cemmap
centre for microdata methods and practice

Testing multivariate economic restrictions using quantiles: the example of Slutsky negative semidefiniteness

Holger Dette
Stefan Hoderlein
Natalie Neumeyer

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP14/11

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL
An ESRC Research Centre

Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE **Tel:** 44 (0)20 7291 4800 **Email:** info@cemmap.ac.uk **www.cemmap.ac.uk**

Testing Multivariate Economic Restrictions Using Quantiles: The Example of Slutsky Negative Semidefiniteness

Holger Dette*
University of Bochum

Stefan Hoderlein †
Boston College

Natalie Neumeyer‡
University of Hamburg

February 8, 2011

Abstract

This paper is concerned with testing rationality restrictions using quantile regression methods. Specifically, we consider negative semidefiniteness of the Slutsky matrix, arguably the core restriction implied by utility maximization. We consider a heterogeneous population characterized by a system of nonseparable structural equations with infinite dimensional unobservable. To analyze the economic restriction, we employ quantile regression methods because they allow us to utilize the entire distribution of the data. Difficulties arise because the restriction involves several equations, while the quantile is a univariate concept. We establish that we may test the economic restriction by considering quantiles of linear combinations of the dependent variable. For this hypothesis we develop a new empirical process based test that applies kernel quantile estimators, and derive its large sample behavior. We investigate the performance of the test in a simulation study. Finally, we apply all concepts to Canadian individual data, and show that rationality is an acceptable description of actual individual behavior.

Keywords: Nonparametric Testing, Heterogeneity, Integrability, Nonseparable Models, Consumer Demand, Quantile Regression.

*Holger Dette, Ruhr-Universität Bochum, Fakultät für Mathematik, 44780 Bochum, Germany, email: holger.dette@rub.de

†Department of Economics, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, USA, Tel. +1-617-552-6042. email: stefan_hoderlein@yahoo.com

‡Natalie Neumeyer, University of Hamburg, Department of Mathematics, Bundesstrasse 55, 20146 Hamburg, Germany, email: neumeyer@math.uni-hamburg.de

1 Introduction

Economic theory yields strong implications for the actual behavior of individuals. In the standard utility maximization model for instance, economic theory places strong restrictions on individual responses to changes in prices and wealth, the so-called integrability constraints. These restrictions are inherently restrictions on individual level: They have to hold for every preference ordering and every single individual, at any price wealth combination. Other than obeying these restrictions, the individuals' idiosyncratic preference orderings may exhibit a lot of differences. Indeed, standard parametric cross section mean regression methods applied to consumer demand data often exhibit R^2 between 0.1 and 0.2. Today, the consensus is that the majority of the unexplained variation is precisely due to unobserved preference heterogeneity. For this reason, the literature has become increasingly interested in exploiting all the information about unobserved heterogeneity contained in the data, in particular using the quantiles of the dependent variable.

To lay out our model, let y denote the $L-1$ vector of quantities demanded. At this stage, we have already imposed the adding up constraint (i.e., out of L goods we have deleted the last). Let p denote the L vector of prices, and x denote income (total expenditure)[1]. For every individual, define the cost function $C(p,u)$ to give the minimum cost to attain utility level u facing the L-vector of prices p. The Slutsky negative semidefiniteness restriction arises from the fact that the cost function is concave, and hence the matrix of second derivatives is negative semidefinite (nsd, henceforth). In slight abuse of notation, define the Marshallian demand function for an individual with preferences $u \in \mathcal{U}$, where $\mathcal{U}$ is a preference space, e.g., the space of r times continuously differentiable utility functions, to be $y = \psi(p,x;u)$. We consider Slutsky nsd for a $(L-1)\times(L-1)$ submatrix of the Hessian of the cost function, denoted $D_{p_{-L}}\psi(p,x;u) + \partial_x\psi(p,x;u)\psi(p,x;u)' = \mathfrak{S}(p,x,u) = (s^{jk}(p,x,u))_{1\le j\le L-1, 1\le k\le L-1}$, where p_{-L} denotes the price vector without the L-th price. If this submatrix of second derivatives is not nsd, the complete matrix involving all L equations cannot be nsd either. The null hypothesis in the underlying unobservable heterogeneous populations is hence that

$$b'\mathfrak{S}(p,x,u)b \le 0,$$

for all $b \in \mathcal{S}_{L-1}$ (where here and throughout $\mathcal{S}_d$ denotes the d-dimensional unit sphere) and any $(u,p,x) \in \mathcal{U}\times \mathbb{R}^{L+1}$. The hypothesis of Slutsky nsd translates therefore to an inequality

[1] This is the income concept commonly used in consumer demand. It is motivated by the assumption of separability of preferences over time and from other decisions (e.g., the labor supply decision). We use the phrases "total expenditure", "income" and "wealth" interchangeably throughout this paper.

restriction on an $L-1$ dimensional system of equations, where each equation is characterized by a nonseparable model.

Obviously, this hypothesis in question has to hold for any preference ordering u. However we do not observe the individual's preference ordering u, and only observe a B dimensional vector of household covariates (denoted q). Specifically, we assume to have n *iid* observations on individuals from an underlying heterogeneous population characterized by random variables U, Y, X, P, Q which have a nondegenerate joint distribution $F_{U,Y,X,P,Q}$.

The question of interest is now as follows: What can we learn from the observable part of this distribution, i.e. $F_{Y,X,P,Q}$, about whether $\mathfrak{S}$ is negative semidefinite across a heterogeneous population, for all values of (p, x, u). In Hoderlein (2010), we consider testing negative semidefiniteness in such a setting with mean and second moment regressions only. However, these lower order moment regressions have the disadvantage that they use only one feature of $F_{Y,X,P,Q}$, and not the entire distribution. Therefore, in this paper we propose to exploit the distributional information by using all the α-quantiles of the conditional distribution of observables, which (with varying α) employ all the information that may be obtained from the data about the economic hypothesis of interest, i.e., Slutsky negative semidefiniteness of the heterogeneous population in question.

There are two immediate difficulties now, and solving them is the major innovation this paper introduces. The first is how to relate a specific economic property in the (unobservable) world of nonseparable functions to observable regression quantiles. The second one is how to use quantiles in systems of equations. The solution for the second difficulty is to consider linear combinations of the dependent variable, i.e. $Y(b) = b'Y$ for all $b \in \mathcal{S}_{L-1}$ and consider the respective conditional α-quantiles of this quantity. This can be thought of as an analogue to the Cramer-Wold device, and is a strategy that is feasible more generally, e.g., when testing omission of variables. As b and α vary, we exploit the entire distribution of observables.

The solution to the first of these two difficulties involves obviously identifying assumptions. To this end, since we are dealing with nonseparable models we require full conditional independence, i.e., we require that $U \perp (P, X) \,|Q$, or versions of this assumption that control for endogeneity. These assumptions are versions of the "selection on observables" assumptions in the treatment effect literature. Essentially they require that, in every subpopulation defined by $Q = q$, preferences as well as prices and income be independently distributed. Although endogeneity is not relevant for our application, our treatment covers the control function approach to handle endogeneity in nonseparable models discussed in Altonji and Matzkin (2005), Imbens and Newey (2009) or Hoderlein (2010), by simply adding endogeneity controls V to the set of household control variables Q. From now on, we denote by W the set of all observable

right hand side variables, i.e., (P', X, Q'), and potentially in addition V, if we are controlling for endogeneity.

Under this assumption and some regularity conditions, our first main contribution is as follows: Let $k(\alpha, b|w)$ denote the conditional α-quantile of $Y(b)$ given $P = p, X = x, Q = q$, ∇_p the gradient with respect to p and ∂_x the partial derivative with respect to x. Then we establish that

$$\mathfrak{S}(p, x, u) \ nsd \quad \Longrightarrow \quad \nabla_p k(\alpha, b|w)'b + \partial_x k(\alpha, b|w) k(\alpha, b|w) \leq 0 \qquad (1.1)$$
$$\text{for all } \ (\alpha, b) \in (0, 1) \times \mathcal{S}_{L-1},$$

and almost everywhere (w, u). The right hand side of this implication depends entirely on observable quantities. Consequently, we can test a null hypothesis in the underlying (unobservable) heterogeneous population model by considering testable implications on observable quantities. These implications take the form of projections, and crucially depend on the employed independence assumption. Observe that the original hypothesis and the testable counterpart are very different objects. In particular, the underlying structural model is characterized by a vector valued functional ψ - which in general will not be a quantile unless U is a scalar, and ψ is strictly monotonic - while the observed model is an univariate quantile, for every b.

The only material assumption that we require to relate the observable object and the underlying heterogeneous population is the conditional independence assumption $U \perp (P, X) \,|Q$

We would like to emphasize at this point that no other material assumption on the function ψ or the distribution of the data has entered the model. In particular, we have not assumed any monotonicity or triangularity assumption; there can be infinitely many unobservables, and they can enter in arbitrarily complicated form. Thus, another way to view this paper is as a clarification about what is the most we can learn from data about an economic hypothesis in a heterogeneous population under economically well specified independence assumptions alone. In this setup characterized by excess heterogeneity, at best averages over the heterogeneous population are identified.

Our second main contribution is proposing a quantile regression based nonparametric test statistic. Specifically, we suggest and analyze the sample counterpart to the right hand side of (1.1), i.e., we consider the test statistic

$$\sup_{\alpha, b} \widehat{\nabla_p k}(\alpha, b|w)'b + \widehat{\partial_x k}(\alpha, b|w)\widehat{k}(\alpha, b|w),$$

where the hats denote appropriate nonparametric estimators. Our main contribution in this part of the paper is the derivation of the large sample properties. We show weak convergence of

a corresponding standardized stochastic process to a Gaussian process and obtain an asymptotically valid hypothesis test. Moreover, we also propose a bootstrap version of our test statistic which is based on a centered version of the stochastic process to avoid the generation of bootstrap observations under the null. We adapt the well known idea of residual bootstrap for our specific model and provide arguments for the validity of the bootstrap. Nonparametric tests involving quantiles are surprisingly scant, and we list the closest references in the following paragraph. Specifically, in a system of equations setup we are the first to propose a quantile based test of an economic hypothesis, and to implement such a test using real world data.

Literature Testing the key integrability constraints that arise out of utility maximization dates back at least to the early work of Stone (1954), and has spurned the extensive research on (parametric) flexible functional form demand systems (e.g., the Translog, cf. Jorgenson, Lau and Stoker (1982), and the Almost Ideal, cf. 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 modelling unobserved heterogeneity. More closely related to our approach is Lewbel (2001) who analyzes integrability constraints in a purely exogenous setting, but does not use distributional information nor suggests or implements an actual test. An alternative method for checking some integrability constraints is revealed preference analysis, see Blundell, Browning and Crawford (2003), and references therein.

While our approach extends earlier work on demand systems, it is very much a blueprint for testing all kinds of economic hypothesis in systems of equations. Due to the nonseparable framework we employ, our approach extends the recent work on nonseparable models - in particular Hoderlein (2010), Hoderlein and Mammen (2007), Imbens and Newey (2009), Matzkin (2003). When it comes to dealing with unobserved heterogeneity, there are two strands in this literature: The first assumes triangularity and monotonicity in the unobservables (Chesher (2003), Imbens and Newey (2009), Matzkin (2003)). The triangularity and monotonicity assumptions are, however, rather implausible for consumer demand, because in general the multivariate demand function is a nonmonotonic function of an infinite dimensional unobservable - the individuals' preference ordering - and all equations depend on this object.

