$.

Assumption 2.1 defines the nonparametric demand system with (potentially) endogenous regressors as a system of nonseparable equations. These models are calles nonseparable, because they do not impose an additive specification for the unobservable random terms (in our case $A$ or $U$). They have been subject of much interest in the recent econometrics literature (Chesher (2003), Matzkin (2003), Altonji and Matzkin (2005), Imbens and Newey (2003), Hoderlein and Mammen (2007), to mention just a few). Since we do not assume monotonicity in unobservables, our approach is more closely related to the latter three approaches.

As is demonstrated there, in the absence of strict monotonicity of $\phi$ in $A$, the function $\phi$ is not identified, however, local average structural derivatives are. Although it will be demonstrated that identification may proceed on this level of abstraction, in the case of endogeneity of $X$ this requires, however, that $U$ be solved for because these residuals have to be employed in a control function fashion. In the application, we specify $\mu$ to be the conditional mean function, and consequently $U$ to be additive mean regression residuals.

Writing $Z = (Q', U)'$, we formalize the notion of independence between instruments and unobservables as follows:

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

$$F_{A|S,Z} = F_{A|Z} \tag{2.5}$$

There are also some differentiability and regularity conditions involved, which are summarized in the appendix in assumption 2.3. However, assumption 2.2 is the key identification assumption and merits a thorough discussion: Assume for a moment all regressors were exogenous, i.e. $S \equiv X$ and $U \equiv 0$. Then this assumption states that $X$, in our case wealth and prices, and unobserved heterogeneity are independently distributed, conditional on individual attributes.

To give an example: Suppose that in order to determine the effect of wealth on consumption, we are given data on the demand of individuals, their wealth and the following attributes: "education in years" and "gender". Take now a typical subgroup of the population, e.g., females having received 12 years of education. Assume that there be two wealth classes for this subgroup, rich and poor, and two types of preferences, type 1 and 2. Then, for both rich and poor women in this subgroup, the proportion of type 1 and 2 preferences has to be identical. This assumption is of course restrictive. Note, however, that preferences and economically interesting regressors may still be correlated across the population. Moreover, any of the $Z$ may be correlated with preferences.

Now turn to the case of endogenous regressors and instruments. Suppose we were again interested in the effect of changes in wealth on the demand of an individual. In the demand literature, wealth equals total expenditure, but the latter is commonly assumed to be endogenous, and hence labor income, denoted by $S_1$, is taken as an instrument. In a world of rational agents, labor income is the result of maximizing behavior by individual agents (consumers and firms). Much as above, we can model $S_1 = \upsilon(Q, A_2)$, where $\upsilon$ is a fixed Borel-measurable scalar valued function defined on the set $\mathcal{Q} \times \mathcal{A}_2$. Here, $A_2$ is a set of unobservables containing variables that govern the decision of the individual's intertemporal optimal labor supply problem, e.g. the attitude towards risk or idiosyncratic information sets.

First, consider the extreme case where $A$ and $A_2$ are independent conditional on $Q$. In addition to independence between $U$ and $S_1$, we assume that $U$ is not a function of $S_1$, i.e. $U = \varphi(Q, A)$. Then, assumption 2.3 is quite plausible as in this case the condition $F_{A|S_1,Z} = F_{A|Z}$ i.e., $F_{A|\upsilon(Q,A_2),\varphi(Q,A),Q} = F_{A|\varphi(Q,A),Q}$, can be derived as follows:

$$F_{A|\upsilon(Q,A_2),\varphi(Q,A),Q} = \frac{F_{A,\upsilon(Q,A_2),\varphi(Q,A)|Q}}{F_{\upsilon(Q,A_2),\varphi(Q,A)|Q}} = \frac{F_{A,\varphi(Q,A)|\upsilon(Q,A_2),Q}}{F_{\varphi(Q,A)|Q}} = \frac{F_{A,\varphi(Q,A)|Q}}{F_{\varphi(Q,A)|Q}} = F_{A|\varphi(Q,A),Q}$$

where only the conditional independence between $A$ and $A_2$ has been used.

In contrast, assume that $A = A_2$, i.e. the unobservable characteristics of the household that govern the two decisions are the same. Then both $S_1$ and $U$ are functions of $A$. Nevertheless, $F_{A|S_1,Z} = F_{A|Z}$ is still not completely implausible as $U$ already reflects some influence of $A$.

## 2.2 Implications for Observable Behavior

Given these assumptions and notations, we concentrate first on the relation of theoretical quantities and the identified objects, specifically $m(\xi, z) = \mathbb{E}[W|X = \xi, Z = z]$ and $M(s, z) = \mathbb{E}[W|S = s, Z = z]$ which denote the conditional mean regression function using either endogenous regressors and controls, or instruments and controls. Finally, let $\sigma\{X\}$ denote the information set ($\sigma$-algebra) spanned by $X$, and $\Xi^-$ be the Moore-Penrose pseudo inverse of a matrix $\Xi$.

More specifically, we focus on the following questions:

1. How are the identified marginal effects (i.e., $D_x m$ or $D_x M$) related to the theoretical derivatives $D_x \phi$?

2. How and under what kind of assumptions do observable elements allow inference on key elements of economic theory? Especially, what do we learn about homogeneity, adding up, as well as negative semidefiniteness and symmetry of the Slutsky matrix, which in the standard consumer demand setup we consider (with budget shares as dependent variables, and log prices and log income as regressors), and in the underling heterogeneous population (defined by $\phi$, $x$, and $v$), takes the form

$$\mathfrak{S}(x, v) = D_p\phi(x, v) + \partial_y\phi(x, v)\phi(x, v)' + \phi(x, v)\phi(x, v)' - diag\{\phi(x, v)\}.$$

Here, $diag\{\phi\}$ denotes the matrix having the $\phi_j$, $j = 1, .., L$ on the diagonal and zero off the diagonal. The second question shall be the subject of Propositions 2.2 and 2.3. We start, however, with Lemma 2.1 which answers the question on the relationship between estimable and theoretical derivatives. All proofs may be found in the appendix.

**Lemma 2.1** *Let all the variables and functions be as defined above. Assume that* $(A2.1)$ - $(A2.3)$ *hold. Then,*

$$\begin{aligned}&(i) \quad \mathbb{E}[D_x\phi(X,V)|X,Z] = D_x m(X,Z) \quad (a.s.),\\&(ii)\ \mathbb{E}[D_x\phi(X,V)|S,Z] = D_s M(S,Z) D_s\mu(S,Z)^- \quad (a.s.).\end{aligned}$$

*Ad* $(i)$ : This result establishes that every 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. Note that it is only the local conditional average of the derivative of $\phi$ that may be identified and not the function $\phi$. Suppose we were given data on consumption, income, "education in years" and "gender" as above. Then by use of the marginal effect $D_x m(X,Z)$, we may identify the average marginal income effect on consumption of, e.g., all female high school graduates, but not the marginal effect of every single individual. Arguably, for most purposes the average effect on the female graduates is all that decision makers care about.

*Ad*$(ii)$ : This result illustrates that assumption $A2.2$ has several implications on observables depending which conditioning set is used. As the preceding discussion shows, conditioning on $X, Z$ seems natural as the subgroups formed have a direct economic interpretation. However, recall that $\sigma\{X,Z\} \subseteq \sigma\{S,Z\}$. Hence, $\sigma\{S,Z\}$ should be employed, because conditional expectations using $\sigma\{S,Z\}$ are closer in $L_2$-norm to the true derivatives. Altonji and Matzkin (2005) derive an estimator for $\mathbb{E}[D_x\phi(X,V)|X,Z]$ by integrating $(i)$ over $U$, but conditioning on $X, Z$. Since our focus is on testing economic restrictions, we avoid the integration step as in many cases it reduces the power of all test statistics. Therefore we give always results using $\sigma\{X,Z\}$ and $\sigma\{S,Z\}$. The former has a more clear cut economic interpretation, the latter yields tests of higher power.

We now turn to the question which economic properties in a 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. We introduce the following notation: Let $e_j$ denote the $j$-th unit vector of length $L+1$, let $E_j = (e_1, .., e_j, 0, .., 0)$ and $\iota$ be the vector containing only 1.

Now we are in the position to state the following proposition, which is concerned with adding up and homogeneity of degree zero, two properties which are related to the linear budget set:

**Proposition 2.2** *Let all the variables and functions be defined as above, and suppose that* $(A2.1)$ *and* $(A2.3)$ *hold. Then,* $\iota'\phi = 1$ $(a.s.) \Rightarrow \iota' m = 1$ *and* $\iota' M = 1$ $(a.s.)$*. If, in addition,* $F_{A|X,Z}(a, \xi, z) = F_{A|X,Z}(a, \xi + \lambda, z)$ *is true for all* $\xi \in X$*, then*

$$\phi(X, V) = \phi(X + \lambda, V) \;\; (a.s.) \;\; \Rightarrow \;\; m(X, Z) = m(X + \lambda, Z) \;\; (a.s.).$$

*Under* $(A2.1) - (A2.3)$ *we obtain that,*

$$\begin{aligned} \phi(X, V) &= \phi(X + \lambda, V) \; (a.s.) \Rightarrow D_p m \iota + \partial_y m = 0 \; (a.s.) \\ \text{and} \;\; \phi(X, V) &= \phi(X + \lambda, V) \; (a.s.) \Rightarrow D_s M D_s \mu^- E_L \iota + D_s M D_s \mu^- e_{L+1} = 0 \; (a.s.). \end{aligned}$$

*Finally, if we also assume* $(A2.1)$*,* $(A2.3)$*,* $\mu(S, Z) = \mu(S + \lambda, Z)$*, as well as,* $F_{A|S,Z}(a, s, z) = F_{A|S,Z}(a, s + \lambda, z)$*, then*

$$\phi(X, V) = \phi(X + \lambda, V) \;\; (a.s.) \;\; \Rightarrow \;\; M(S, Z) = M(S + \lambda, Z) \;\; (a.s.)$$

**Remark 2.2:** Note that we do not require any type of independence for adding up to carry through to the world of observables. This is very comforting since - in the absence of any direct way of testing this restriction - adding up is imposed on the regressions by deleting one variable.

The homogeneity part of Proposition 2.2 is ordered according to the severity of assumptions: Homogeneity carries through to the regression conditioning on endogenous regressors and controls under a homogeneity assumption on the cdf. This assumption is weaker than $A2.2$ as it is obviously implied by conditional independence. The derivative implications are generally true under conditional independence. This is particularly useful for testing homogeneity using the regression including instruments, as for this regression to inherit homogeneity an implausible additional homogeneity assumption, $\mu(S, Z) = \mu(S + \lambda, Z)$, has to be fulfilled.

The following proposition is concerned with the Slutsky matrix. We need again some notation. Let $\mathbb{V}[G, H|\mathcal{F}]$ denote the conditional covariance matrix between two random vectors $G$ and $H$, conditional on some $\sigma$-algebra $\mathcal{F}$, and $\mathbb{V}[H|\mathcal{F}]$ be the conditional covariance matrix of a random vector $H$. We will also make use of the second moment regressions, i.e. $m_2(\xi, z) = \mathbb{E}[WW'|X = \xi, Z = z]$ and $M_2(s, z) = \mathbb{E}[WW'|S = s, Z = z]$. Moreover, denote by $vec$ the operator that stacks a $m \times q$ matrix columnwise into a $mq \times 1$ column vector, and by $vec^{-1}$ the operation that stacks an $mq \times 1$ column vector columnwise into a $m \times q$ matrix. Finally, for any square matrix $B$, let $\overline{B} = B + B'$.