Hence we follow the second route. Extending earlier work in Hoderlein (2010), Hoderlein and Mammen (2007) establish interpretation of the derivative of the conditional quantile (a scalar valued function!) if there is more than one unobservable[2]. The upshot of this work is that in a

[2]Specifically, assume that $Y = \phi(X, A)$ is the structural model and that $k_\alpha^{Y|X}(x)$ denotes the conditional α-quantile of Y given X. The main result of Hoderlein and Mammen (2007) is that $\partial_x k_\alpha^{Y|X}(x) = \mathbb{E}[\partial_x \phi(X, A)|X =$

world of excess heterogeneity, at best conditional average effects are identified, see also Altonji and Matzkin (2005) and Imbens and Newey (2009). We use a similar notion on this paper, but apply it to systems of equations in which we analyze economic questions. To the best of our knowledge we are the first to analyze major economic restrictions using quantiles, as well as provide guidance in nonseparable system of equations without triangularity and monotonicity. Other than the novelty in the identification framework, we would also like to emphasize that the test statistic is in itself a contribution which generalizes beyond the specific economic question at hands. Specifically, our test pioneers nonparametric hypothesis tests for economic hypothesis using quantiles in cross section setups. In the statistics literature, the closest work we are aware of includes the testing procedures proposed by Zheng (1998), Sun (2006), Escanciano and Velasco (2010), and Dette, Volgushev and Wagener (2011).

Structure of the Paper: The exposition of this paper is as follows. In the next section, we introduce our model formally, state some assumptions, derive the main identification result (1.1) in a rigorous fashion, and discuss it. In the third section, we propose a nonparametric test for the economic hypothesis of Slutsky nsd, analyze its large sample behavior and propose a bootstrap procedure to derive the critical values. We investigate the performance of the bootstrap procedure for moderate sample sizes in a simulation study in section 4. In the fifth section, we apply these concepts to Canadian expenditure 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 regularity assumptions, proofs, graphs and summary statistics.

2 Deriving Quantile Restrictions of Economic Behavior in a Heterogeneous Population

2.1 Building Blocks of the Model and Assumptions

Our model of consumer demand in a heterogeneous population consists of several building blocks. As is common in consumer demand, we assume that - for a fixed preference ordering - there is a causal relationship between quantities, a real valued random L-vector denoted by Y, and regressors of economic importance, namely prices P and total expenditure X, real valued random vectors of length L and 1, respectively. As already mentioned, we assume a linear budget constraint as well as nonsatiation of preferences, which implies the adding up

$x, Y = k_\alpha^{Y|X}(x)]$.

constraint. To avoid the singularity associated with this constraint, we impose it from the outset, so that we obtain an $L-1$ vector of dependent variables. We also assume homogeneity of degree zero, so that we can omit the L-th price (that of the residual category), all prices are relative to the residual price, and total expenditure is normalized to be real total expenditure. To capture the notion that preferences vary across the population, we assume that there is a random variable $U \in \mathcal{U}$, where $\mathcal{U}$ is a Borel space[3], 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 K-vector Q. Hence, we let $U = \vartheta(Q, A)$, where ϑ is a fixed $\mathcal{U}$-valued map 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. To fix ideas, think of A as the genom of an individual.

As already mentioned, we want to allow for infinitely many individual preference orderings each of which may be characterized by an infinite dimensional parameter. Therefore we formalize the heterogeneous population as $Y = \psi(P, X, U) = \phi(P, X, Q, A)$, for a general map ϕ and an infinite dimensional vector A. Obviously, neither ϕ nor the distribution of A are nonparametricaly identified. Still, for any fixed value of A, say a_0, we obtain a demand function having standard properties. Moreover, to show that our approach can handle endogeneity which arises because economic decisions are related, we treat the more general case and introduce additional instruments, denoted S. To keep the exposition simple, we focus on the scalar case, i.e. we assume that it is only X that is endogenous, and that there is exactly one additional instrument S. This is in line with the demand literature, see Lewbel (1999), but may be easily adapted to cover endogeneity in prices. Note, however, that in our application endogeneity does not play a major role, and this is done solely for expositional purposes.

The first assumption collects all definitions and specifies the DGP formally:

Assumption 1. *Let $(\Omega, \mathcal{F}, P)$ be a complete probability space on which are defined the random vectors $A : \Omega \to \mathcal{A}$, $\mathcal{A} \subseteq \mathbb{R}^\infty$, and $(Y, P, X, Q, S, V) : \Omega \to \mathcal{Y} \times \mathcal{P} \times \mathcal{X} \times \mathcal{Q} \times \mathcal{S} \times \mathcal{V}$, $\mathcal{Y} \subseteq \mathbb{R}^{L-1}, \mathcal{P} \subseteq \mathbb{R}^{L-1}, \mathcal{X} \subseteq \mathbb{R}, \mathcal{Q} \subseteq \mathbb{R}^K, \mathcal{S} \subseteq \mathbb{R}$, $\mathcal{V} \subseteq \mathbb{R}$, with L and K finite integers, such that*

$$\begin{aligned} Y &= \phi(P, X, Q, A), \\ X &= \mu(P, Q, S, V) \end{aligned}$$

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

where $\phi : \mathcal{P} \times \mathcal{X} \times \mathcal{Q} \times \mathcal{A} \to \mathcal{Y}$ *and* $\mu : \mathcal{P} \times \mathcal{Q} \times \mathcal{S} \times \mathcal{V} \to \mathcal{X}$ *are Borel functions, and realizations of* (Y, P, X, Q, S) *are observable, whereas those of* (A, V) *are not. Moreover,* μ *is invertible in its last argument, for every* (p, q, s).

Assumption **A1** defines the nonparametric demand system with (potentially) endogenous regressors as a system of nonseparable equations. These models are called nonseparable, because they do not impose an additive specification for the unobservable random terms (in our case A). They have been subject of much interest in the recent econometrics literature (Chesher (2003), Matzkin (2003), Altonji and Matzkin (2005), Imbens and Newey (2009), Hoderlein (2010), 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 five approaches. As is demonstrated there, in the absence of strict monotonicity of ϕ in A, the function ϕ 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 V be solved for because these residuals have to be employed in a control function fashion. In the application, we specify μ to be the conditional mean function, and consequently V to be the additive mean regression residuals, but this is only one out of several possibilities.

Given that we have all major elements of our model defined and in place, we specify the independence conditions required for identification. We introduce the notation $Z = (Q', V)'$, and $\mathcal{Z} = \mathcal{Q} \times \mathcal{V}$.

Assumption 2. *The random vectors* A *and* (P, X) *are independent conditional on* Z.

Assumption **A2** is the only material assumption that we require in order to identify the marginal effect of interest, and thus being able to test the economic restriction of interest, in our case Slutsky negative semidefiniteness. Therefore it merits a thorough discussion: Assume for a moment all regressors were exogenous, i.e. $S \equiv X$, $Z \equiv Q$ and $V \equiv 0$. Then this assumption states that wealth and prices be independently distributed of unobserved heterogeneity A, 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 for all levels of wealth. 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. Finally, if it is suspected that regressors and unobservables are not independent, we may still introduce instruments in a control function fashion.

In the following, we make use of the notation $W = (P', X, Q', V')'$. Moreover, for a given vector b we introduce the conditional quantile $k(\alpha, b|w)$ of the distribution of $Y(b) = b'Y$ given $W = w$, i.e. for $0 < \alpha < 1$ the quantity $k(\alpha, b|w)$ is defined by $\mathbb{P}(Y(b) \leq k(\alpha, b|w) \mid W = w) = \alpha$, or, upon substitution,

$$\mathbb{P}(b'\phi(p, x, q, A) \leq k(\alpha, b|w) \mid W = w) = \alpha,$$

where $w = (p', x, q', v')$. We will also require a set of regularity assumptions, largely differentiability and boundedness conditions, which can be found in the appendix (see assumption **A3**). Given these assumptions and notations, we concentrate first on the relation of theoretical quantities and identified (and hence estimable) objects. Specifically, we are concerned with the question of how quantiles allow inference on key elements of economic theory. In particular, we want to learn about negative semidefiniteness of the Slutsky matrix. In the standard consumer demand setup we consider, the Slutsky matrix in the underlying heterogeneous population (defined by ϕ, x and v), takes the form

$$\mathfrak{S}(p, x, u) = D_p\psi(p, x, u) + \partial_x\psi(p, x, u)\psi(p, x, u)' \quad \text{for all } (p, x, u) \in \mathcal{P} \times \mathcal{X} \times \mathcal{U}.$$

The following theorem provides an answer about what we can learn from regression quantiles.

Theorem 1. *Let assumptions **A1–A3** hold. Then*

$$\mathfrak{S}(p, x, u) \ \textit{nsd} \Rightarrow \nabla_p k(\alpha, b|w)'b + \partial_x k(\alpha, b|w)k(\alpha, b|w) \leq 0$$

for all $(\alpha, b) \in (0, 1) \times \mathcal{S}_{L-1}$, *and all* $(p, x, u, z) \in \mathcal{P} \times \mathcal{X} \times \mathcal{U} \times \mathcal{Z}$. *Moreover,*

$$\nabla_p k(\alpha, b|w)'b \ + \ \partial_x k(\alpha, b|w)k(\alpha, b|w) \quad = \quad \mathbb{E}\left[\mathfrak{S} \mid P = p, X = x, Z = z, Y(b) = k(\alpha, b|w)\right].$$

Discussion of Theorem 1: This result establishes the link between negative semidefiniteness in a heterogeneous population characterized by excess heterogeneity, and the joint distribution of the data as characterized by the various regression quantiles of $Y(b)$ for all $b \in \mathcal{S}_{L-1}$. As already discussed above, it characterizes all we can learn from data about the economic hypothesis of interest, and also characterizes the object by which we can, i.e., $g(\alpha; b, p, x, z) = \nabla_p k(\alpha, b|p, x, z)'b + \partial_x k(\alpha, b|p, x, z)k(\alpha, b|p, x, z)$. To see the economic content of $g(\alpha; p, x, z)$, note that the second part of the theorem establishes that this quantity is related to the LASD

of Hoderlein (2010), and Hoderlein and Mammen (2007). To continue with our economic example, suppose again we were given data on consumption, wealth, "education in years" and "gender" as above. Then by considering $g(\alpha; b, p, x, z)$, we may identify, for fixed b, the average over the Slutsky matrix of a subpopulation characterized by a certain level of prices, wealth and covariates, e.g., all female high school graduates earning 50K, whose value of a weighted average of their demands has a certain value. However, due to the *excess heterogeneity* we are not able to identify the Slutsky matrix of every single individual. Thus, since we consider $b'Y$, our averages use more information than simply the one embedded in the regressors; by variation of b we consider these averages across the entire distribution of the data and hence use all information available.

There are limitations to this approach that employs minimal assumptions and these limitations suggest interesting directions for future research. One particular issue is the following: The equation (1.1) provides testable implications for each fixed value (p, x, z), and all quantiles. In this setup, even the entire information contained in the joint distribution of observables does not suffice to trace out the distribution of unobservables - there is excess heterogeneity. We would like to emphasize that absent additional functional form or homogeneity assumptions, this is the most we can learn from the data.

This has the following implication for the economic hypothesis at hand: Suppose that we found the property not rejected for a certain subpopulation. There may still be a part of the subpopulation that does not behave rationally, but the violations are overcompensated by the part of the subpopulation that behaves rationally. As such, we may suffer from low power. While this is true for every subpopulation, to assess whether rationality holds across the entire population, we will evaluate the hypothesis at a set of representative positions (i.e., subpopulations, actually random draws from the sample). As a characterization of our empirical results, we will also give the percentage of positions at which we reject the null. Since we are evaluating the population at a large number of local averages over relatively small neighborhoods, we may expect low power at any given position to nevertheless result in occasional rejections. In the spirit of our approach, we could argue that the percentage of rejections obtained using the best approximations given the data is itself a best approximation to the underlying fraction of irrational individuals, given all the information to our disposal. We conclude this cautionary discussion by emphasizing that this is not a defect of our specific approach: In the absence of additional information - either through assumptions or additional data - this is simply the limit of what we can learn about an economic property in a heterogeneous population.

3 From Hypothesis to Test Statistic

In this section we assume we have observed independent data (Y_i, P_i, X_i), $i = 1, \ldots, n$, with the same distribution as $(Y, P, X) \in \mathbb{R}^{L-1} \times \mathbb{R}^{L-1} \times \mathbb{R}$. We do not treat the additional conditioning on Z_i as standard nonparametric results extend straightforwardly to this setup by simply adapting the rates appropriately and - in the case of V - having generated regressors, whose estimation does not affect first order asymptotics if we are willing to assume enough smoothness in the regression of endogenous regressors on instruments, see Sperlich (2009). Since including Z only makes the notation more cumbersome, all results extend in a straightforward fashion, and we consider a homogeneous subpopulation in our application, we omit it from now on.
For each fixed $w = (p, x) \in \mathbb{R}^{L-1} \times \mathbb{R}$ we want to test negative semi-definiteness of the Slutsky matrix. To this end we use the notations

$$\begin{aligned} k_{p_\ell}(\alpha, b \mid w) &= \partial_{p_\ell} k(\alpha, b \mid w) \\ k_x(\alpha, b \mid w) &= \partial_x k(\alpha, b \mid w), \end{aligned}$$

where $k(\alpha, b \mid w)$ denotes again the α-quantile of the conditional distribution of $Y(b) = b'Y$, given $W = (P, X) = w$. As we have seen from the previous subsection, the null hypothesis of rationality transforms to

$$\begin{aligned} H_0 \quad : \quad & \sum_{\ell=1}^{L-1} b_\ell k_{p_\ell}(\alpha, b \mid w) + k_x(\alpha, b \mid w) k(\alpha, b \mid w) \leq 0 \\ & \forall \alpha \in (0,1), b = (b_1, \ldots, b_{L-1})' \in \mathcal{S}_{L-1}. \end{aligned} \tag{3.1}$$

Now let $\mathbb{A}$ be a closed subset of $(0, 1)$. We define the test statistic by

$$T_n = \sqrt{nh^{L+2}} \sup_{\alpha \in \mathbb{A}, b \in \mathcal{S}_{L-1}} R_n(\alpha, b \mid w), \tag{3.2}$$

where

$$R_n(\alpha, b \mid w) = \sum_{\ell=1}^{L-1} b_\ell \hat{k}_{p_\ell}(\alpha, b \mid w) + \hat{k}_x(\alpha, b \mid w) \hat{k}(\alpha, b \mid w). \tag{3.3}$$

Here, with notations $W_i = (P_i', X_i)'$ and $\tau_\alpha(u) = u(\alpha - I\{u < 0\})$, the estimators are obtained from the kernel quantile estimation approach

$$\begin{aligned} & (\hat{\mu}_0, \hat{\mu}_1, \hat{\mu}_2) \\ = \; & \arg \min_{\substack{(\mu_0, \mu_1, \mu_2) \in \\ \mathbb{R} \times \mathbb{R}^L \times \mathbb{R}^{L \times L}}} \sum_{i=1}^n \tau_\alpha \Big(Y_i' b - \mu_0 - \mu_1'(W_i - w) - (W_i - w)' \mu_2 (W_i - w) \Big) K\Big(\frac{W_i - w}{h} \Big) \end{aligned}$$

as $\hat{k}(\alpha, b \mid w) = \hat{\mu}_0$, $\hat{k}_{p_\ell}(\alpha, b \mid w) = \hat{\mu}_{1,\ell}$, $\ell = 1, \ldots, L-1$, and $\hat{k}_x(\alpha, b \mid w) = \hat{\mu}_{1,L}$, see Yu and Jones (1998), Koenker (2005) or Hoderlein and Mammen (2009) among others.
Under the conditions stated in assumption **A4** in the appendix, $R_n(\alpha, b \mid w)$ consistently estimates

$$R(\alpha, b \mid w) = \sum_{\ell=1}^{L-1} b_\ell k_{p_\ell}(\alpha, b \mid w) + k_x(\alpha, b \mid w) k(\alpha, b \mid w)$$

and we have the following weak convergence result.

Theorem 2. *The process*

$$\sqrt{nh^{L+2}}\Big(R_n(\alpha, b \mid w) - R(\alpha, b \mid w)\Big)_{\alpha \in \mathbb{A}, b \in \mathcal{S}_{L-1}}$$

converges (for w fixed) weakly in $\ell^\infty(\mathbb{A} \times \mathcal{S}_{L-1})$ to a Gaussian process $G(\alpha, b \mid w)_{\alpha \in \mathbb{A}, b \in \mathcal{S}_{L-1}}$ with covariance

$$\begin{aligned}
& Cov\Big(G(\alpha, b \mid w), G(\tilde{\alpha}, \tilde{b} \mid w)\Big) \\
= \; & \Big[\mathbb{P}\Big(Y(b) \le k(\alpha, b \mid w), Y(\tilde{b}) \le k(\tilde{\alpha}, \tilde{b} \mid w) \;\Big|\; (P, X) = w\Big) \\
& - F_{Y(b)|P,X}\Big(k(\alpha, b \mid w) \;\Big|\; w\Big) F_{Y(\tilde{b})|P,X}\Big(k(\tilde{\alpha}, \tilde{b} \mid w) \;\Big|\; w\Big)\Big] \\
& \times \frac{\int K^2(\overline{p}, \overline{x})(b'\overline{p} + k(\alpha, b \mid w)\overline{x})(\tilde{b}'\overline{p} + k(\tilde{\alpha}, \tilde{b} \mid w)\overline{x})\, d(\overline{p}, \overline{x})}{f_{Y(b)|P,X}(k(\alpha, b \mid w) \mid w) f_{Y(\tilde{b})|P,X}(k(\tilde{\alpha}, \tilde{b} \mid w) \mid w) f_{P,X}(w)(\int u^2 \kappa(u)\, du)^2},
\end{aligned}$$

where $f_{P,X}$ denotes the density of (P, X) and $f_{Y(b)|P,X}(\cdot \mid w)$, $F_{Y(b)|P,X}(\cdot \mid w)$ denote the conditional density and distribution function of $Y(b)$, given $(P, X) = w$, respectively.

From the theorem we cannot obtain the asymptotic distribution of the test statistic T_n under H_0 to approximate critical values for the test, but we obtain the asymptotic distribution of a "centered" version, i. e.

$$\tilde{T}_n = \sqrt{nh^{L+2}} \sup_{\alpha \in \mathbb{A}, b \in \mathcal{S}_{L-1}} (R_n(\alpha, b \mid w) - R(\alpha, b \mid w)) \tag{3.4}$$

both under the null hypothesis and under fixed alternatives. As corollary from Theorem 2 it follows that, for each $c \in \mathbb{R}$,

$$\mathbb{P}(\tilde{T}_n > c) \overset{n \to \infty}{\longrightarrow} \mathbb{P}(\sup_{\alpha, b} G(\alpha, b) > c). \tag{3.5}$$

Because of the specific structure of the null hypothesis, $H_0 : R(\cdot, \cdot \mid w) \le 0$, it follows that $\mathbb{P}(T_n > c) \le \mathbb{P}(\tilde{T}_n > c)$ under H_0 and, hence, we obtain an asymptotically level γ test by

rejecting H_0 whenever $T_n > c_\gamma$, where $\mathbb{P}(\sup_{\alpha,b} G(\alpha, b) > c_\gamma) = \gamma$. Note that this construction may lead to conservative testing procedures.

Because of the complicated covariance structure in Theorem 2, in applications we suggest to approximate the asymptotic quantile c_γ by the following bootstrap procedure.

The idea of any bootstrap procedure for testing hypotheses for independent samples is to generate new data (Y_k^*, P_k^*, X_k^*), $k = 1, \ldots, n$, (independent, given the original sample (Y_i, P_i, X_i), $i = 1, \ldots, n$) which follow a model as similar as possible to the original data model under the null hypothesis. However, given the specific structure of the null hypothesis it is not clear how to generate data that fulfill H_0. Instead we use bootstrap versions of the "centered" test statistic $\tilde{T}_n$ as follows. As is common in regression models, to generate new data we keep the covariates and define $(P_k^*, X_k^*) = (P_k, X_k)$. For each fixed covariate (P_k, X_k) we then generate Y_k^* from a distribution which approximates the conditional distribution of Y, given $(P, X) = (P_k, X_k)$, i. e. $F_{Y|P,X}(\cdot \mid P_k, X_k)$. Note that for one-dimensional observations Y this method coincides with the bootstrap procedure suggested in Hoderlein and Mammen (2009). To estimate the conditional distribution $F_{Y|P,X}(\cdot \mid P_k, X_k)$ we apply the usual kernel approach and define

$$\hat{F}_{Y|P,X}(y \mid P_k, X_k) = \sum_{i=1}^{n} I\{Y_{i,1} \leq y_1, \ldots, Y_{i,L-1} \leq y_{L-1}\} \frac{k\left(\frac{W_i - W_k}{g}\right)}{\sum_{j=1}^{n} k\left(\frac{W_j - W_k}{g}\right)}$$

for $y = (y_1, \ldots, y_{L-1})$, $Y_i = (Y_{i,1}, \ldots, Y_{i,L-1})$, $i = 1, \ldots, n$, with kernel function k and bandwidth g. The bootstrap version of the test statistic T_n defined in (3.2) is

$$T_n^* = \sqrt{nh^{L+2}} \sup_{\alpha \in \mathbb{A}, b \in \mathcal{S}_{L-1}} \Big(R_n^*(\alpha, b \mid w) - R_n(\alpha, b \mid w) \Big), \tag{3.6}$$

where R_n is defined in (3.3) and R_n^* is defined analogously, but based on the bootstrap sample (Y_k^*, P_k^*, X_k^*), $k = 1, \ldots, n$. Both under the null hypothesis and under alternatives, the conditional distribution of T_n^*, given the original sample, approximates the distribution of $\tilde{T}_n$ defined in (3.4) and (3.5). We approximate the asymptotic quantile c_γ by $c_{n,\gamma}^*$, where

$$\mathbb{P}\Big(T_n^* > c_{n,\gamma}^* \;\Big|\; (Y_i, P_i, X_i), i = 1, \ldots, n \Big) = \gamma.$$

With the same argument as before (under regularity conditions given in appendix II in **A4** and some assumptions on the kernel k and bandwidth g) we obtain an asymptotically level γ test by rejecting H_0 whenever $T_n > c_{n,\gamma}^*$. Moreover, the test is consistent (against alternatives representing the negation of H_0), because under a fixed alternative $\sup_{b,\alpha} R_n(b, \alpha \mid w)$ converges to $\sup_{b,\alpha} R(b, \alpha \mid w) > 0$ in probability, such that T_n converges to infinity. On the other hand, $c_{n,\gamma}^*$ approximates the $(1-\gamma)$-quantile of $\tilde{T}_n$ which converges to $c_\gamma \in \mathbb{R}$ by (3.5). Hence, the probability of rejection converges to one.

4 Monte Carlo Experiments

To analyze the finite sample performance of our test statistic and to get a feeling for the behavior of the test in our application, we simulate data from a joint distribution that has similar features at least in terms of observables. We specify the DGP to be a linear random coefficients specification, which is arguably the most straightforward model of a heterogeneous population, and choose the distributions of coefficients such that under the null the entire population is rational, while under the alternative there is a fraction that does not have a negative semidefinite Slutsky matrix. We apply the test developed above to answer the question whether there is a significant fraction that does not behave rationally.