**Proposition 2.3**: *Let all the variables and functions be defined as above. Suppose that* $(A2.1)$*-*$(A2.3)$ *hold. Then, the following implications hold almost surely:*

$\mathfrak{S}$ *nsd* $\Rightarrow \overline{D_p m} + \partial_y m_2 + 2\left(m_2 - diag\left\{m\right\}\right)$ *nsd, and*

$\mathfrak{S}$ *nsd* $\Rightarrow \overline{D_s M D_s \mu^- E_L} + vec^{-1}\left\{D_s vec\left[M_2\right] D_s \mu^- e_{L+1}\right\} + 2\left(M_2 - diag\left\{M\right\}\right)$ *nsd*

*However, if and only if* $\mathbb{V}\left[\partial_y \phi, \phi | X, Z\right]$, *respectively* $\mathbb{V}\left[\partial_y \phi, \phi | S, Z\right]$, *are symmetric we have*

$\mathfrak{S}$ *symmetric* $\Rightarrow\ D_p m + \partial_y m m'$ *symmetric, and*

$\mathfrak{S}$ *symmetric* $\Rightarrow\ D_s M D_s \mu^- E_L + D_s M D_s \mu^- e_{L+1} M'$ *symmetric*

*almost surely.*

**Remark 2.3:** The importance of this proposition lies in the fact that it allows for testing the key elements of rationality without having to specify the functional form of the individual demand functions or their distribution in a heterogeneous population. Indeed, the only element that has to be specified correctly is a conditional independence assumption.

Suppose now we see any of these conditions rejected in the observable (generally nonparametric) regression at a position $x, z$. 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. An interesting question is when the reverse implications hold as well, i.e. one can deduce from the properties of the observable elements directly on $\mathfrak{S}$. This issue is related to the concept of completeness raised in Blundell, Chen and Kristensen (2007). It is our conjecture that some of the concepts may be transfered, but a detailed treatment is beyond the scope of this paper and will be left for future research.

Proposition 2.3 illustrates clearly that appending "an additive error capturing unobserved heterogeneity" and proceeding as if the mean regression $m$ has the properties of individual demand is not the way to solve the problem of unobserved heterogeneity. Note that we may always append a mean independent additive error, since $\phi = m + (\phi - m) = m + \varepsilon$. The crux is now that the error is generally a function of $y$ and $p$. For instance, the potentially nonsymmetric part of the Slutsky matrix becomes

$$\mathfrak{S} = D_p m + \partial_y m m' + D_p \varepsilon + \left(\partial_y m\right) \varepsilon' + \left(\partial_y \varepsilon\right) m' + \left(\partial_y \varepsilon\right) \varepsilon' \ ,$$

and the last four terms will not vanish in general.

Returning to Proposition 2.3, 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. To obtain a similar result for symmetry, we have to invoke an additional assumption about the conditional covariance matrix $\mathbb{V}\left[\partial_y \phi, \phi | X, Z\right]$. 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 assumption does not hold, we are able to test at most for homogeneity, adding up and Slutsky negative semidefiniteness. This amounts 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) and Quah (2000). Furthermore, note 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. 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 $\mathbb{V}[\partial_y \phi, \phi]$ has been used in this literature (as "increasing dispersion", see Härdle, Hildenbrand and Jerison (1989)).

# 3 Empirical Implementation

In this section we discuss all matters pertaining to the empirical implementation: We give first a brief sketch of the econometrics methods, an overview of the data, mention some issues regarding the econometric methods, and present the results.

## 3.1 Econometric Specifications and Methods

From our identification results in proposition 2.2 and 2.3, we are able to characterize testable implications on nonparametric mean regressions. It is imperative to note that these mean regressions are only the reduced form model, and not a specification of the structural model. As discussed above, the implications will depend on the independence assumption that a researcher deems realistic. In the demand literature, it is log total expenditure that is taken to be endogenous, and labor income is taken as additional instrument, see Lewbel (1999). The basic reason is preference endogeneity: since the broad aggregates of goods typically considered (e.g., "Food") explain much of total expenditure, it is suspected that the preferences that determine demand for these broad categories of goods, and those how determine total nondurable consumption are dependent. Prices, in turn are assumed to be exogenous, because unlike in IO approaches were the demands for individual goods is analyzed, and price endogeneities may be suspected on grounds that firms charge different consumers differently, the broad categories of good typically analyzed are invariant to these types of considerations. Still, one may question this exogeneity assumption, and we can only emphasize that nothing in our chain of argumentation precludes handling this endogeneity in precisely the same way than total expenditure endogeneity.

As discussed in the introduction, for purpose of recovering $U$ we have to specify the equation relating the endogenous regressors total expenditure $S$ to instruments only. In this paper, we have settled for the general nonparametric form $Y = \mu(S, Z) = \psi(S, Q) + \sigma_2(S, Q)U$, with the normalization $\mathbb{E}[U|S, Q] = 0$ and $\mathbb{V}[U|S, Q] = 1$. In supplementary material that may be downloaded from the author's webpage, we discuss issues in the implementation of estimators and test statistics in detail. Here we just mention briefly that we estimate all regression functions by local polynomials, and form pointwise nonparametric tests using these estimators. We evaluate all tests at a grid of $n = 3000$ observations that are *iid* draws from the data. To derive the asymptotic distribution of our test statistic at each of these observations, we apply bootstrap procedures. In the case of testing homogeneity, we use the fact that the hypothesis, i.e. $m(P, Y, Z) = m(P + \lambda, Y + \lambda, Z)$ may be reformulated as $m(\tilde{P}, Y, Z) = m(\tilde{P}, Z)$, where $\tilde{P} = P - Y$, i.e., the test reduces the a nonparametric system of equations omission of variables test. Consequently, constructing a bootstrap sample under the null, i.e., with homogeneity imposed, is straightforward. The limiting distribution, as well as arguments showing the consistency of the boostrap, are then straightforward, and may be related to Ait-Sahalia, Bickel and Stoker (2001).

The case of testing symmetry is more involved, as there is no easy way of constructing symmetry restricted residuals. However, by similar arguments as in Haag, Hoderlein and Pendakur (2007), it is possible to obtain the limiting distribution of the test statistic. Moreover, adaptation of their idea of deriving a bootstrap distribution under the null (by exploiting the structure of the Kernel estimator, as well as the fact that under the null the bias vanishes) to our pointwise testing problem is straightforward. This has the added benefit that the consistency of the bootstrap may be established along similar lines. Finally, a bootstrap procedure for testing negative semidefiniteness may be devised using a similar idea by Härdle and Hart (1993). Essentially, the idea is to look at the bootstrap distribution of the largest eigenvalue. This procedure is consistent, provided there is no multiplicity of eigenvalues, which is not a problem in our application. For technical details we refer again to the supplementary material that may be downloaded from the author's webpage.

## 3.2 Data

The FES reports a yearly cross section of income, expenditures, demographic composition and other characteristics of about 7,000 households. We use the years 1974-1993, but exclude the respective Christmas periods as they contain too much irregular behavior. Like the parametric literature, because of measurement error we focus on the subpopulation of two person households, both adults, at least one is working and the head of household is a white collar worker, see Lewbel (1999). We provide a summary statistic of our data in table in the appendix.

The expenditures of all goods are grouped into three categories. The first category is related to food consumption and consists of the subcategories food bought, food out (catering) and tobacco, which are self explanatory. The second category contains expenditures which are related to the house, namely housing (a more heterogeneous category; it consists of rent or mortgage payments), furniture as well as household goods and services. Finally, the last group consists of motoring and fuel expenditures, categories that are often related to energy prices. For brevity, we call these categories food, housing and energy. These broader categories are formed since more detailed accounts suffer from infrequent purchases (recall that the recording period is 14 days) and are thus often underreported. Together they account for 50-60% of total expenditure on average, leaving a forth residual category. We removed outliers by excluding the upper and lower 2.5% of the population in the three groups.

Income is constructed as in the definition of household below average income study (HBAI). It is roughly defined as net income after taxes, but including state transfers. We include the remaining household covariates as regressors. Specifically, we use principal components to reduce $Q$ to some three orthogonal approximately continuous components, mainly because we require continuous covariates for nonparametric estimation. While this has some additional advantages, it is arguably ad hoc. However, we performed some robustness checks like alternating the components or adding parametric indices to the regressions, and the results do not change in an appreciable fashion.

## 3.3 Empirical Results in Detail

Although they are not in the focus of this paper, because they are building blocks for our test statistics we display in the appendix some nonparametric estimates of the function and the derivatives. In figures 1-3 in the appendix we show the budget shares of the three categories of goods against log total expenditure. Note that the food budget share is downward sloping in income, whereas the others are weakly increasing or constant across the expenditure range. Moreover, we also show two compensated own price effects as well as one compensated cross price effect in figures 4 - 6. They show that the compensated own price effects of food and housing are negative, as predicted by theory. The cross price effect is very small in comparison, but note the rather large confidence bands around all the derivatives. Both observations are indicative that negative semidefniteness may not be rejected. All functions are plotted at the mean level of all other regressors, and obviously at other values of the regressors the picture varies. The largely negative compensated own price effects, however, remain preserved.

Turning to the test statistics, they are constructed such that the implications of economic theory are true under the null. Moreover, the tests are to be performed pointwise at the individual observations. A rejection means that the original condition, e.g., negative semidefiniteness in the underlying heterogeneous population, cannot hold in the neighborhood of an individual. Although this gives an accurate picture of the behavior on individual level, it results in a flood of results, and we aggregate across the population as a means of condensing the result.

Before discussing the results in detail, two important issues have to be clarified. The first concerns the functional form. Using the test of Haag and Hoderlein (2006), we reject the null that the regression is a Quadratic Almost Ideal. More precisely, the value of the test statistic is 7.32, while the 0.95 quantile of the bootstrap distribution is 3.65. Indeed, this finding strongly suggests using a flexible nonparametric form - otherwise rejections of economic hypotheses may occur just because we have chosen the wrong functional form.

The second issue is endogeneity: We use total expenditure as regressor, which is potentially endogenous. Employing again a Haag and Hoderlein (2006) type test, we reject the hypothesis that $U$ should be excluded from the regression with a $p$-value of 0.01. Although this strongly suggests that a control function correction for endogeneity should be adopted, below we will see that in terms of economic restrictions the results are largely comparable.

The main results are condensed in tables 4.1 to 4.3. We report the results using total expenditure under both exogeneity and endogeneity, where we handle the latter by including control function residuals. The results using the instruments plus controls regressions produce similar results, and are therefore not reported here.