More specifically, to test whether H_0 in (3.1) is valid, we first estimate T_n as described in equation (3.2) and the following passage. In particular, in (3.3) we use a local quadratic estimator with a product kernel where the individual kernels are standard univariate Epanechnikov kernels. The bandwidth is selected by using a slightly larger bandwidth than the bandwidth that is selected by cross validation of the corresponding nonparametric median regression, to account for the fact that we largely use derivatives. Since our simulation setup is calibrated to the application, we use somewhat differing variances for the respective regressors, see below. To account for this when choosing the bandwidth, we scale the individual bandwidths for every dimension by dividing through the standard deviation σ_{W_j} of the respective regressor W_j, i.e., the bandwidth for dimension j is of the form $h_j = h\sigma_{W_j}$, and h is selected by cross validation to be approximately 0.25.

To obtain the distribution of T_n^* from (3.6) we apply the same estimators, but now use 100 bootstrap samples generated as described in section 3. The bandwidth we use is slightly smaller than the one used in estimation (by a factor of 0.8). Since b is supposed to have unit length, the grid of bs is chosen such that $b_1^2 + b_2^2 + b_3^2 = 1$, and $b_j^2 \in \{0, 0.1, ..., 0.9, 1\}$, for all j. We have evaluated all quantile regressions at a set of 15 equally spaced positions, i.e., $\mathbb{A} = \{0.05, ..., 0.95\}$ of α-quantiles of $Y(b)$, for every b. We design the test to have nominal level of 0.05, as mentioned above we will have a slight size distortion by construction. This completes the description of the econometric tools we apply, and we now turn to the details of the data generating process: To keep the results simple and transparent, we assume that the data generating process contains no income effect, and only prices $P = (P_1, P_2, P_3, P_4)'$, $P \sim \mathcal{N}(\mu_P, \Sigma_P)$, where $\mu_P = (1, 1, 1, 1)'$ and $\Sigma_P = diag(1.6, 1.1, 1.1, 1.6)$. This is justified as the income effects turn out to be rather minor in our application, because the individual categories account for moderate sections of total expenditures only, and the income elasticities are not very large, see the next section.

In addition to the stochastic regressors, the DGP contains random heterogeneity parameters

$A = (A_0, A_1, A_2, A_3, A_4)'$, independent of P, for which we specify that $A \sim \mathcal{N}(\mu_A, \Sigma_A)$. We choose $\mu_A = (20, 20, 20, 30, 10)'$; the second to fourth reproduce the average absolute values of the own price effects of prices 1-3 in our application, and $\Sigma_A = diag(5, 5, 5, 5, 2)$ in the low variance setup, and $\Sigma_A = diag(5, 10, 5, 5, 2)$. Those values are chosen, together with the random coefficients A_0 and A_4, to roughly reproduce the R^2 in the respective regressions. Moreover, we assume that there is a fixed parameter $\lambda \in \mathbb{R}^+$ which we introduce to model deviations from the null of rationality (negative semidefinitness).

The quantities $Y = (Y_1, Y_2, Y_3)'$ are then generated as

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \begin{bmatrix} 1.5(A_0 - 20) + 20 \\ A_0 \\ 0.8A_0 \end{bmatrix} + \underbrace{\begin{bmatrix} -A_1 + \lambda & 0.4A_2 & 0 \\ 0.2A_1 & -A_2 & 0 \\ 0 & 0 & -A_3 \end{bmatrix}}_{Slutsky} \begin{bmatrix} P_1 \\ P_2 \\ P_3 \end{bmatrix} + \begin{bmatrix} A_4 \\ A_4 \\ A_4 \end{bmatrix} P_4,$$

Observe that in this model the Slutsky matrix is given by:

$$\mathfrak{S}(A) = \begin{bmatrix} -A_1 + \lambda & 0.4A_2 & 0 \\ 0.2A_1 & -A_2 & 0 \\ 0 & 0 & -A_3 \end{bmatrix},$$

For $\lambda = 0$, the entire population is rational, i.e., its Slutsky matrix is negative semidefinite. However, as λ increases, parts of the population become indefinite, which in the language of this paper means that they cease to be rational. The following table illustrates how the proportion of the population which is not rational increases, as illustrated by the largest (nonnegative) eigenvalue of the symmetrized version of the Slutsky matrix:

λ	0	10	15	17.5	20	22.5	25	30
% Population not Rational	0.0001	0.041	0.250	0.443	0.647	0.822	0.928	0.996

This exercise corresponds to a mean shift in the distribution of the first random coefficient. Alternatively, more nonrational types may also be generated by increasing the variance, however, for the above reason (proximity to our application), we focus on this specification. From this model, we draw an *iid* sample of $n = 3000$ observations.

The following table shows the result of our procedure as described in the previous paragraphs, with $n = 3000$ observations, a bandwidth of 0.25, and at mean values of the regressors, $\mu = (1, 1, 1, 1)$. The size of the test is 0.03, and the power is displayed at the following alternatives:

λ	10	15	17.5	20	22.5	25	30
% Population not Rational	0.041	0.250	0.443	0.647	0.822	0.928	0.996
Power	0.120	0.280	0.410	0.590	0.770	0.900	0.970

Figure 1 in the appendix gives a graphical representation of these results, as function of λ, while fig.2 displays the power as a function of the proportion of the population that violates rationality. If we increase the bandwidth somewhat, the results do not change significantly, however, eventually the power decreases. In contrast, the results are sensitive to the choice of bandwidth in the sense that we get size distortions if we choose too small a bandwidth, more precisely, the size becomes 0.110. The following table shows the result of decreasing h from 0.25 to 0.2.

λ	10	15	17.5	20	22.5	25	30
% Population not Rational	0.041	0.250	0.443	0.647	0.822	0.928	0.996
Power	0.190	0.310	0.390	0.580	0.770	0.880	0.940

The results are graphically compared in fig. 3 in the appendix. Finally, to show the consistency of the test, we display the result with $n = 6000$. At the now somewhat smaller bandwidth of $h = 0.2$, we obtain a size of 0.04 and the following results on power:

λ	10	15	17.5	20	22.5	25	30
% Population not Rational	0.041	0.250	0.443	0.647	0.822	0.928	0.996
Power	0.240	0.380	0.530	0.640	0.800	0.910	0.980

A graph showing again the increase in power is displayed in fig. 4 in the appendix. Obviously the test exhibits power and is consistent. Moreover, the power should also be seen in the context that the true model is a linear random coefficient model. It is well known from the nonseparable models literature that there is a tight connection between quantiles and nonlinear models with one monotonic heterogeneity factor, a class of models that is very different from our DGP. Finally, note, that these deviations from rationality are rather mild; even amongst the non rational people most of the individuals exhibit small positive values of the Slutsky matrix. It is easy to design Monte Carlo experiments in which a larger positive eigenvalue of a small fraction of the population would generate even more power for our test, however, we do not believe that this represents a feature of our application, and hence desist from doing so here. In summary, we would also not expect that our test exhibits too much power against this specification, and the results appear reasonable.

However, this specification of the DGP allows easier comparison with standard practise, without which our results are hard to interpret. Since, to the best of our knowledge, there is no work which even mildly resembles what we propose, we compare our approach with a stylized version of standard parametric methods in this setup. Specifically, we run a parametric regression using a FGLS approach. Given our setup, the FGLS estimator incorporates the entire information

about the model, as it is the ML estimator[4]. We then compute the largest eigenvalue of the Slutsky matrix, say, $\hat{\chi}$. Finally, we implement a bootstrap procedure to obtain standard errors, which is appropriate if there is no multiplicity of eigenvalues. The point estimate of the largest eigenvalue for $\lambda = 17.5$ is -0.871, with 95% confidence interval $[-1.189; -0.549]$. For $\lambda = 22.5$, the point estimate is 3.809, with 95% confidence interval $[3.480; 4.129]$. To perform a one sides test, we moreover construct critical values $CV = 2\hat{\chi} - Q(.95, F_{\hat{\chi}^* - \hat{\chi}})$, where $Q(.95, F_{\hat{\chi}^* - \hat{\chi}})$ denotes the 95% quantile of $F_{\hat{\chi}^* - \hat{\chi}}$. For $\lambda = 17.5$, we obtain an average $CV = -1.193$, and zero rejections, whereas for $\lambda = 22.5$, we obtain an average $CV = 3.488$ with universal rejections. Standard practise would hence not reject, if 45% of the population is not rational and always reject with 80% nonrational individuals.

An interesting conclusion out of this comparison is that standard practise in parametric models picks up deviations from rationality only if it has finally an impact on the mean. In contrast, our test exhibits power already if the fraction of the population being not rational is rather small (e.g., with 4% non rational individuals and $n = 6000$, a typical size in a cross section application, we reject one quarter of times). Rather than significant parts of the population being wildly non rational, we believe that at best parts of the population deviate from rationality in a rather mild fashion, and given the simulation results we feel comfortable that our test will be able to detect these deviations in an application, at least if we perform it at a large set of independent positions. Let us therefore now turn to such an application.

5 Empirical Implementation

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

5.1 Econometric Specifications and Methods

5.2 Data

The data used in this paper come from the following public use sources: (1) the Canadian Family Expenditure Surveys 1969, 1974, 1978, 1982, 1984, 1986, 1990, 1992 and 1996; (2) the Surveys of Household Spending 1997, 1998 and 1999; and (3) Pendakur (1999). Price and expenditure data are available for 12 years in 5 regions (Atlantic, Quebec, Ontario, Prairies

[4]We are indebted to Dennis Kristensen for this suggestion

and British Columbia) yielding 60 distinct price vectors. Prices are normalized so that the price vector facing residents of Ontario in 1986 is $(1, ..., 1)$. The data is hence a repeated cross section; every individual is only sampled once.

Table 4 gives (unweighted) summary statistics for 6952 observations of rental-tenure unattached individuals aged 25-64 with no dependents. Estimated nonparametric densities (not reported, but available from the authors) for log-prices and log-expenditures are approximately normal, as is typically found in the demand literature. Analysis is restricted to these households to minimize demographic variation in preferences. Demographic variation could be added to the model by conditioning all levels, log-price derivatives and log-expenditure derivatives on demographic covariates. Rather than pursue this strategy, we use a sample with very limited demographic variation. Haag, Hoderlein and Pendakur (2009) do not find evidence of endogeneity in this data set; we hence do not pursue any control function strategy in this paper (even though it would be straightforward given the theoretical results).

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

Table 1: The Data		Min	Max	Mean	Std Dev
Expenditure Shares	Food at Home	0.02	0.84	0.23	0.11
	Food Out	0.00	0.75	0.11	0.10
	Rent	0.01	0.97	0.54	0.14
	Clothing	0.00	0.61	0.12	0.09
Total Expenditure		640	40270	8596	4427
Prices	Food at Home	0.2436	1.4000	1.0095	0.315
	Food Out	0.2328	1.7050	1.1260	0.412
	Rent	0.2682	1.4423	0.9312	0.321
	TotExp in 10K$	0.0640	4.0270	0.8600	0.428

As another descriptive means to characterize the data, we report the results of three log-log mean regressions. In our framework, the coefficients do at best reflect some averages of effects (in the case of endogeneity, not even this is warranted), and are only meant to be informative, as well as provide consistency checks for the data.