Table 4.1 shows results for tests of the negative semidefiniteness hypothesis. With respect to the individual results we report both the uncompensated and compensated price effect (i.e., Slutsky) matrices. Specifically, it shows the percentage of the population for which negative semidefiniteness can not be rejected at the 0.95 percent confidence level.

| *Hypothesis* | Uncomp. PE | Comp. PE |
|---|---|---|
| Negative Semidefiniteness under Exogeneity | 0.81 | 0.69 |
| Negative Semidefiniteness under Exogeneity and Homogeneity | 0.81 | 0.82 |
| Negative Semidefiniteness under Endogeneity and Homogeneity | 0.92 | 0.93 |

Table 4.1. Percentage of the Population in Accordance with Slutsky Semidefiniteness

From these results it is obvious that negative semidefiniteness is well accepted. Note from the first row the somewhat unexpected result that the uncompensated price effect matrix is more often negative semidefinite than the compensated one. This seems to disagree with economic theory. However, a careful examination of the individual results shows that matrices that were only barely negative semidefinite get slightly perturbed by the relatively strong income effect matrix (i.e., the difference between the two matrices).

This curiosity disappears once we impose homogeneity, because it is hard to imagine circumstances under which negative semidefiniteness holds, while homogeneity does not. Imposing homogeneity, the Slutsky matrices appear to be more negative semidefinite than the uncompensated price effect matrices. Finally, the inclusion of endogeneity improves the result even further, so that we arrive at an almost complete compliance of individuals with negative semidefiniteness. Moreover, there is no obvious structure in the remaining rejections, i.e. no clustering at certain household characteristics.

With respect to homogeneity, our results look similar. As can be seen from table 4.2, it is widely accepted. The table shows in the first column again the hypothesis that is being tested, while the second and third row display the number of non rejections at the 0.90 and 0.95 confidence level.

| *Hypothesis* | 0.90 | 0.95 |
|---|---|---|
| Homogeneity under Exogeneity | 0.73 | 0.89 |
| Homogeneity under Endogeneity | 0.71 | 0.87 |

Table 4.2. Percentage of the Population in Accordance with Homogeneity

Here we see that exogeneity and endogeneity do not generate materially different results, although that of endogeneity looks slightly worse. Again, no clear pattern of violations arises.

Finally, the results for symmetry are displayed in table 4.3. As before, the percentage of non rejections at the significance levels 0.90 and 0.95 are being displayed.

| *Hypothesis* | 0.90 | 0.95 |
|---|---|---|
| Symmetry under Exogeneity | 0.93 | 0.95 |
| Symmetry under Exogeneity and Homogeneity | 0.86 | 0.92 |
| Symmetry under Endogeneity and Homogeneity | 0.91 | 0.95 |

Table 4.3. Percentage of the Population in Accordance with Slutsky Symmetry

It is interesting to note that accounting for homogeneity worsens the result, while correcting for endogeneity generally improves the results. However, the results do not differ substantially.

As the main caveat for our analysis, the following point should be emphasized: Indeed, some point estimates of the Slutsky matrix appear to be far from symmetric. The same is, perhaps to a lesser extent, also true of negative semidefiniteness. However, it is especially the cross price effects that are measured with very low precision, and hence have huge standard errors. This familiar problem of parametric demand analysis (cf. Lewbel (1999)) is aggravated by the high dimensional nonparametric approach taken here. Hence, we should stress the fact that we *merely were not able to reject these hypotheses.*

Nevertheless, it is noteworthy that overall economic theory provides an acceptable hypothesis for the population, as is indicated by results beyond the 80% mark. Hence, at least in this specific subpopulation, economic theory fares well.

# 4 Conclusion

In this paper we have established that it is really only necessary to impose one substantial restriction in order to perform most of demand analysis empirically. This is the conditional independence assumption $A2.2$. Once this assumption holds nothing else apart from regularity conditions has to be assumed to test conclusively the main elements of demand theory, in particular homogeneity and negative semidefiniteness. It is a key feature of our approach that no material assumptions about functional form or heterogeneity of the population have to be imposed.

Symmetry of the Slutsky matrix turns out to be the only major implication of rationality that will only ever be testable under additional identification assumptions. The doubts on the empirical verifiability of this property suggests that economic theory should perhaps concentrate on a model of demand that is entirely testable, such as in Kihlstrom, Mas-Colell and Sonnenschein (1976) or Quah (2000).

The conditional independence assumption plays a critical role in this paper. However, it is sufficiently general to nest a variety of scenarios including control function IV, and proxy variables, which is not detailed in this paper but straightforward. The main task of the researcher is to choose out of a set of independence assumptions the one that he believes to be most realistic on economic grounds. This paper has established that from this starting point on empirical economic analysis can proceed without any major additional restrictions.

# Appendix

## A.1 Assumption 2.3: Regularity Conditions

Assume that the demand functions $\phi$ be continuously differentiable in $x$ for all $x \in \mathcal{X} \subseteq \mathbb{R}^{L+1}$. This restricts preferences to be continuous, strictly convex and locally nonsatiated, with associated utility functions everywhere twice differentiable. Assume in addition that $\mu$ be continuously differentiable in $s$ for all $s \in \mathcal{S} \subseteq \mathbb{R}^{K+1}$, and that $D_s\mu$ has full column rank almost surely. Assume that preferences be additively separable over time, which justifies the use of total expenditure as wealth. Moreover, we confine ourselves to observationally distinct preferences, i.e. if $v_j, v_k \in \mathcal{V}$ and $v_j \neq v_k$, then there exists a set $\mathcal{X} \subseteq \mathbb{R}^{L+1}$ with $\mathbb{P}(\mathcal{X}) > 0$, such that $\forall x \in X : D_x\phi(x, v_1) \neq D_x\phi(x, v_2)$. Finally, we require the following condition for dominated convergence: there exists a function $g$, s. th. $\|vec\,[D_x\phi(x, \vartheta(q, a))]\| \leq g(a)$, with $\int g(a)F_A(da) < \infty$, uniformly in $(x, q)$.

## A.2 Proofs of Lemmata and Propositions

**Proof of Lemma 2.1:**

*Ad* $(i),(iii)$ First recall that, by definition, $0 \leq W \leq 1$. Thus, the expectation exists and $\mathbb{E}[|W|] \leq c < \infty$, where $c$ is a generic constant (the same holds for the second moment). From this it follows that all conditional expectations exist as well, and are bounded.

Next, let $x, z$ be fixed, but arbitrary.

$$\begin{aligned} D_x m(x,q,u) &= D_x \int_{\mathcal{A}} \phi(x, \vartheta(q,a)) F_{A|X,Q,U}(da, x, q, u) \\ &= D_x \int_{\mathcal{A}} \phi(x, \vartheta(q,a)) F_{A|Q,U}(da, q, u), \end{aligned}$$

due to $A2.2 \Longrightarrow F_{A|X,Q,U} = F_{A|Q,U}$. Using dominated convergence, we obtain that the rhs equals

$$\int_{\mathcal{A}} D_x \phi(x, \vartheta(q,a)) F_{A|Q,U}(da, q, u) \tag{A.1}$$

But due to $A2.2$ this is (a version of) $\mathbb{E}[D_x\phi|X = x, Z = z]$. Upon inserting random variables for the fixed $z, y$ the statement follows. For $(iii)$, by the same arguments $D_s M(s,z) = \mathbb{E}[D_x \phi|S = s, Z = z] D_s \mu(s,z)$. Postmultiplying with $D_s\mu(s,z)^-$ produces the result. *Q.E.D.*

*Ad* $(ii)$ To see the "if" part, simply note that if $V$ is $Z$-measurable,

$$\mathbb{E}[D_x \phi(X, \vartheta(Q,A))|X, Z] = D_x \phi(X, \theta(Z))$$

for some function $\theta$.

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 $\omega_j \in \Omega_0, \omega_k \in \Omega_1$, $k \neq j$, $Z(\omega_k) = Z(\omega_j)$, $X(\omega_k) = X(\omega_j)$, but $V(\omega_k) \neq V(\omega_j)$ since otherwise $V$ would be $Z$-measurable on $\Omega_0 \cup \Omega_1$ or $X$ and $V$ not independent conditional on $Z$. By the observational distinctness condition it follows that $D_x\phi(\omega_k) \neq D_x\phi(\omega_j)$, for all $\omega_j \in \Omega_0, \omega_k \in \Omega_1$, $k \neq j$, although

$$\mathbb{E}[D_x\phi(X,V)|X, Z, \{\omega \in \Omega_0\}] = \mathbb{E}[D_x\phi(X,V)|X, Z, \{\omega \in \Omega_1\}]$$

is possible. Hence, $D_x m(\omega) \neq D_x\phi(\omega)$ for some $\Omega_s \subseteq \Omega_0 \cup \Omega_1$ with $\mathbb{P}[\Omega_s] > 0$ *Q.E.D.*

**Proof of Proposition 2.2:** Adding up follows trivially by taking conditional expectations of $\iota'\phi = 1$ $(a.s)$. To see homogeneity, recall that

$$m(x+\lambda, q, u) = \int_{\mathcal{A}} \phi(x+\lambda, \vartheta(a,q)) F_{A|X,Q,U}(da, x+\lambda, q, u)$$

Inserting $\phi(x+\lambda, v) = \phi(x, v)$ as well as $F_{A|X,Q,U}(a, x+\lambda, q, u) = F_{A|X,Q,U}(a, x, q, u)$ produces

$$\int_{\mathcal{A}} \phi(x+\lambda, \vartheta(a,q)) F_{A|X,Q,U}(da, x+\lambda, q, u) = \int_{\mathcal{A}} \phi(x, \vartheta(a,q)) F_{A|X,Q,U}(da, x, q, u) = m(x, q, u).$$

The same argumentation holds also for

$$M(s+\lambda, q, u) = \int_{\mathcal{A}} \phi(\mu(s+\lambda, q, u) + \lambda, \vartheta(a,q)) F_{A|S,Q,U}(da, s, q, u),$$

using additionally that $\mu(s+\lambda, q, u) = \mu(s, q, u)$. Finally, the statements abouts the derivatives follow straightforwardly from the fact that homogeneity implies that $D_p\phi\iota + \partial_y\phi = 0$ $(a.s.)$, in connection with Lemma 2.1. *Q.E.D.*

**Proof of Proposition 2.3:** *Ad Negative Semidefiniteness* : Let $A(\omega)$, $\omega \in \Omega$, denote any random matrix. If $p'A(\omega)p \leq 0$ for all $\omega \in \Omega$ and all $p \in \mathbb{R}^L$, then, upon taking expectations w.r.t. an arbitrary probability measure $F$, it follows that

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

From this $\mathfrak{S}$ $nsd$ $(a.s.) \Rightarrow \mathbb{E}[\mathfrak{S}|X, Z]$ $nsd$ $(a.s.)$ is immediate. Let $\mathbb{E}[\mathfrak{S}|X, Z] = B$, 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

$$p'\bar{B}p = 2p'Bp \text{ for all } p \in \mathbb{R}^L,$$

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

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

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

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

From Lemma 2.1 it is apparent that $\bar{B}_1 = D_pm + D_pm'$. To see that $\bar{B}_2 = \partial_y m_2(x, z)$, first note that due to the boundedness of $W$ the second moments and conditional moments exist, so that

$$\begin{aligned} \partial_y m_2(x, z) &= \partial_y \int_{\mathcal{A}} \phi(x, \vartheta(q,a))\phi'(x, \vartheta(q,a)) F_{A|X,Z}(da; x, z) \\ &= \mathbb{E}[\partial_y(\phi\phi')|X = x, Z = z] \\ &= \mathbb{E}[\partial_y\phi\phi' + \phi\partial_y\phi'|X = x, Z = z] = \bar{B}_2 \end{aligned}$$

$B_3$ and $B_4$ are trivial. Upon inserting random variables, the statement follows. The proof using the regression with instruments and controls follows by the same arguments in connection with Lemma 2.1 ($iii$). *Q.E.D.*

*Ad Symmetry* To see the "if" direction, note first that $\mathfrak{S}$ symmetric iff $K = D_p\phi + \partial_y \phi\phi'$ is symmetric, which implies that $\mathbb{E}[K|Y,Z]$ is symmetric since $A_{ij} = \mathbb{E}[K_{ij}|Y,Z] = \mathbb{E}[K_{ji}|Y,Z] = A_{ji}$, where the subscript $ij$ denotes the $ij$-th element of the matrix. This implies in turn that

$$\begin{aligned}
\mathbb{E}[K|Y,Z] &= \mathbb{E}[D_p\phi|Y,Z] + \mathbb{E}[\partial_y\phi\phi'|Y,Z] \qquad \text{(A.2)}\\
&= \mathbb{E}[D_p\phi|Y,Z] + \mathbb{E}[\partial_y\phi|Y,Z]\,\mathbb{E}[\phi'|Y,Z] + \mathbb{V}[\partial_y\phi,\phi|Y,Z]
\end{aligned}$$

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 if $\mathbb{V}[\partial_y\phi,\phi|Y,Z]$ is assumed to be symmetric. By Lemma 2.1. this equals $D_pm + \partial_y mm'$.