Table A.1 reports the result of the regression of log food in on the four log prices and log total expenditure. The own price elasticity is around -1, which is in line with reported results for other good (e.g., gasoline, see Hausman and Newey (1995)). Food in and food out are strong substitutes, as is to be expected, while the substitution patterns with the other goods are much less pronounced. In fact, if anything, rent and food in seem to be complements; a fact that may be related to a common lifestyle. Finally, the rather low total expenditure elasticity is also well documented in other studies, see Lewbel (1999) for an overview. It is shared by other necessities, e.g., gasoline (Hausman and Newey (1995)). Given the large number of observations, the estimates are fairly precise.

The other two regressions mirror these findings. Food out is less of a necessity than food in, consequently, its own price elasticity is larger in absolute values, see table A.2. Moreover, there is only a very weak relationship between food out and rent, which is confirmed by both regressions in tables A.2 and A.3. In both instances, the total expenditure elasticities are larger in absolute value, which is not surprising given that satiation is less of an issue with food out and rent, which consequently have more of a luxury character. The own price elasticities are both solidly negative and dominate in absolute value.

Linear quantile regressions essentially reproduce these results. As an example, we have included the median regression of food out on the same variable in table A.4. Obviously, the results are very comparable. The variations across quantiles are also surprisingly low. In particular, the dominant negative diagonal in the Slutsky matrix is well preserved throughout the range of quantiles. Though our analysis is nonparametric, uses levels as opposed to logarithms, and considers the compensated as opposed to the uncompensated price effect matrix, the descriptive results foreshadow our main finding: the data are largely consistent with the Slutsky matrix being negative semidefinite across the population.

5.3 Methodology and Results

As discussed in previous sections, we consider the statistic R_n defined in (3.3), where for the kernel quantile estimator we use a product of standard Epanechnikov kernels. The bandwidth is selected by using a slightly larger bandwidth than the bandwidth that was selected by cross validation of the corresponding nonparametric median regression, to account for the fact that

we largely use derivatives. More specifically, we scaled the individual bandwidths for every dimension by dividing through the standard deviation of the respective regressor, i.e., the bandwidth for dimension j is of the form $h_j = h\sigma_{W_j}$, and h was selected by cross validation to be 0.85. Given that features of the simulation were chosen to coincide with the data, this suggests that the model deviates significantly from the random coefficient model with normally distributed coefficients; if we force the bandwidth to be closer to zero we obtain very unstable behavior.

Standard errors were obtained via the bootstrap, as described in section 3, the bandwidth we used was slightly smaller than the one used in estimation (by a factor of 0.8). We have used 100 bootstrap replications, the grid of bs was selected as in section 4, and we have evaluated all quantile regressions at the same equally spaced grid of 15 support points, i.e., the set $\mathbb{A} = \{0.05, ..., 0.95\}$ of quantiles of $Y(b)$, for every b satisfying the above normalization condition. Moreover, we impose homogeneity of degree zero, which means that we work with three relative prices (relative to clothing), and with normalized total expenditure.

When we apply this machinery to the data, we obtain the following results: On a grid of 64 values of w whose outer limits contain roughly 95% of the population in terms of W, the point estimates for $T_n(\alpha, b \mid w)$ are largely negative, with only three insignificant exceptions. The point estimates range from -5.166 to 2.874, with a mean of -1.472 and a median of -1.638. Most of the values are clustered around -1.5. Table 2 provides results at selected values of W.

Table 2: Results	Values of W	T_n	$T_n - c_\gamma$
	$(1.1, 1.1, 1.1, 1.0)$	-0.979	-5.602
	$(0.7, 1.1, 1.1, 1.0)$	-3.766	-13.753
	$(1.3, 1.1, 1.1, 1.0)$	-1.653	-8.029
	$(1.1, 0.7, 1.1, 1.0)$	-0.403	-5.112
	$(1.1, 1.3, 1.1, 1.0)$	-1.64	-8.294
	$(1.1, 1.1, 0.7, 1.0)$	$+2.224$	-3.467
	$(1.1, 1.1, 1.3, 1.0)$	-1.672	-8.559
	$(0.9, 0.9, 0.9, 1.0)$	-1.828	-8.199
	$(1.3, 0.7, 0.7, 1.0)$	$+2.874$	-3.105
	$(0.7, 0.7, 0.7, 1.0)$	-0.924	-13.074

These affirmative results are then corroborated if we employ our formal analysis, as outlined and analyzed above and in the finite sample exercise. In particular, neither of the three positive

values is significantly positive, the difference $T_n - c_\gamma$ is always below -2.8. Given the results we can of course not reject rationality at any of the subpopulations (i.e., values of independent variables) considered. Since we can think of these positions as being asymptotically independent, even with rather small fraction of the population we would expect occassional violations of rationality. The fact that these almost never occur allows us to reach the conclusion that rationality is widely accepted in this data set.

6 Summary and Outlook

Rationality of economic agents is the central paradigm of economics. Yet, within this paradigm individuals can vary widely in their actual behavior; only the qualitative properties of individual behavior are constrained, but not the heterogeneity across individuals. Indeed, in many data sets there are large differences in observed consumer choices even for individuals which are equal in terms of their observed household covariates, like age, gender, educational background etc.

One of the core qualitative restrictions of Economics is the negative semidefiniteness of the Slutsky matrix. It is the core restriction arising out of (static) utility maximization subject to a linear budget constraint. This paper discusses how to test this property using the entire conditional distribution of the data when individuals are assumed to be rational, but otherwise are allowed to be completely different from each other. The key insight is that quantile regressions based on linear combinations of the original dependent variables may be used to test the property of interest. While some of the insights of this paper may be generalized to related questions like omission of variables in system of equations (e.g., supply and demand systems), our focus in this paper remains on negative semidefiniteness. We derive the large sample behavior of the specific test statistic we consider, and analyze its small sample behavior in a simulation study.

Our empirical findings emphasize the affirmative tendency in the studies of rationality in Blundell, Pashardes and Weber (1993), Hoderlein (2010), and Haag, Hoderlein and Pendakur (2009). Using Canadian data, we find negative semidefiniteness to be widely accepted. As a caveat, we have seen from the simulation study that the test may not detect very small fractions of irrational individuals in the population. While this may be a minor issue given the large set of independent conditions we are considering, it should nevertheless be seen as encouragement to perform a similar analysis with other data sets. Also, it may be interesting to search for semiparametric structures, e.g., random coefficient models that allow to test negative semidefiniteness in a heterogeneous population with tests of higher power. Similarly, the structure of

panel data may be exploited with new models that are either more efficient due to a repeated observations structure, or less prone to model misspecification as they allow for correlated time invariant factors. We hope that this research will encourage future work in this direction.

Acknowledgements The authors have received helpful comments from Andrew Chesher, Roger Koenker, Dennis Kristensen, Arthur Lewbel, Rosa Matzkin, Ulrich Mueller, as well as seminar participants at Boston College, Princeton and the Conference on Nonparametrics and Shape Constraints at Northwestern University. We are particularly indebted to Krishna Pendakur to provide us with the data. We would also like to thank Martina Stein, who typed parts of this manuscript with considerable technical expertise. This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823) of the German Research Foundation (DFG).

Appendix

Appendix I - Log-Log Regressions

Table A.1: Standard log-log Regression, Dependent Variable Food in, OLS estimates

	Coefficient	Standard Error	t-Value	p-Value
intercept	2.77650	0.01299	213.733	< 0.000001
log price food in	-1.06962	0.09856	-10.852	< 0.000001
log price food out	1.02478	0.12840	7.981	< 0.000001
log price rent	-0.38870	0.04568	-8.509	< 0.000001
log price clothing	-0.57160	0.13624	-4.195	< 0.000001
log total expenditure	0.53954	0.01653	32.635	< 0.000001

R^2	0.195			
adjusted R^2	0.194			
$d.o.f.$	5793			

Table A.2: Standard log-log Regression, Dependent Variable Food Out, OLS estimates

	Coefficient	Standard Error	t-Value	p-Value
intercept	2.49445	0.02295	108.673	< 0.000001
log price food in	1.01379	0.17415	5.821	< 0.000001
log price food out	-1.64400	$1 - 0.22688$	-7.246	< 0.000001
log price rent	-0.05472	0.0807	-0.678	0.4980145
log price clothing	-1.37527	0.24074	-5.713	< 0.000001
log total expenditure	1.73781	0.02921	59.490	< 0.000001

R^2	0.4176			
adjusted R^2	0.4171			
$d.o.f.$	5793			

Table A.3: Standard log-log Regression, Dependent Variable Rent, OLS estimates

	Coefficient	Standard Error	t-Value	p-Value
intercept	3.913712	0.008012	488.461	< 0.000001
log price food in	-0.229934	0.060790	-3.782	0.000157
log price food out	-0.009231	0.079195	-0.117	0.907214
log price rent	-0.737412	0.028174	-26.173	< 0.000001
log price clothing	0.326560	0.084032	3.886	0.000103
log total expenditure	0.927734	0.010197	90.982	< 0.000001

R^2	0.604			
adjusted R^2	0.603			
$d.o.f.$	5793			

Table A.4: Linear Quantile log-log Regression, Dependent Variable Food Out, Median

	Coefficient	Standard Error	t-Value	p-Value
intercept	2.61447	0.02739	95.46997	< 0.000001
log price food in	1.03697	0.21259	4.87776	< 0.000001
log price food out	-1.65577	0.28291	-5.85254	< 0.000001
log price rent	-0.04897	0.09972	-0.49107	0.62340
log price clothing	-1.48867	0.29715	-5.00975	< 0.000001
log total expenditure	1.84493	0.03762	49.03717	< 0.000001

Appendix II - Assumptions

Regularity Conditions for Theorem 1

Assumption 3. *For fixed values* $(p^*, x^*, z^*) \in \mathcal{P} \times \mathcal{X} \times \mathcal{Z}$ *and* $0 < \alpha < 1$ *we will make use of the following assumptions:*
(i) The conditional distribution of $Y(b)$ *given* (P, X, Z) *is absolutely continuous w.r.t. the Lebesgue measure for* (p, x) *in a neighborhood of* (p^*, x^*) *and for* $z = z^*$.
(ii) The conditional density $f_{Y(b)|PXZ}(y|p, x, z^*)$ *of* $Y(b)$ *given* (P, X, Z) *is continuous in* (y, p, x) *at the point* $(k_\alpha(w^*; b), w^*)$.
(iii) The conditional density $f_{Y(b)|PXZ}(y|w^*, z^*)$ *of* $Y(b)$ *given* $W = w^*$ *is bounded in* $y \in \mathbb{R}$.
(iv) Let $w_1 = (p', x)'$. $k_\alpha(w_1, z; b)$ *is partially differentiable with respect to any component of* w_1 *at* $(w_1, z) = (w_1^*, z^*)$. *Moreover, there exist measurable functions* Δ_k, $k = 1, .., L$, *satisfying*

$$\mathbb{P}[|\phi(w_{1k}^* + \delta, w_{-1k}^*, z^*, A) - \phi(w_1^*, z^*, A) - \delta\Delta_k(A)| \geq \varepsilon\delta_k \mid X = x^*] = o(\delta_k)$$

for $\delta_k \to 0$ *and fixed* $\varepsilon > 0$. *We write* $\partial_{w_{1k}}\phi(w_1^*, z^*, a)$ *for* $\Delta_k(a)$ *and* $\partial_{w_k}\phi$ *for* $\Delta_k(A)$, *for all* $k = 1, .., L$.
(v) The conditional distribution of $(Y(b), \partial_{w_{1k}}\phi)$, *given* (P, X, Z), *is absolutely continuous w.r.t. the Lebesgue measure for* $(p, x, z) = (p^*, x^*, z^*)$ *and all* (b, k). *For the conditional density* $f_{Y(b),\partial_{w_{1k}}\phi|PXZ}$ *of* $(Y(b), \partial_{w_{1k}}\phi)$ *given* (P, X, Z), *we require that* $f_{Y(b),\partial_{w_{1k}}\phi|PXZ}(y, y'|p^*, x^*, z^*) \leq Cg(y')$, *where* C *is a constant and* g *a positive density on* $\mathbb{R}$ *with finite mean (i.e.* $\int |y'|g(y')\, dy' < \infty$*).*