To establish the "only if" direction, we have to show that $\mathbb{V}[\partial_y\phi,\phi|Y,Z]$ not symmetric implies that $\mathfrak{S}$ symmetric does not imply that $D_pm + \partial_y mm'$ be symmetric. To this end, assume again that $\mathfrak{S}$ be symmetric, and consider (A.2). In this case, $\mathbb{E}[K|Y,Z]$ is symmetric, but due to $\mathbb{V}[\partial_y\phi,\phi|Y,Z]$ not symmetric we obtain that

$$D_pm + \partial_y mm' = \mathbb{E}[K|Y,Z] - \mathbb{V}[\partial_y\phi,\phi|Y,Z]$$

has to be not symmetric as well. *Q.E.D.*

## Appendix 2

## Summary Statistics of Data: Household Charcteristics, Income and Budget Shares after Oulier Removement

| Variable | Minimum | 1st Quartile | Median | Mean | 3rd Quartile | maximum |
|---|---|---|---|---|---|---|
| number of female | 0 | 1 | 1 | 1.073 | 1 | 2 |
| number of retired | 0 | 0 | 0 | 0.051 | 0 | 1 |
| number of earners | 0 | 1 | 2 | 1.692 | 2 | 2 |
| Age of HHhead | 19 | 31 | 49 | 46 | 58 | 90 |
| Fridge | 0 | 1 | 1 | 0.987 | 1 | 1 |
| Washing Machine | 0 | 1 | 1 | 0.882 | 1 | 1 |
| Centr. Heating | 0 | 1 | 1 | 0.804 | 1 | 1 |
| TV | 0 | 1 | 1 | 0.874 | 1 | 1 |
| Video | 0 | 0 | 0 | 0.407 | 1 | 1 |
| PC | 0 | 0 | 0 | 0.792 | 0 | 1 |
| number of cars | 0 | 1 | 1 | 1.351 | 2 | 10 |
| number of rooms | 1 | 4 | 5 | 5.455 | 6 | 26 |
| ln.HHincome | 3.779 | 4.667 | 5.216 | 5.171 | 5.681 | 6.453 |
| BS GROUP 1 | 0 | 0.140 | 0.198 | 0.216 | 0.275 | 0.784 |
| Food bought | 0 | 0.082 | 0.130 | 0.148 | 0.195 | 0.784 |
| Catering | 0 | 0.014 | 0.035 | 0.046 | 0.065 | 0.719 |
| Tobacco | 0 | 0 | 0 | 0.020 | 0.032 | 0.426 |
| BS GROUP 2 | 0 | 0.205 | 0.297 | 0.317 | 0.406 | 0.959 |
| Housing | 0 | 0.102 | 0.179 | 0.198 | 0.272 | 0.874 |
| HHgoods | 0 | 0.018 | 0.042 | 0.075 | 0.091 | 0.939 |
| HHservices | 0 | 0.020 | 0.031 | 0.043 | 0.050 | 0.836 |
| BS GROUP 3 | 0 | 0.087 | 0.150 | 0.182 | 0.242 | 0.859 |
| Motoring | 0 | 0.035 | 0.096 | 0.130 | 0.186 | 0.841 |
| Fuel | 0 | 0.025 | 0.041 | 0.052 | 0.064 | 0.640 |

## Graphs

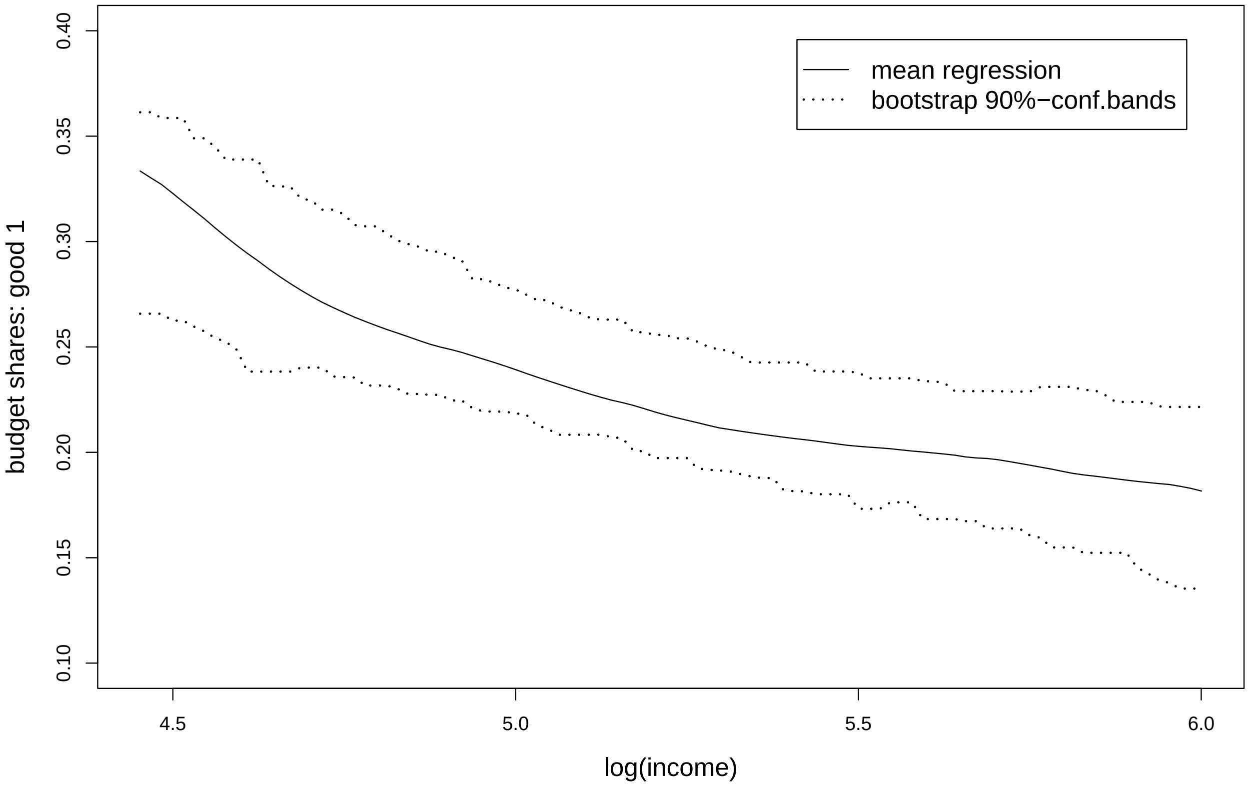


Figure 1: Mean regression budget share “food” on log income controlling for covariates and prices.

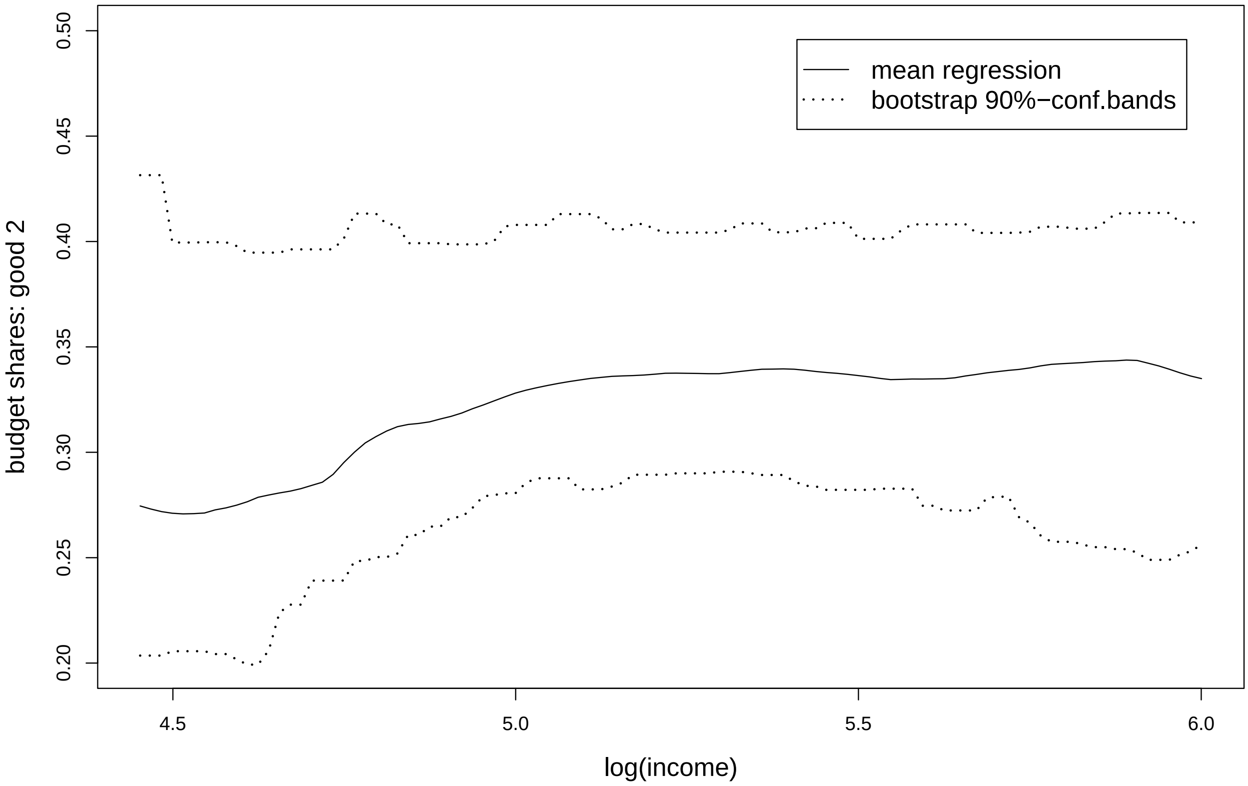


Figure 2: Mean regression budget share "housing" on log income controlling for covariates and prices.

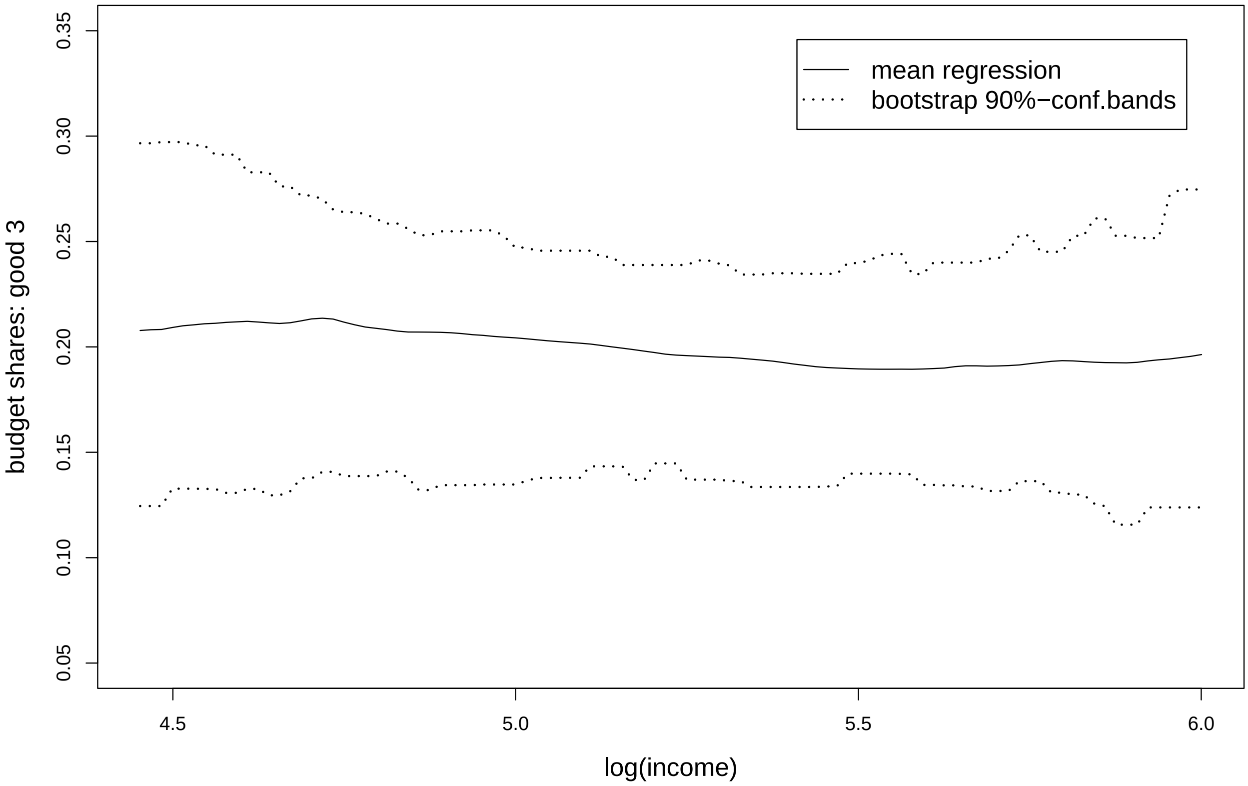


Figure 3: Mean regression budget share “energy” on log income controlling for covariates and prices.

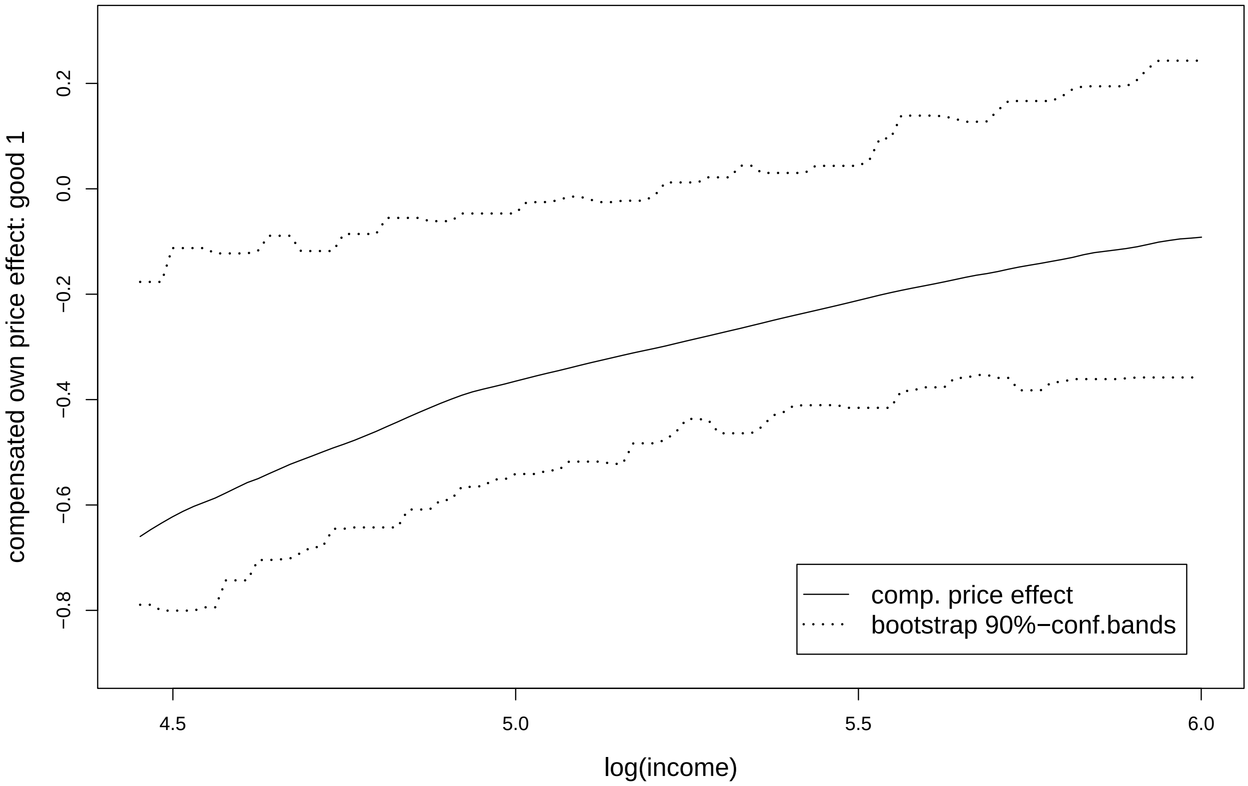


Figure 4: Compensated own price effect of “food”.

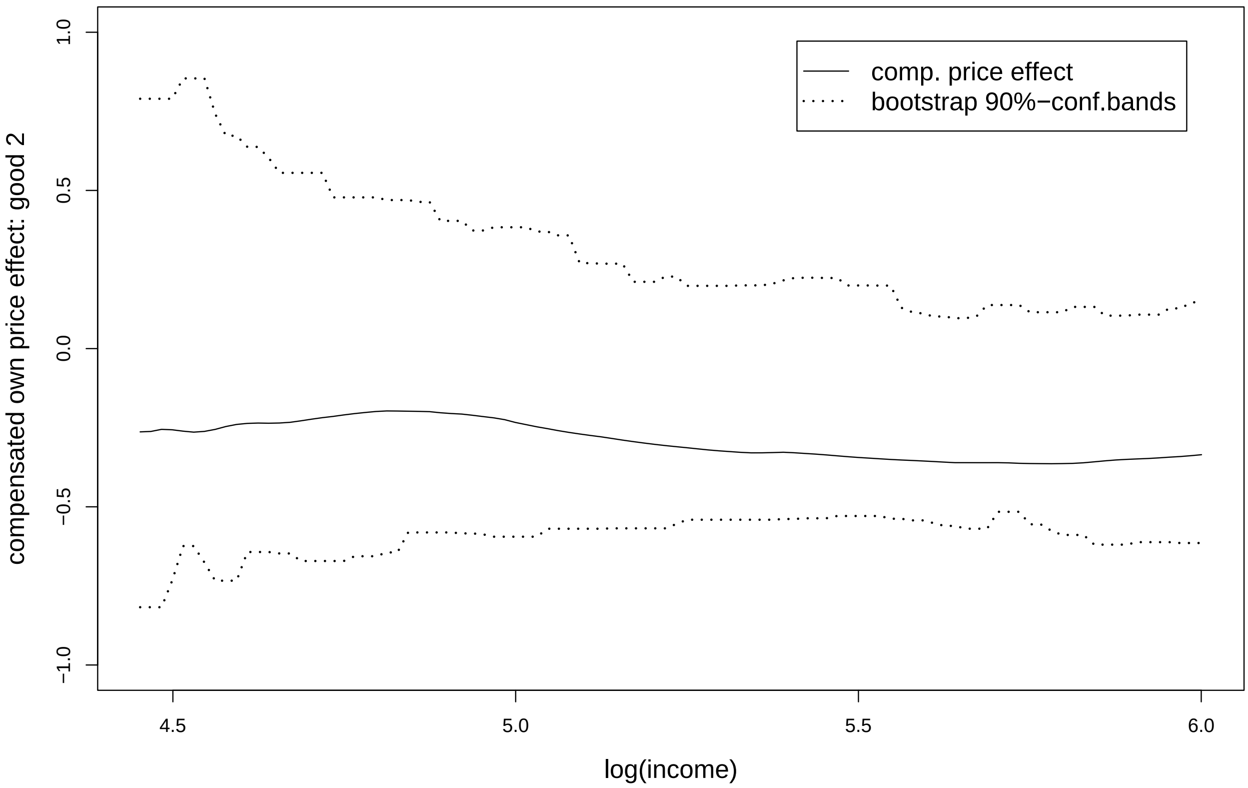


Figure 5: Compensated own price effect of “housing”.