Assumptions for Theorem 2

Assumption 4. *In the following let* $\mathbb{A}$ *denote a closed subset of* $(0, 1)$ *and* $\mathcal{S}_{L-1} = \{b \in \mathbb{R}^{L-1} \mid ||b|| = 1\}$. *Let* $\mathcal{W} \subset \mathbb{R}^L$ *denote the support of* $W = (P, X)$ *and let* $w \in \mathcal{W}$ *be fixed. Let* $F_{Y(b)|W}(\cdot \mid \tilde{w})$ *and* $f_{Y(b)|W}(\cdot \mid \tilde{w})$ *denote the conditional distribution and density functions of* $Y(b) = b'Y$, *given* $W = \tilde{w}$, *respectively. For* $\delta > 0$, $M > 0$, *let* $C_M^\delta(\mathcal{W})$ *denote the class of smooth functions with partial derivatives up to order* $\underline{\delta}$ *(the greatest integer smaller than* δ*) uniformly bounded by* M, *whose highest partial derivatives are Lipschitz continuous of order* $\delta - \underline{\delta}$ *[see van der Vaart and Wellner (1996), p. 154/155]. We will make use of the following assumptions:*
(i) Let K *be an* L*-variate product kernel of the univariate bounded symmetric density* κ *with bounded support such that* $\int_{\mathbb{R}} \kappa(u)u^2\, du < \infty$.
(ii) Let $h = h_n$ *denote a sequence of positive bandwidths such that* $nh^{L(1+\frac{1}{L+5})} \longrightarrow \infty$, $nh^{L+6} \longrightarrow 0$ *for* $n \to \infty$.

(iii) Let $\mathcal{W}$ be compact and convex and let the density f_W of W be bounded away from zero in a neighborhood $\mathcal{N}$ of w and be bounded and continuous. Let the second moments of W exist.
(iv) Assume that the function $\mathbb{A} \times \mathcal{S}_{L-1} \to \mathbb{R}$, $(\alpha, b) \mapsto k(\alpha, b \mid w)$ is bounded by a constant C and uniformly continuous.
(v) For each fixed $\tilde{w} \in \mathcal{W}$ let the function $\mathbb{A} \times \mathcal{S}_{L-1} \to \mathbb{R}$, $(\alpha, b) \mapsto F_{Y(b)|W}(k(\alpha, b \mid w) \mid \tilde{w})$ be continuous. Assume that there exist $M > 0$ and $\delta > \frac{L}{2}$ such that for all $c \in \mathbb{R}$, $|c| \leq C$ the function $\mathcal{W} \to \mathbb{R}$, $\tilde{w} \mapsto F_{Y(b)|W}(c \mid \tilde{w})$ belongs to $C_M^\delta(\mathcal{W})$. Let the function $(y, \tilde{w}) \mapsto f_{Y(b)|W}(y \mid \tilde{w})$ be uniformly continuous in y and $\tilde{w}$ and be bounded and bounded away from zero. Further assume that $\inf_{(\alpha,b,\tilde{w}) \in \mathbb{A} \times \mathcal{S}_{L-1} \times \mathcal{N}} f_{Y(b)|W}(k(\alpha, b \mid \tilde{w}) \mid \tilde{w}) > 0$.

Appendix III - Proof of Theorem 1

Let $A(\omega)$, $\omega \in \Omega$, denote any random matrix. If $b'A(\omega)b \leq 0$ for all $\omega \in \Omega$ and all $b \in \mathbb{R}^L$, then, upon taking expectations w.r.t. an arbitrary probability measure F, it follows that

$$\int b'A(\omega)bF(d\omega) \leq 0 \Leftrightarrow b' \int A(\omega)F(d\omega)b \leq 0, \text{ for all } b \in \mathcal{S}_{L-1}.$$

From this, $\mathfrak{S}(p, x, u)$ *nsd* $\Rightarrow \mathbb{E}\left[\mathfrak{S} | P = p, X = x, Z = z, Y(b) = k(\alpha, b|w)\right]$ *nsd* for all $(p, x; z) \in \mathcal{P} \times \mathcal{X} \times \mathcal{Z}$ is immediate. Let $\mathbb{E}\left[\mathfrak{S} | P, X, Z, Y(b)\right] = B$, and note that since the definition of negative semidefiniteness of a square matrix B of dimension $L-1$ involves the quadratic form, $b'Bb \leq 0$. Next, observe that

$$\begin{aligned} B &= \mathbb{E}\left[\mathfrak{S} | P = p, X = x, Z = z, Y(b) = k(\alpha, b|w)\right] \\ &= \mathbb{E}\left[D_p\phi | P = p, X = x, Z = z, Y(b) = k(\alpha, b|w)\right] \\ &\quad + \mathbb{E}\left[\partial_x\phi\phi' | P = p, X = x, Z = z, Y(b) = k(\alpha, b|w)\right] \\ &= B_1 + B_2, \end{aligned}$$

as well as

$$\begin{aligned} b'B_1b &= b'\mathbb{E}\left[D_p\phi | P = p, X = x, Z = z, Y(b) = k(\alpha, b|w)\right] b \\ &= \mathbb{E}\left[\nabla_p\eta | P = p, X = x, Z = z, Y(b) = k(\alpha, b|w)\right] b \\ &= \nabla_p k(\alpha, b|w)'b, \end{aligned}$$

where $\eta = \phi' b$ and the last equality follows from the theorem in Hoderlein and Mammen (2007). Moreover, by arguments as above

$$\begin{aligned} b'B_2 b &= \mathbb{E}\left[b'\partial_x \phi\phi' b | P = p, X = x, Z = z, Y(b) = k(\alpha, b|w)\right] \\ &= \mathbb{E}\left[\partial_x \eta\eta | P = p, X = x, Z = z, Y(b) = k(\alpha, b|w)\right] \\ &= \mathbb{E}\left[\partial_x \eta | P = p, X = x, Z = z, Y(b) = k(\alpha, b|w)\right] k(\alpha, b|w) \\ &= \partial_x k(\alpha, b|w) k(\alpha, b|w), \end{aligned}$$

where the second to last equality follows from $Y(b) = \eta$ and the last equality again from Hoderlein and Mammen (2007). Using the definition of negative semidefiniteness,

$$b'\left(B_1 + B_2\right) b \leq 0,$$

we obtain,

$$\nabla_p k(\alpha, b|w)' b + \partial_x k(\alpha, b|w) k(\alpha, b|w) \leq 0$$

and the result follows. □

Appendix IV - Proof of Theorem 2

Auxiliary results

Lemma 6.1. *Uniformly with respect to $\alpha \in \mathbb{A}, b \in \mathcal{S}_{L-1}$ we have*

$$\sqrt{nh^{L+2}}\Big(R_n(\cdot, \cdot \mid w) - R(\cdot, \cdot \mid w)\Big) = \tilde{R}_n + o_p(1),$$

where

$$\tilde{R}_n(\alpha, b) = \sqrt{nh^{L+2}}\Big(\sum_{\ell=1}^{L-1} b_\ell(\hat{k}_{p_\ell}(\alpha, b \mid w) - k_{p_\ell}(\alpha, b \mid w)) + (\hat{k}_x(\alpha, b \mid w) - k_x(\alpha, b \mid w))k(\alpha, b \mid w)\Big).$$

Proof of Lemma 6.1. The assertion follows from uniform consistency of the estimator $\hat{k}_x(\alpha, b \mid w)$ and

$$\sup_{\alpha, b} |\hat{k}(\alpha, b \mid w) - k(\alpha, b \mid w)| = o_p(\frac{1}{\sqrt{nh^{L+2}}}).$$

□

Lemma 6.2. *(Bahadur expansion) Uniformly with respect to $\alpha \in \mathbb{A}, b \in \mathcal{S}_{L-1}$ we have*

$$\begin{aligned} \tilde{R}_n(\alpha, b) &= -\frac{1}{\sqrt{nh^{L+2}}} \frac{1}{f_{Y(b)|P,X}(k(\alpha, b \mid w) \mid w) f_W(w) \int u^2 \kappa(u)\, du} \\ &\quad \sum_{i=1}^{n} K\Big(\frac{W_i - w}{h}\Big)\Big(I\{Y_i(b) \leq k(\alpha, b \mid w)\} - F_{Y(b)|W}(k(\alpha, b \mid w) \mid W_i)\Big) \\ &\quad \times \Big[b'(P_i - p) + k(\alpha, b \mid w)(X_i - x)\Big] + o_p(1). \end{aligned}$$

Proof of Lemma 6.2. Obviously it is sufficient to prove the representation for $\hat{k}_{p_\ell}(\alpha, b \mid w) - k_{p_\ell}(\alpha, b \mid w)$, $\ell = 1, \ldots, L-1$, and $\hat{k}_x(\alpha, b \mid w) - k_x(\alpha, b \mid w))k(\alpha, b \mid w)$ in the representation of $\tilde{R}_n(\alpha, b)$ in Lemma 6.1, and the assertion of the Lemma follows by building linear combinations. For the sake of brevity, we restrict ourselves to the first component of the vector P; all other cases are treated in the same way.

Using similar arguments as in the proof of Theorem A.1 by Hoderlein and Mammen (2009) we obtain the expansion

$$\begin{aligned} R_n^{(1)}(\alpha, b) &= \sqrt{nh^{L+2}}(\hat{k}_{p_1}(\alpha, b \mid w) - k_{p_1}(\alpha, b \mid w) \\ &= -\frac{1}{\sqrt{nh^{L+2}}} \frac{1}{f_{Y(b)|P,X}(k(\alpha, b|w)|w)\ f_W(w) \int u^2 \kappa(u) du} \\ &\quad \times \sum_{i=1}^n K(\frac{W_i - w}{n})\ \{I\{Y_i^*(b) < 0\} - \alpha\}\ (P_{1i} - p_1) + o_p(1) \end{aligned}$$

uniformly with respect to $\alpha \in \mathbb{A},\ b \in \mathcal{S}_{L-1}$, where P_{1i} denotes the first component of the vector P_i $(i = 1, \ldots, n)$ and the random variables $Y_i^*(b)$ are defined by $Y_i^*(b) = Y_i(b) - \overline{k}_i(\alpha, b|w)$ with

$$\overline{k}_i(\alpha, b, w) = k(\alpha, b|w) - \nabla_w\ k(\alpha, b|w)(W_i - w) - \frac{1}{2}(W_i - w)^T \nabla_{ww}\ k(\alpha, b|w)(W_i - w)\ .$$

This yields the decomposition

$$R_n^{(1)}(\alpha, b) = -\frac{S_{1,n}(\alpha, b) + S_{2,n}(\alpha, b) + S_{3,n}(\alpha, b)}{f_{Y(b)|P,X}(k(\alpha, b|w)|w) f_W(w) \int u^2 \kappa(u) du}\ , \tag{6.1}$$

where

$$\begin{aligned} S_{1,n}(\alpha, b) &= \frac{1}{\sqrt{nh^{L+2}}} \sum_{i=1}^n K(\frac{W_i - w}{h})(P_{1i} - p_1)\Big\{I\{Y_i(b) < k(\alpha, b|w)\} - \mathbb{P}(Y_i(b) < k(\alpha, b|w)|W_i)\Big\} \\ S_{2,n}(\alpha, b) &= \frac{1}{\sqrt{nh^{L+2}}} \sum_{i=1}^n K(\frac{W_i - w}{h})(P_{1i} - p_1)\Big\{[\ I\{Y_i^*(b) < 0\} - I\{Y_i(b) < k(\alpha, b|w)\}] \\ &\quad -[\mathbb{P}(Y_i^*(b) < 0|W_i) - \mathbb{P}(Y_i(b) < k(\alpha, b|w)|W_i)]\Big\} \\ S_{3,n}(\alpha, b) &= \frac{1}{\sqrt{nh^{L+2}}} \sum_{i=1}^n K(\frac{W_i - w}{h})(P_{1i} - p_1)\{\mathbb{P}(Y_i^*(b) < 0|W_i) - \alpha\}\ . \end{aligned}$$