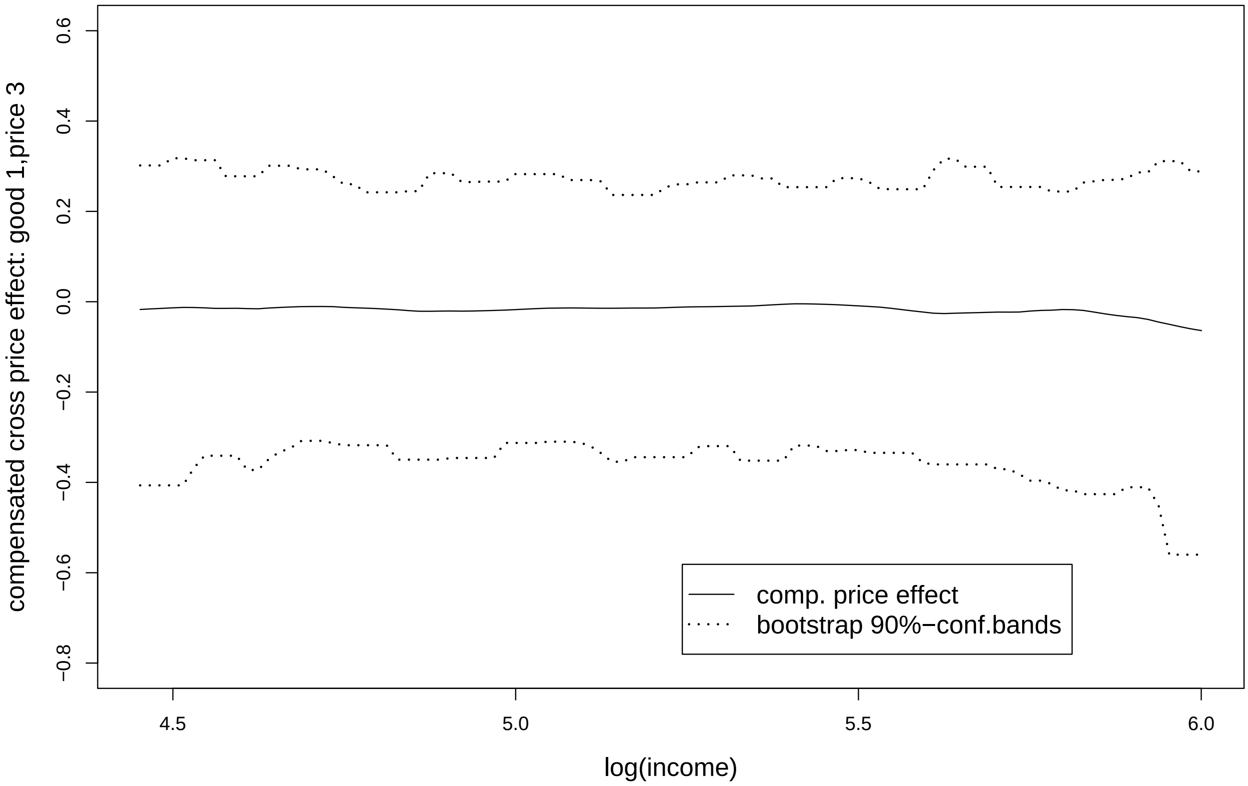


Figure 6: Compensated cross price effect of “energy” prices on “food” demand

# References


[1] ALTONJI, J., and R. MATZKIN (2005); Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors, *Econometrica*, 73, 1053 - 1103.

[2] BLUNDELL, R., BROWNING, M., and I. CRAWFORD (2003); Nonparametric Engel Curves and Revealed Preference, *Econometrica*, 71, 205-240.

[3] BLUNDELL, R., CHEN, X., and D. KRISTENSEN (2007); Semi-Nonparametric Estimation of Shape Invariant Engel Curves under Endogeneity, *Econometrica*, forthcoming.

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

[5] HAAG, B. and HODERLEIN, S. (2006); Testing Economic Hypotheses in Systems of Equations, Working Paper, Mannheim University.

[6] HÄRDLE, W., HILDENBRAND, W., and M. JERISON (1991); Empirical Evidence on the Law of Demand, *Econometrica,* 59, 1525-1549.

[7] IMBENS, G. and W. NEWEY (2003), Identification and Estimation of Triangular Simultaneous Equations Models without Additivity, Working Paper, MIT.

[8] 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.

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

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

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

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

[13] STOKER, T. (1989); Tests of Additive Derivative Constraints, *Review of Economic Studies,* 56, 535-552

# Appendix 3

# Econometric Specification

In this seperate appendix we establish the asymptotic properties of an estimator for the nonparametric demand systems, which is going to involve a system of local polynomial regression. Moreover, we discuss the pointwise test statistics for all economic hypotheses in the paper. As already pointed out in the text, the first order asymptotic distribution theory involves quantities that are cumbersome to estimate, which was the reason for using alternatively wild bootstrap procedures for testing all hypotheses. Here, we give arguments for the consistency of the bootstrap for the test statistics' distribution.

## The Nonparametric Demand System as Local Linear Systems of Equations

In the following subsection, we concentrate on the case of one endogenous variable, namely total expenditure. This is in line with the parametric literature, see Lewbel (1999). To treat the exogenous and the endogenous scenario under the same format, we consider the regression of $W$ on $P, B$ and $Z$, where $B$ may now denote either total expenditure or labor income, which were denoted as $Y$ or $S$, respectively, and recall that $Z$ may include control function residuals $U$. This extension is not trivial, as this regressor has to be replaced by a pre-estimated version which involves, e.g., a Nadaraya Watson pre-estimator $\hat{\mu}$ for $\mu$. Finally, we assume $\mu$ to be an additive function of the error, but our specification will allow for heteroscedasticity. We assume that

$$Y = \mu(S, Z) = \psi(S, Q) + \sigma_2(S, Q)U, \tag{4.1}$$

with the normalization $\mathbb{E}[U|S, Q] = 0$ and $\mathbb{V}[U|S, Q] = 1$. This specific structure will lead to slightly more specific implications for the test statistics. In particular, we will require an estimator for the scedastic function, $\partial_s \sigma_2$.

In what follows we assume to have an *iid* sample, and we index all random variables by a subscript $i$, while superscript $j$ indexes goods (equations). Following standard notational convention, let $\mathcal{X}_i = [P_i', B_i, Z_i']'$ denote all regressors, where $Z_i = [Q_i', U_{ni}]'$ contains the estimated control function residuals from the regression of total expenditure on instruments, formally $U_{ni} = B_i - \hat{m}_L(S_i, Q_i)$, where $\hat{m}_L$ denotes a Nadaraya Watson estimator. Hence, the set of all regressors has dimension $d = L + 1 + K + 1$.

The task of estimating an empirical object which has some economically interpretable structure involves the model

$$W_i = m(\mathcal{X}_i) + \Sigma_1(\mathcal{X}_i)\eta_i,$$

where $m(\cdot)$ and $\Sigma_1(\cdot)$ are now assumed to be Borel-measurable, smooth, $L-1$ vector valued mean regression[2], and $L-1 \times L-1$ matrix valued scedastic functions, respectively. We assume that the $\eta_i$ are *iid* mean zero unit variance random vectors. In addition to the $L-1$ budget shares we consider additional dependent variables denoted as $R_{ni}$, because we require an estimator for the mean regression $\psi$, as well as the scedastic function $\sigma_2$ in mode (4.1). Consequently, let $J_i = [W_i', Y_i, U_{ni}^2]' = [W_i', R_{ni}']'$ denote the vector of dependent variables, which is of length $\overline{L} = L + 1$.

The estimation of the mean regression functions, the scedastic function and their derivatives at a particular point $x_0$ is based on local polynomial modelling. To this end, consider the systems local linear (SLL) estimator which solves the following weighted least squares minimization problem:

$$\min_{\theta(z)} n^{-1} \sum_{i=1}^{n} K_{H_n}\left(\mathcal{X}_i - x_0\right) \xi_i' \xi_i, \tag{A.1}$$

where, $\theta(x_0) = \{\theta^1(x_0), ..., \theta^{\overline{L}}(x_0)),$and for all $j = 1, .., \overline{L}$, $\theta^j(x_0) = \{m^j(x_0), h\partial_{p_l} m^j(x_0), h\partial_y m^j(x_0),$ $h\partial_{z_k} m^j(x_0)\}$, $j = 1, .., \overline{L}$, $l = 1, .., L$, $k = 1, .., K$. The scaling of the marginal effects by $h$ has been performed to keep track of the difference in speed of convergence. Moreover, $\xi_i = [\xi_i^1, ..., \xi_i^{\overline{L}}]'$, where, for all $j = 1, .., \overline{L}$,

$$\xi_i^j = J_i^j - m^j(x_0) - \sum_{l=1}^{L} h\partial_{p_l} m^j(x_0) \frac{P_i^l - p^l}{h} - h\partial_b m^j(x_0) \frac{B_i - b}{h} - \sum_{k=1}^{K} h\partial_{z_k} m^j(x_0) \left( \frac{Z_i^k - z^k}{h} \right).$$

In addition, let $K_{H_n}(\psi) = |H_n|^{-1/2} K(H_n^{-1/2}\psi)$ where $K$ is an $L$-variate kernel such that $\int K(\psi) d\psi = 1$, and $H_n$ is an $L \times L$ symmetric positive definite bandwidth matrix depending on $n$. For simplicity of exposition, we shall use a product Kernel and a diagonal bandwidth matrix, with $H_n = h_n^2 I_L$. The asymptotic distribution of this systems-of-equations, regression-plus-scedastic function estimator with generated regressors is given by the following theorem:

**Proposition A.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. Moreover, $bias(x_0)$ contains the leading bias term detailed in the proof, $A$ is a fixed $(d+1) \times (d+1)$ matrix given by*

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

$A = f_X(x_0)^{-1}B^{-1}CB^{-1}$, $f_X(x_0) > 0$ *denotes the joint distribution of all regressors, and the fixed matrices* $B$ and $C$ *are defined as*

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

*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: See Appendix.*

**Remarks:** This result extends standard local polynomial estimators to systems of equations and our setup which includes the control function residuals as generated regressors. The latter is perhaps the main innovation. To tackle it, in addition to standard assumptions we require that $\psi$ and $\sigma_2$ be four times differentiable and the Kernel be of fourth order and differentiable.

## Testing Hypotheses in this Framework

This subsection is concerned with devising tests for the various hypotheses. We focus in particular on devising bootstrap versions of our test statistics, firstly because they are straightforward to implement, but secondly because the first order asymptotic analysis may give a poor approximation to the true finite sample behavior of the test statistics.

### Homogeneity

Recall the testable implications of the assumptions that homogeneity holds in a heterogeneous population, as given in $P2.2$; adapted to our scenario:

$$D_p m(x_0)\,\iota + \partial_y m(x_0) = 0 \text{ for all } x_0 \in supp(\mathcal{X}). \tag{4.2}$$

and $D_p M(x_0)\,\iota + \partial_s M(x_0)\left(\partial_s\psi(x_0) + \partial_s\sigma_2(x_0)\,U\right)^{-1} = 0$ for all $x_0 \in supp(\mathcal{X})$. This hypothesis is easily rewritten as $R\theta(x_0) = 0$, where $R$ denotes the $L-1 \times d(L-1)$ matrix $R = I_{L-1} \otimes \begin{bmatrix} 0 & \iota_{L+1} & 0_K \end{bmatrix}$, which together with $P3.1$ suggests the test statistic

$$\widehat{\Psi}_{Hom,Ex}(x_0) = \left[R\left[\hat{\theta}(x_0) - h^2\widehat{bias}(x_0)\right]\right]'\left[R'\left[\widehat{\Sigma_1\Sigma_1'}(x_0)\otimes\hat{A}(x_0)\right]R\right]^{-1}R\left[\hat{\theta}(x_0) - h^2\widehat{bias}(x_0)\right],$$

where $\hat{A}(x_0) = \hat{f}_X(x_0)^{-1}B^{-1}CB^{-1}$ and $\widehat{bias}(x)$ is a pre-estimator for the bias[3]. Then, by a trivial corollary to proposition 3.1, $\widehat{\Psi}_{Hom,Ex} \xrightarrow{d} \chi^2_{L-1}$.

This test statistic could be used in principle, however it's disadvantage is that involves the estimation of the multivariate scedastic function $\widehat{\Sigma_1\Sigma_1'}$ and an pre-estimator of the bias, both of which may be hard to obtain in practise. Hence, we consider the following test statistic instead:

$$\widehat{\Gamma}_{Hom,Ex}(x_0) = \left(R\hat{\theta}(x_0)\right)' R\hat{\theta}(x_0).$$

It's asymptotic distribution may be derived through the fact that for any multinormal random $L$-vector $\zeta$, $\zeta \sim N(0,\Sigma)$ the statistic $\Xi = \zeta'\zeta$ has moment generating function $mgf_\Xi(t) = \det\{I - 2t\Sigma\}^{-1/2}$. However, as with the above test $\widehat{\Psi}_{Hom,Ex}$ this distribution is hardly useable in practise, and we will apply the bootstrap to obtain an approximation to the test's true distribution instead.

More specifically, the following bootstrap procedure appears natural: 1. Calculate (multivariate) residuals $\hat{\varepsilon}_i = W_i - \hat{m}(\tilde{P}_i, Y_i, Z_i)$. 2. For each $i$ randomly draw $\varepsilon_i^* = (\varepsilon_i^{1*}, \ldots, \varepsilon_i^{L-1,*})'$ from a distribution $\hat{F}_i$ that mimics the first two moments of $\hat{\varepsilon}_i$. 3. Generate the bootstrap sample $(W_i^*, \tilde{P}_i^*, Z_i^*), i = 1, \ldots, n$ by $W_i^* = \hat{m}_{\tilde{h}}(\tilde{P}_i, Z_i) + \varepsilon_i^*$ and $\tilde{P}_i^* = \tilde{P}_i, Z_i^* = Z_i$. Here $\hat{m}_{\tilde{h}}(\tilde{P}_i, Z_i)$ denotes the restricted estimator. 4. Calculate $\widehat{\Gamma}^*_{Hom,Ex}$ from the bootstrap sample $(W_i^*, \tilde{P}_i^*, Z_i^*), i = 1, \ldots, n$. 5. Repeat steps 2 to 4 $B$ times to obtain critical values for $\widehat{\Gamma}_{Hom,Ex}$.

To see that the wild bootstrap provides a consistent estimator for the test statistic's distribution, consider the following: First, the asymptotic distribution of the SLLE is derived as above. Second, the bootstrap version of the SLLE, $\hat{\theta}^*(x_0)$ converges under the same conditions, as well as under standard assumptions on the distribution of $\varepsilon_i^*$ against the same limiting distribution. Third, since $\widehat{\Gamma}_{Hom,Ex}(x_0)$ is a simple quadratic form of the original estimators, apply the continuous mapping theorem to see that the properly normalized version of the test statistic and its bootstrap version converge against the same limiting distribution. In Haag and Hoderlein (2006), we examine in detail the asymptotic behavior of a more general test that is a sum of the homogeneity test statistics, ie $\hat{\tau} = n^{-1}\sum_{i=1,..,n}\widehat{\Gamma}_{Hom,Ex}(\mathcal{X}_i)$, including a formal examination of the consistency of the bootstrap. Many of the arguments can be directly transferred to our setup, however, since large sample theory is not the focus of the paper we desist from this here.

The test becomes slightly more involved if we want to scrutinize homogeneity in the endogenous case. First, note that $\partial_s\sigma_2 = \partial_s\sigma_2^2\left[2\left(\sigma_2^2\right)^{-1/2}\right]^{-1}$. Then, $G(\theta, x_0) = (G_1(\theta, x_0), .., G_{L-1}(\theta, x_0))' =$

[3] 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.

0, where $\forall\ j$, $G_j\left(\theta, x_0\right) = \left\{\beta_y^{\overline{L}}(x_0) + 0.5\beta_y^{\overline{L}}(x_0)\alpha^{\overline{L}}\left(x_0\right)^{-1/2} u\right\} \sum_l \beta_l^j(x_0) + \beta_y^j(x_0)$. This leads to,

$$\widehat{\Gamma}_{Hom,End}(x_0) = \left[G\left(\hat{\theta}, x_0\right)\right]' G\left(\hat{\theta}, x_0\right),$$

for which critical values are again obtained via the bootstrap.

**Symmetry**

Now the test statistics for the hypothesis of symmetry in a heterogeneous population - as given in $P2.3$ - are being analyzed. Recall that under the additional identification assumption that $\mathbb{V}\left\{\partial_y \phi, \phi | X, Z\right\}$ be symmetric, the matrices $D_p m(x_0) + \partial_y m(x_0) m(x_0)'$ and $D_p M(x_0) + \partial_s M(x_0)$ $\left\{\partial_s \psi(x_0) + \partial_s \sigma_2(x_0) U\right\}^{-1} M(x_0)'$ are shown to be symmetric for all $x_0 \in$supp$(\mathcal{X})$. To derive a test for the first case, stack the $1/2L(L-1)$ nonlinear restrictions $R_{kl}(\theta, x_0)$, at a fixed position $x_0$,

$$R_{kl}(\theta, x_0) = \partial_{p_k} m^l(x_0) + \partial_y m^l(x_0) m^k(x_0) - \left(\partial_{p_l} m^k(x_0) + \partial_y m^k(x_0) m^l(x_0)\right) = 0, \quad k, l = 1, .., L-1, k > l,$$

into a vector $R$. Consequently, "$D_p m(x_0) + \partial_y m(x_0) m(x_0)'$ symmetric for all $x_0 \in supp\left(\mathcal{X}\right)$" becomes "$R(\theta, x_0) = 0$ for all $x_0 \in supp\left(\mathcal{X}\right)$". This suggest using the quadratic form

$$\widehat{\Gamma}_{Sym,Ex}(x_0) = \left[R(\hat{\theta}, x_0)\right]' R(\hat{\theta}, x_0),$$

and check whether this is significantly bigger than zero. Again, building upon $P3.1$, a standard distribution theory could be derived. However, looking at the complicated structure of the test statistic for symmetry, the asymptotic approach to implement the test might even be less reliable than it was for the tests of homogeneity, and bootstrap seems to be the method of choice. But to use the wild bootstrap to calculate critical values for $\hat{\Gamma}_{Sym,Ex}$, the procedure has to be changed since we have no restricted estimator to add on the unrestricted residuals. In contrast, we plug the residuals directly in the test statistic. To see how this can be sensibly done, note that the estimator of the derivative can be written as a weighted average $\partial_{p_k} \hat{m}_h^l(x_0) = \sum_{i=1}^n \tilde{V}_i^{kl}(x_0) W_i^l$, where $\tilde{V}_i^{kl}(x_0)$ denote weights which, when applied to $W_i^l$, yield an estimator for the price derivative, and analogously $\partial_y \hat{m}_h^l(x_0) = \sum_{i=1}^n \tilde{V}_i^{yl}(x_0) W_i^l$ for income. Using this in the definition of $\hat{\Gamma}_{Sym,Ex}(x_0)$ we obtain

$$\hat{\Gamma}_{Sym,Ex}(x_0) = \sum_{l=1}^{L-2} \sum_{k=l+1}^{L-1} \left(\sum_{j=1}^{n} V_j^{kl}(x_0) W_j^l - V_j^{lk}(x_0) W_j^k\right)^2$$

with $V_j^{kl}(x_0) = \tilde{V}_j^{kl}(x_0) + \hat{m}_h^k(x_0) \tilde{V}_j^{yl}(x_0)$. Here we use that the estimator of the function converges faster than the estimator of the derivative. Substituting $W_l^l = m^l(Z_l) + \varepsilon_l^l$ and noting that

$$\sum_{j=1}^{n} V_j^{lk}(x_0) m^l(X_j, Z_j) + V_j^{kl}(x_0) m^k(X_j, Z_j) \approx \partial_{p_k} m^j(x_0) + m^k(x_0) \partial_y m^l(x_0) - \partial_{p_l} m^k(x_0) - m^l(x_0) \partial_y m^k(x_0) = 0$$

for large $n$ and under $H_0$, the test statistic can be approximated by

$$\hat{\Gamma}_{Sym,Ex}(x_0) \approx \sum_{l=1}^{L-2}\sum_{k=l+1}^{L-1}\left(\sum_{j=1}^{n} V_j^{kl}(x_0)\varepsilon_j^l - V_j^{lk}(x_0)\varepsilon_j^k\right)^2$$

The bootstrap is based on this equation and is described as follows: 1. Construct (multivariate) residuals $\hat{\varepsilon}_j = W_j - \hat{m}_h(X_j)$. 2. For each $i$ randomly draw $\varepsilon_i^*$ from a two point distribution that mimics the first two moments of $\hat{\varepsilon}_i$. 3. Calculate $\hat{\Gamma}^*_{Sym,Ex}(x_0)$ from the bootstrap sample $(\varepsilon_j^*, X_j)$, $j = 1, \ldots, n$ by

$$\hat{\Gamma}^*_{Sym,Ex}(x_0) = \sum_{l=1}^{L-2}\sum_{k=l+1}^{L-1}\left(\sum_{j=1}^{n} V_j^{kl}(x_0)\varepsilon_j^{*l} - V_j^{lk}(x_0)\varepsilon_j^{*k}\right)^2,$$

4. Repeat this often enough to obtain critical values. The test for symmetry under endogeneity follows by the same arguments with $U$ as additional regressor.

To see that the wild bootstrap provides a consistent estimator for the statistic's asymptotic distribution in the case of symmetry, first note that (under standard conditions and using similar arguments as in the proof of the asymptotic distribution of the LPEs) the limiting distributions of $\hat{\Gamma}_{Sym,Ex}(x_0)$ and $\tilde{\Gamma}_{Sym,Ex}(x_0)$ coincide under the null up to terms of smaller order than the leading bias term. Second, under the same conditions, as well as standard assumptions on the distribution of $\varepsilon_i^*$, $\tilde{\Gamma}_{Sym,Ex}(x_0)$ and $\hat{\Gamma}^*_{Sym,Ex}(x_0)$ converge against the same limiting distribution. For further technical details we refer now to Haag, Hoderlein and Pendakur (2006), where we examine the asymptotic behavior of the more general test that is a sum of the symmetry test statistics, ie $\hat{\tau} = n^{-1}\sum_{i=1,..,n}\widehat{\Gamma}_{Sym,Ex}(\mathcal{X}_i)$.

A test for symmetry under endogeneity follows by similar arguments, noting that the instrument equation is not affected by the restriction and therefore sampling under the restriction equals sampling without the restriction.

### Negative Semidefiniteness

Negative Semidefiniteness poses an added difficulty, as (much like symmetry) no restricted estimator is available, but in addition we do not have a test statistic that is a linear (or asymptotically linear) combination of the parameters. Instead, we focus on the bootstrap distribution of the largest eigenvalue of the Slutsky matrix in the following fashion: 1. Construct residuals $\hat{\varepsilon}_j = W_j - \hat{m}_h(X_j, Z_j)$. 2. For each $i$ randomly draw $\varepsilon_j^*$ from a distribution that mimics the first two moments of $\hat{\varepsilon}_i$. 3. Calculate the Slutsky matrix $S^*$ and largest eigenvalue $\lambda^*$ from the bootstrap sample $(\varepsilon_j^*, X_j, Z_j)$, $j = 1, \ldots, n$. 4. Repeat this often enough to obtain critical values. 5. Reject the hypothesis of negative semidefiniteness if the 0.05 quantile of the distribution of the largest eigenvalue exceeds 0.

This procedure is similar then the one proposed by Härdle and Hart (1993), and we refer the reader to their paper for technical details regarding the consistency of the bootstrap.

## Proof of Proposition A.1:

In this proof we treat the asymptotics for the systems local polynomial estimator, including 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.