Obviously we have $E[S_{2,n}(\alpha, b)] = 0$ and for the variance of this random variable we obtain by a straightforward but tedious calculation

$$\begin{aligned} Var(S_{2,n}(\alpha, b)) &\le \frac{1}{h^{L+2}} E\Big[K^2(\frac{W_i - w}{h})(P_{1i} - p_1)^2 E[(I\{Y_i^*(b) < 0\} - I\{Y_i(b) < k(\alpha, b|w)\})^2|W_i]\Big] \\ &\le \frac{1}{h^{L+2}} E\Big[K^2(\frac{W_i - w}{h})(P_{1i} - p_1)^2 \Big|F_{Y(b)|W}(k(\alpha, b|w)|W_i) - F_{Y(b)|W}(\overline{k}_i(\alpha, b|w)|W_i)\Big|\Big] \\ &= O(h^3), \end{aligned}$$

which shows that $S_{2,n}(\alpha,b) = o_p(1)$ for all (α,b). Moreover, the centered process $S_{2,n}$ can be represented as

$$S_{2,n}(\alpha,b) \;=\; \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi_{n,\alpha,b}(W_i,Y_i),$$

where with the same methods as in the proof of Theorem 2 below the class of functions $\mathcal{F}_n = \{\varphi_{n,\alpha,b} \mid (\alpha,b)\in\mathbb{A}\times\mathcal{S}_{L-1}\}$ can be embedded into a class $\tilde{\mathcal{F}}_n = \{\tilde{\varphi}_{n,c} \mid c\in\mathcal{C}\}$ which is Donsker in the sense that the process

$$S_n(c) \;=\; \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\varphi}_{n,c}(W_i,Y_i),\; c\in\mathcal{C},$$

weakly converges to a Gaussian process $S(c)$, $c\in\mathcal{C}$. Hence, weak convergence also follows for the restricted function class, i. e. for $S_{2,n}(\alpha,b)$, $(\alpha,b)\in\mathbb{A}\times\mathcal{S}_{L-1}$. However, as we have shown that $S_{2,n}(\alpha,b) = o_p(1)$ for all $(\alpha,b)\in\mathbb{A}\times\mathcal{S}_{L-1}$, the limit is degenerate and uniform convergence follows, i. e.

$$\sup_{(\alpha,b)\in\mathbb{A}\times\mathcal{S}_{L-1}} |S_{2,n}(\alpha,b)| = o_p(1).$$

Similarly, it follows for the term $S_{3,n}(\alpha,b)$ (note that $\mathbb{P}(Y_i(b) < k(\alpha,b|W_i)|W_i) = \alpha$)

$$E[S_{3,n}(\alpha,b)] = O(\sqrt{nh^{L+6}}) = o(1)\;; \qquad Var(S_{3,n}(\alpha,b)) = O(h^6) = o(1)$$

which shows that $S_{3,n}(\alpha,b) = o_p(1)$ for all (α,b). Uniform convergence, i. e. $\sup_{(\alpha,b)\in\mathbb{A}\times\mathcal{S}_{L-1}} |S_{2,n}(\alpha,b)| = o_p(1)$, can be shown with the same arguments as before.
Therefore we obtain from (6.1) that

$$R_n^{(1)}(\alpha,b) = -\frac{\sum_{i=1}^{n} K(\frac{W_i-w}{h})(P_{1i}-p_1)\{I\{Y_i(b) < k(\alpha,b|w)\} - F_{Y(b)|W}(k(\alpha,b|w)|W_i)\}}{\sqrt{nh^{L+2}} f_{Y(b)|P,X}(k(\alpha,b|w)|w) f_W(w)\int u^2\kappa(u)du} + o_p(1)\;,$$

uniformly with respect to (α,b), and a similar argument for the remaining terms in $\tilde{R}_n(\alpha,b)$ yields the assertion of Lemma 6.2. □

Proof of Theorem 2 Consider for fixed w the process

$$\bar{R}_n(\alpha,b) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi_{n,\alpha,b}(W_i,Y_i)$$

where

$$\begin{aligned}\varphi_{n,\alpha,b}(W,Y) \;=\; & \frac{1}{h^{(L+2)/2}} K\Big(\frac{W-w}{h}\Big)\Big(I\{b'Y \le k(\alpha,b \mid w)\} - F_{Y(b)|W}(k(\alpha,b \mid w) \mid W)\Big) \\ & \times \Big[b'(P-p) + k(\alpha,b \mid w)(X-x)\Big]\end{aligned}$$

such that

$$\tilde{R}_n(\alpha, b) = -\frac{1}{f_{Y(b)|P,X}(k(\alpha, b \mid w) \mid w) f_{P,X}(w) \int u^2 \kappa(u)\, du} \bar{R}_n(\alpha, b) + o_p(1)$$

Consider the function class

$$\mathcal{F}_n = \{\varphi_{n,\alpha,b} \mid \alpha \in \mathbb{A}, b \in \mathcal{S}_{L-1}\}.$$

Under our assumptions, $k(\alpha, b \mid w)$ is bounded. Hence, $\mathcal{F}_n$ is a subclass of $\mathcal{G}_n - \mathcal{H}_n$ defined as follows,

$$\begin{aligned}
\mathcal{G}_n &= \Big\{(W, Y) \mapsto \frac{1}{h^{(L+2)/2}} K\Big(\frac{W-w}{h}\Big) I\{Y \in H\} c'(W-w) \;\Big|\; H \text{ halfspace in } \mathbb{R}^{L-1}, c \in \mathbb{R}^L, \\
&\qquad ||c|| \le C\Big\} \\
\mathcal{H}_n &= \Big\{(W, Y) \mapsto \frac{1}{h^{(L+2)/2}} K\Big(\frac{W-w}{h}\Big) \varphi(W) c'(W-w) \;\Big|\; \varphi \in C_M^{\delta}(\mathcal{W}), c \in \mathbb{R}^L, ||c|| \le C\Big\}
\end{aligned}$$

for some suitable $C > 0$. Let s denote a constant such that $||W - w|| \le s$ with probability 1 (under our assumptions the support of the covariates W is bounded).

The classes $\mathcal{G}_1 = \{Y \mapsto I\{Y \in H\} \mid H \text{ halfspace in } \mathbb{R}^{L-1}\}$ and $\mathcal{G}_2 = \{W \mapsto c'(W-w) \mid c \in \mathbb{R}^L, ||c|| \le C\}$ are VC-classes with (under our assumptions) constant envelopes $G_1 \equiv 1$, $G_2 \equiv Cs$ and VC-dimension $L+1$ and $L+2$, respectively, see van der Vaart and Wellner (1996), Problem 14, p. 152, and Lemma 2.6.15, p. 146. Hence, from Theorem 2.6.7, van der Vaart and Wellner (1996), p. 141, we obtain the following result for covering numbers,

$$\sup_Q N(\epsilon, \mathcal{G}_1, L_2(Q)) \le d_1 \epsilon^{-2L}, \quad \sup_Q N(\epsilon Cs, \mathcal{G}_2, L_2(Q)) \le d_2 \epsilon^{-2(L+1)}$$

for constants d_1, d_2 not depending on the arbitrary distribution Q of (W, Y) (such that second moments exist), nor on ϵ.

From Kosorok (2008), proof of Theorem 9.15, p. 158, it follows for the covering number of the product class $\mathcal{G}_1\mathcal{G}_2$ with envelope Cs that

$$\sup_Q N(\epsilon Cs, \mathcal{G}_1\mathcal{G}_2, L_2(Q)) \le \sup_Q N\Big(\frac{\epsilon}{2}, \mathcal{G}_1, L_2(Q)\Big) \sup_Q N\Big(\frac{\epsilon}{2} Cs, \mathcal{G}_2, L_2(Q)\Big) \le d_1 d_2 (\frac{\epsilon}{2})^{-2(2L+1)}.$$

Now, for

$$g_n(Y, W) = \frac{1}{h^{(L+2)/2}} K\Big(\frac{W-w}{h}\Big)$$

we have $\mathcal{G}_n = g_n \mathcal{G}_1 \mathcal{G}_2$ with envelope $G_n = g_n Cs$ and again from Kosorok (2008), proof of Theorem 9.15., it follows that

$$\sup_Q N(\epsilon ||g_n||_{Q,2} Cs, \mathcal{G}_n, L_2(Q)) \le \sup_Q N(\epsilon Cs, \mathcal{G}_1\mathcal{G}_2, L_2(Q)) \le d_1 d_2 (\frac{\epsilon}{2})^{-2(2L+1)}.$$

Hence, it follows that

$$\int_0^{\delta_n} \sup_Q \sqrt{\log N(\epsilon ||G_n||_{Q,2}, \mathcal{G}_n, L_2(Q))}\, d\epsilon \longrightarrow 0 \text{ for } \delta_n \to 0, n \to \infty.$$

Similarly, the same can be shown for the class $\mathcal{H}_n$ where we use the same argumentation as before with g_n and $\mathcal{G}_2$ as before, but $\mathcal{G}_1 = C_M^\delta(\mathcal{W})$ with constant envelope M, such that the covering number fulfills

$$\sup_Q \log N(\epsilon, \mathcal{G}_1, L_2(Q)) \leq d\Big(\frac{1}{\epsilon}\Big)^{L/\delta},$$

see van der Vaart and Wellner (1996), Theorem 2.7.1, p. 155.

Altogether, we obtain

$$\int_0^{\delta_n} \sup_Q \sqrt{\log N(\epsilon ||F_n||_{Q,2}, \mathcal{F}_n, L_2(Q))}\, d\epsilon \longrightarrow 0 \text{ for } \delta_n \to 0, n \to \infty.$$

for the function class $\mathcal{F}_n$, which is a subclass of $\mathcal{G}_n - \mathcal{H}_n$, see Lemma 9.14 Kosorok (2008) and Lemma 16 by Nolan and Pollard (1987), where the envelope is defined as

$$F_n(W,Y) = \frac{2C}{h^{L/2}} K\Big(\frac{W-w}{h}\Big) \frac{||W-w||}{h}.$$

Let $\mathbb{P}$ denote the distribution of (W,Y) under the assumptions listed in **A**4. Then,

$$\begin{aligned}
\mathbb{P}F_n^2 &= E[F_n(W,Y)^2] = \frac{2C}{h^L} \int_{\mathbb{R}^L} K^2\Big(\frac{t-w}{h}\Big) \Big|\Big|\frac{t-w}{h}\Big|\Big|^2 f_W(t)\, dt \\
&= 2C \int_{\mathbb{R}^L} K^2(u) ||u||^2 f_W(w+hu)\, du = O(1).
\end{aligned}$$

Similarly, one has $\mathbb{P}F_n^2 I\{F_n \geq \eta\sqrt{n}\} \to 0$ for every $\eta > 0$ applying $nh^L \to \infty$.