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

(A1) The $\eta_i$ are zero mean and unity diagonal variance random vectors s.t. $\eta_i$ is independent of $\mathcal{X}_1, \cdots, \mathcal{X}_i, \eta_1, \cdots, \eta_{i-1}, , \varepsilon_1, \cdots, \varepsilon_i$ for each $i \geq 1$ and every element of $\eta_i \eta_i'$ is uniformly integrable. The $\varepsilon_i$ are zero mean and unit variance random vectors s.t. $\varepsilon_i$ is independent of $\mathcal{X}_1, \cdots, \mathcal{X}_i, \eta_1, \cdots, \eta_i, , \varepsilon_1, \cdots, \varepsilon_{i-1}$ for each $i \geq 1$. Moreover, the $\varepsilon_i$ have finite fourth moments and $\mathbb{E}\left[\varepsilon_i^l \eta_i\right]$, $l = 1, 2, 3$ are finite is uniformly integrable. Finally, $\varepsilon_i$ and $\varepsilon_i^l \eta_i$, $l = 1, 2, 3$ are uniformly integrable.

(A2) The $\mathcal{X}_i$ are iid with common density $f_{\mathcal{X}}$.

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

(A4) $f_{\mathcal{X}}$ is bounded, as well as $f_{\mathcal{X}}(x) > 0 \ \forall x$.

(A5) Every element of the Hessian of $m_j$, $j = 1, .., L$ and of $\sigma_2^2$ is bounded.

(A6) Every element of $\Xi$ 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 A.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.*

**Proof of Propostion A.1:** The estimator given by the first order conditions is

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

Here, J is the $\bar{L}n \times 1$ vector given by J$= (J_1', .., J_{M-1}')'$ with $J_j = (K_1^{\frac{1}{2}} J_{j1}, .., K_n^{\frac{1}{2}} J_{jn})'$ $\forall j = 1, .., \bar{L}$, 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 $\mathbb{X} = (\mathcal{X}^0, .., \mathcal{X}^d)$ and
$\mathcal{X}^m = (K_1^{\frac{1}{2}} \left(\mathcal{X}_1^m - x_0^m\right)/h, .., K_n^{\frac{1}{2}} \left(\mathcal{X}_n^m - x_0^m\right)/h)'$ $\forall m = 1, .., d$ as well as
$\mathcal{X}^0 = (K_1^{\frac{1}{2}}, .., K_n^{\frac{1}{2}})'$. Moreover, we need the notation
$\mathrm{X}_i - x_0 = (1, \left(\mathcal{X}_i^1 - x_0^1\right)/h, .., \left(\mathcal{X}_i^d - x_0^d\right)/h))'$
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 $\tilde{B}_j = (K_1^{\frac{1}{2}} \frac{h^2}{2} \iota' \mathcal{H}_m(\mathcal{X}_{r1}) \iota, .., K_n^{\frac{1}{2}} \frac{h^2}{2} \iota' \mathcal{H}_m(\mathcal{X}_{rn}) \iota)'$ $\forall j = 1, .., L$, where
$\mathcal{H}_m(\mathcal{X}_{r1})$ is the Hessian of $m$ at an intermediate position, $\mathcal{X}_{r1} = x_0 + \lambda(\mathcal{X}_i)'(\mathcal{X}_i - x_0)$
and $\tilde{B}_{\bar{L}} = (K_1^{\frac{1}{2}} \frac{h^2}{2} \mathcal{H}\sigma_2^2(\mathcal{X}_{r1})'\iota, .., K_n^{\frac{1}{2}} \frac{h^2}{2} \mathcal{H}\sigma_2^2(\mathcal{X}_{rn})'\iota)'$,
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
$G_i = 2\sigma_2(\mathcal{X}_i) U_i (\hat{\mu}(\mathcal{X}_i) - \mu(\mathcal{X}_i)) + (\hat{\mu}(\mathcal{X}_i) - \mu(\mathcal{X}_i))^2$
Finally, $\mathbb{U}$ is the $\overline{L}n \times 1$ vector given by $\mathbb{U} = (\tilde{U}_1', .., \tilde{U}_{\bar{L}}')'$ with
$\tilde{U}_j = (K_1^{\frac{1}{2}} \Sigma_{1,j}(\mathcal{X}_1)\eta_1, .., K_n^{\frac{1}{2}} \Sigma_{1,j}(\mathcal{X}_n)\eta_n)'$ $\forall j = 1, .., L-1$,
$\tilde{U}_L = (K_1^{\frac{1}{2}} \sigma_2(\mathcal{X}_1)\varepsilon_1, .., K_n^{\frac{1}{2}} \sigma_2(\mathcal{X}_n)\varepsilon_n)'$ and
$\tilde{U}_{\overline{L}} = (K_1^{\frac{1}{2}} \sigma_2^2(\mathcal{X}_1)\left(\varepsilon_1^2 - 1\right), .., K_n^{\frac{1}{2}} \sigma_2^2(\mathcal{X}_n)\left(\varepsilon_n^2 - 1\right))'$.

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**

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

The second lemma treats the first bias expression:

**Lemma A.2**

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

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**

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

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. *Q.E.D.*

**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

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

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

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

Thus,

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

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

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

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. *Q.E.D.*

**Proof of Lemma A.4**

First, recall the notation $\mathcal{X}_i - x_0 = (1, (\mathcal{X}_i^1 - x_0^1)/h, .., (\mathcal{X}_i^d - x_0^d)/h))'$.

We show: $(\phi_{ni}, \mathcal{F}_{ni})$, $i = 1, \ldots, n$, $n \geq 1$, with $\phi_{ni} = (\phi'_{ni1}, .., \phi'_{ni\bar{L}})'$ and

$\phi_{nij} = \frac{1}{\sqrt{nh^d}} K((\mathcal{X}_i - x_0)/h)(\mathcal{X}_i - x_0)\Sigma_{1,j}(\mathcal{X}_i)\eta_{ni}$, $\forall j = 1, .., L-1$,

$\phi_{niLs} = \frac{1}{\sqrt{nh^d}} K((\mathcal{X}_i - x_0)/h)(\mathcal{X}_i - x_0)\sigma_2(\mathcal{X}_i)\varepsilon_{ni}$,

$\phi_{ni\bar{L}s} = \frac{1}{\sqrt{nh^d}} K((\mathcal{X}_i - x_0)/h)((\mathcal{X}_i - x_0)/h)\sigma_2^2(\mathcal{X}_i)(\varepsilon_{ni}^2 - 1)$,

$\mathcal{F}_{ni} = \mathcal{F}_i$,

is a martingale difference array such that

(i) $\operatorname{plim}_{n\to\infty} \sum_{i=1}^n \mathbb{E}[\phi_{ni}\phi'_{ni}|\mathcal{F}_{i-1}] = \Xi(x_0) \otimes C$,

(ii) $\operatorname{plim}_{n\to\infty} \sum_{i=1}^n \mathbb{E}\left[\|\phi_{ni}\|^2 \mathbf{1}_{\{\|\phi_{ni}\|>\delta\}}|\mathcal{F}_{i-1}\right] = 0$ 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

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

where $\Sigma_{jk}$ denotes the $jk$-th element of $\Sigma_1$. The same holds for terms involving $\sigma_2, \sigma_2^2$. Therefore, by similar reasoning as above,

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

As to (ii), note first that

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

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 $b_{ni} = K_i(\mathcal{X}_i - x_0)$, we have

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

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 $\mathcal{X}_i$ is outside a $h$-neighborhood of $x_0$. Therefore the $||b_{ni}||^2 \Sigma_{1,11}(\mathcal{X}_i)$ are uniformly bounded by a constant for $h$ small enough, if the realizations of $\mathcal{X}_i$ 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 $\eta_i^2$, 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

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

where $V$ is a nonrandom matrix, the last term tends to zero in probability as well. *Q.E.D.*

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

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

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

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

and let $\hat{D}_n = \dfrac{1}{n}\sum_{i=1}^n K_i \left[\left(X_i^j - x_0^j\right)/h\right]^2 \left(\hat{\mu}(\mathcal{X}_i) - \mu(\mathcal{X}_i)\right)^2.$

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 $f_{JXZ}(j,x,z)$ denote the joint density of $(J_i, X_i, Z_i)$, and let $f_{XZ}(x,z)$ denote the joint density of $(X_i, Z_i)$. Let

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

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

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

Let

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

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

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

and

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

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

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

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

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

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

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 $\Psi$ 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,

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

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

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

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 $s = 0, 1$,

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

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

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

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 $\Gamma$ 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

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

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)

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

Let

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

Integration by parts yields,

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

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

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

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.

$$\frac{2}{\sqrt{nh^d}}\sum_{i=1}^n K_iU_i\left(X_i-x_0\right)\left(\hat{\mu}(X_i)-\mu(X_i)\right). \tag{4.3}$$

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 $r \geq 2, d \geq 3$. *Q.E.D.*

## 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 $U_i = X_i - \mu(Z_i)$ are replaced by $U_{ni} = X_i - \hat{\mu}(Z_i)$, 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 $\mu$ be four times continuously differentiable.

Finally, recall the expansion

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

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

**Changes to Lemma A.1:** Let

$$\mathbb{X}'\mathbb{X} = \left[\begin{array}{cc} \sum_{i=1}^n K_{ni}\left(\mathcal{X}_i - x_0\right)\left(\mathcal{X}_i - x_0\right)'/h^2 & \sum_{i=1}^n K_{ni}\left(\mathcal{X}_i - x_0\right)\left(U_{ni} - u_0\right)/h^2 \\ \sum_{i=1}^n K_{ni}\left(U_{ni} - u_0\right)\left(\mathcal{X}_i - x_0\right)'/h^2 & \sum_{i=1}^n K_{ni}\left(U_{ni} - u_0\right)^2/h^2 \end{array}\right],$$

where $K_{ni} = K_{ni}(x_0, u_0) = K\left(\frac{\mathcal{X}_i - x_0}{h}\right) K\left(\frac{U_i - u_0}{h}\right)$. 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,

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

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

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

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 $\left(\left(\mu_i - \hat{\mu}_i\right)/h\right)$ are actually more benign, because $\|\mu_i - \hat{\mu}_i\| = o_p(h)$. In contrast, more problematic are terms that are linear in $\left(\left(\mu_i - \hat{\mu}_i\right)/h\right)$.

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. *Q.E.D.*

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

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

where $K_{\mathcal{X}i}K'_{Ui} = K\left(\frac{\mathcal{X}_i - x_0}{h}\right)K'\left(\frac{U_i - u_0}{h}\right)$. Taking again the first element of $(\mathcal{X}_i - x_0)$, 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 $r \geq 2$. *Q.E.D.*

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

**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

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

$a, b, c \in \{0, 1, 2\}$, $l \in \{1, 2\}$ $g \in \{1, 2, 3, 4\}$, $\zeta(x) = x$ or $x^2 - 1$, and $\mathbb{E}\left[ \zeta(\varepsilon_i) | \mathcal{X}_i, U_i \right] = 0$. Consider again the linear in $\left( \left( \mu_i - \hat{\mu}_i \right) / h \right)$, and $a = 1, b = c = 0$. Using $\int \varepsilon dF_{\varepsilon|\mathcal{X},U} = 0$, 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. *Q.E.D.*