It remains to show that

$$\sup_{\substack{||b-\tilde{b}|| \leq \delta_n \\ |\alpha - \tilde{\alpha}| \leq \delta_n}} \mathbb{P}(\varphi_{n,\alpha,b} - \varphi_{n,\tilde{\alpha},\tilde{b}})^2 \longrightarrow 0 \text{ for every } \delta_n \to 0, n \to \infty.$$

A straightforward estimation yields for $||b - \tilde{b}|| \leq \delta_n$, $|\alpha - \tilde{\alpha}| \leq \delta_n$, $\delta_n \to 0$,

$$\begin{aligned}
&\mathbb{P}(\varphi_{n,\alpha,b} - \varphi_{n,\tilde{\alpha},\tilde{b}})^2 \\
\leq\ & 2E\Big[\frac{1}{h^{L+2}} K^2\Big(\frac{W-w}{h}\Big)\Big((b - \tilde{b}'(P-p) + (k(\alpha,b \mid w) - k(\tilde{\alpha},\tilde{b} \mid w))(X-x)\Big)^2\Big] \\
&+ E\Big[\frac{1}{h^{L+2}} K^2\Big(\frac{W-w}{h}\Big)\Big(I\{Y(b) \leq k(\alpha,b \mid w)\} - I\{Y(\tilde{b}) \leq k(\tilde{\alpha},\tilde{b} \mid w)\} \\
&\qquad - F_{Y(b)|W}(k(\alpha,b \mid w) \mid W) + F_{Y(\tilde{b})|W}(k(\tilde{\alpha},\tilde{b} \mid w) \mid W)\Big)^2 C^2 ||W-w||^2\Big] \\
\leq\ & 4\Big(\delta_n^2 + \sup_{\substack{||b-\tilde{b}|| \leq \delta_n \\ |\alpha - \tilde{\alpha}| \leq \delta_n}} (k(\alpha,b \mid w) - k(\tilde{\alpha},\tilde{b} \mid w))^2\Big) E\Big[\frac{1}{h^L} K^2\Big(\frac{W-w}{h}\Big) \frac{||W-w||^2}{h^2}\Big] \\
&+ 4E\Big[\frac{1}{h^L} K^2\Big(\frac{W-w}{h}\Big) \frac{C^2 ||W-w||^2}{h^2} \Big|F_{Y(b)|W}(k(\alpha,b \mid w) \mid W) - F_{Y(\tilde{b})|W}(k(\tilde{\alpha},\tilde{b} \mid w) \mid W)\Big|\Big] \\
=\ & o(1).
\end{aligned}$$

Then weak convergence of the (centered) process $\bar{R}_n$ to a Gaussian process follows from Theorem 2.11.22 in van der Vaart and Wellner (1996), p. 220.
The calculation of the covariance structure is straightforward and therefore omitted. □

References

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

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

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

[4] CHESHER, A. (2003), Identification in nonseparable models, *Econometrica*, 71, 1405–1443.

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

[6] DETTE, H., VOLGUSHEV, S. and WAGENER, J. (2011); Comparing conditional quantile curves, *Scandinavian Journal of Statistics*, forthcoming.

[7] ESCANCIANO, J. C. and VELASCO, C. (2010); Specification tests of parametric dynamic conditional quantiles, *Journal of Econometrics,* 159, 209-221.

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

[9] HAAG, B., HODERLEIN, S. and K. PENDAKUR (2009); Testing and Imposing Slutsky Symmetry, *Journal of Econometrics*, 153, 33-50,

[10] HAUSMAN, J. and W. NEWEY, 1995. Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss, *Econometrica*, 1445-1476.

[11] HODERLEIN, S. (2010): "How many Consumers are Rational?" *Working Paper, Boston College.*

[12] HODERLEIN, S. and MAMMEN, E. (2007): Identification of Marginal Effects in Nonseparable Models without Monotonicity, *Econometrica,* 75, 1513–1518.

[13] HODERLEIN, S. and MAMMEN, E. (2009): Identification and Estimation of Marginal Effects in Nonseparable, Nonmonotonic Models, *Econometrics Journal*, **12**, 1-25.

[14] IMBENS, G. and NEWEY, W. (2009), Identification and Estimation of Triangular Simultaneous Equations Models without Additivity, *Econometrica*, forthcoming.

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

[16] KOSOROK, M. R. (2008); Introduction to Empirical Processes and Semiparametric Inference, Springer.

[17] KOENKER, R. (2005); Quantile Regression, Cambridge University Press.

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

[19] LEWBEL, A. (2001); Demand Systems With and Without Errors, *American Economic Review*, 91, 611-618.

[20] MATZKIN, R., (2003). Nonparametric estimation of nonadditive random functions, *Econometrica* **71**, 1339-1375.

[21] NOLAN, D. and POLLARD, D. (1987); U–Processes: Rate of convergence, *Annals of Statistics*, 15, 780-799.

[22] PENDAKUR, K. (1999). Nonparametric estimation and testing of base-independent equivalence scales, *Journal of Econometrics* 88, 1-40.

[23] SPERLICH, S., (2009), A note on non-parametric estimation with predicted variables, *Econometrics Journa*l, 12, 382–395

[24] STONE, R., (1954); Linear Expenditure Systems and Demand Analysis: An Application to the Pattern of British Demand, *Economic Journal* 64, 511-527.

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

[26] SUN, Y. (2006); A consistent nonparametric equality test of conditional quantile functions, *Econometric Theory,* 22, 614-632.

[27] VAN DER VAART, A. W. and WELLNER, J. A. (1996); Weak convergence and empirical processes, Springer.

[28] YU, K. and JONES, M. C. (1998); Local linear quantile regression, *Journal of the American Statistical Association*, 93, 228-237.

[29] ZHENG, J. X. (1998); A consistent nonparametric test of parametric regression models under conditional quantile restrictions, *Econometric Theory*, 14, 123-138.

Power of Test Statistic for N=3000 as Function of Lambda

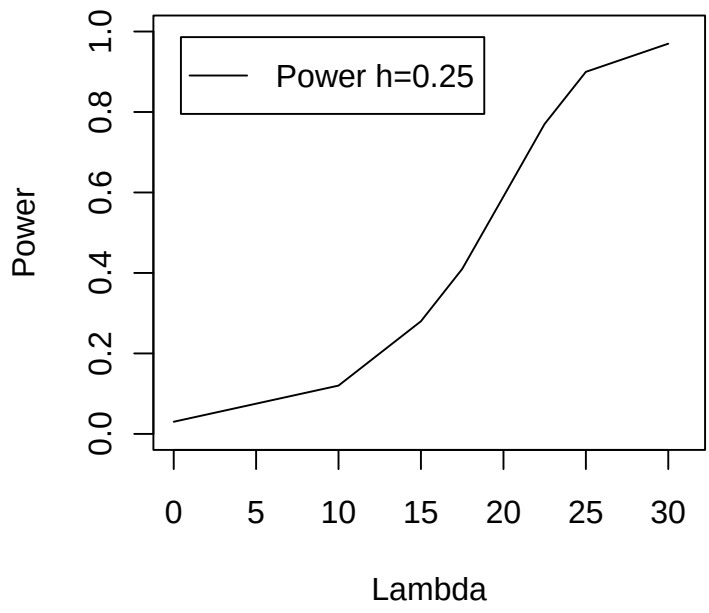

Figure 1

Power of Test Statistic as Function of Proportion Nonrational Individuals

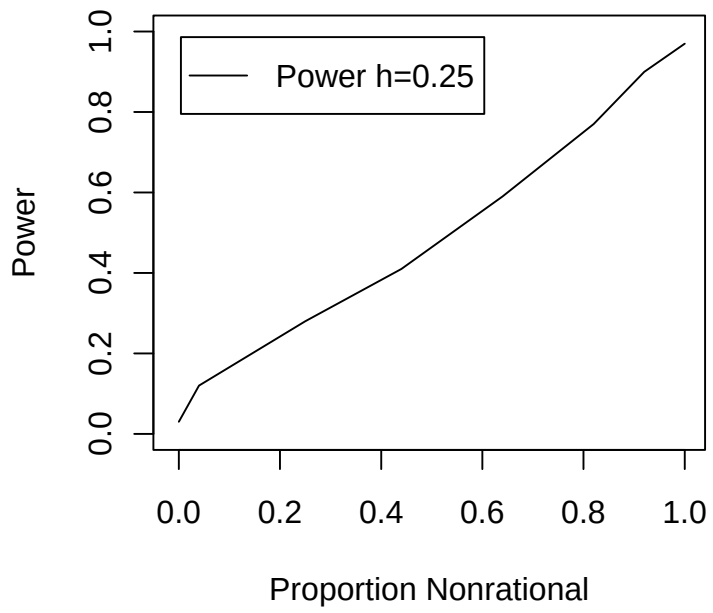

Figure 2

Power of Test Statistic as Function of Lambda
Size Distortion due to Small Bandwidth

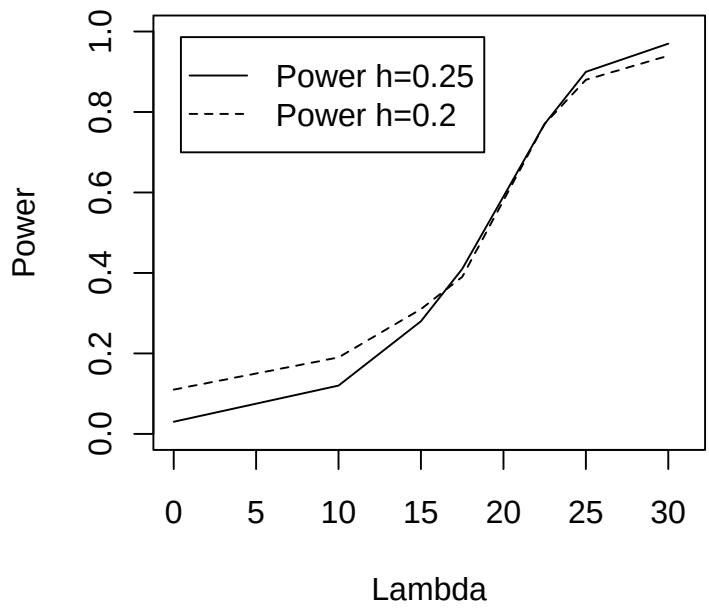

Figure 3

Power of Test Statistic as Function of Lambda
Effect of Sample Size - Consistency of Test

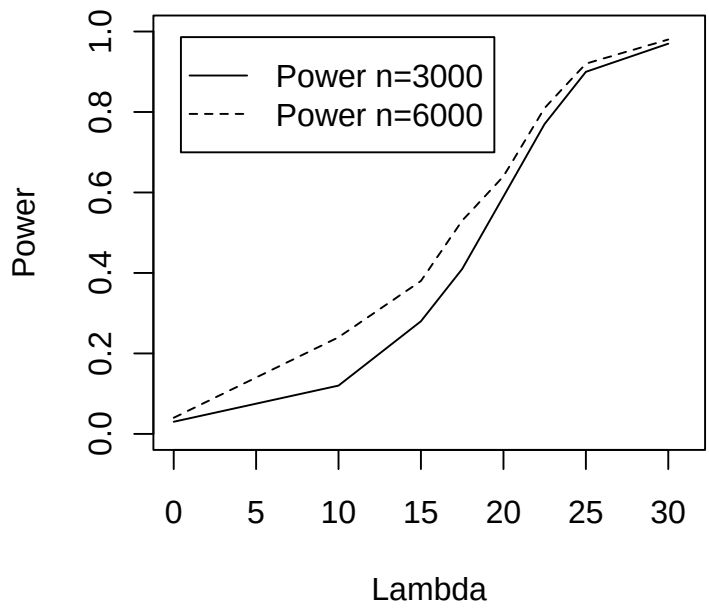

Figure 4