Local rank tests in a multivariate nonparametric relationship *

Natércia Fortuna
CEMPRE[†] Faculdade de Economia, Universidade do Porto

January 30, 2004

Abstract

Consider a multivariate nonparametric model where the unknown vector of functions depends on two sets of explanatory variables. For a fixed level of one set of explanatory variables, we provide consistent statistical tests, called local rank tests, to determine whether the multivariate relationship can be explained by a smaller number of functions. We also provide estimators for the smallest number of functions, called local rank, explaining the relationship. The local rank tests and the estimators of the local rank are based on the asymptotics of the eigenvalues of some matrix. This matrix is estimated by using kernel-based methods and the asymptotics of its eigenvalues is established by using the so-called Fujikoshi expansions along with some techniques of the theory of U-statistics. We present a simulation study which examines small sample properties of local rank tests. We also apply the local rank tests and the local rank estimators of the paper to a demand system given by a newly constructed data set. Our results can be viewed as localized counterparts of tests for a number of factors in a nonparametric relationship introduced by Donald.

Keywords: nonparametric relationship, local rank, local rank estimation, kernel smoothing, consistent tests, demand systems.

JEL classification: C12, C13, C14, D12.

1 Introduction

This study was motivated by the theory of ranks of demand systems. Recall that a demand system in economics is a functional relation $y = (y_1, \ldots, y_J)' = f(x, z) = (f_1(x, z), \ldots, f_J(x, z))'$ where y_j, $j = 1, \ldots, J$, is the proportion of the total expenditures for the jth good, called a budget share for the jth good, x is total expenditures (income, in short) and $z = (z_1, \ldots, z_J)$ are prices of J goods faced by a consumer. Introduced by Gorman (1981) and later developed by Lewbel (1991), the rank of a demand system can be either local or global. The *local rank* $\mathrm{rk}\{f(\cdot, z)\}$ at a fixed value of z is defined as the dimension of the function space spanned by the coordinate functions $f_1(x, z), \ldots, f_J(x, z)$ of $f(x, z)$ when z is fixed. The *global rank* is the maximum of local ranks taken over all values of z. In other words, the rank is the smallest number of functions needed to explain the demand system

*The paper is based on part of the author's Ph.D. thesis. The author would like to thank Stephen G. Donald and Vladas Pipiras for helpful comments and suggestions. Thanks also to James Spencer and Pierre Bahizi.

†CEMPRE - Centro de Estudos Macroeconómicos e Previsão - is supported by the Fundação para a Ciência e a Tecnologia, Portugal, through the Programa Operacional Ciência, Tecnologia e Inovação (POCTI) of the Quadro Comunitário de Apoio III, which is financed by FEDER and Portuguese funds.

$y = f(x, z)$, either locally at z or globally over all values of z. Ranks turn out to be of great interest in Economic Theory where demand systems are derived through a utility maximization principle. For example, Gorman (1981) showed that commonly used exactly aggregable demand systems, when derived through a utility maximization principle, have always rank less than or equal to 3. Lewbel (1991) showed that ranks have important implications on functional structure and aggregation of demand systems. Some further theoretical studies related to ranks can be found in Lewbel (1989), Russell and Farris (1993) and Lewbel and Perraudin (1995).

Parallel to understanding its implications for Economic Theory, the rank of a demand system has been also studied from the point of view of a statistical estimation. A statistical model for a demand system is assumed to have a stochastic form $Y_i = f(X_i, Z_i) + \epsilon_i$, where Y_i, X_i and Z_i are the shares of goods, the income and the prices faced by the ith consumer, and ϵ_i is the noise term. The rank of a stochastic system is defined in the same way as in the deterministic situation by using the coordinate functions of a vector $f(x, z)$. The function f is assumed to have either a nonparametric or a (semi)parametric form. The goal is to estimate the rank of a demand system from the observations Y_i, X_i and Z_i. Under a semiparametric model, the rank of a demand system is typically expressed as the rank of some matrix. The problem then becomes that of estimating the rank of a matrix. This can be done by using one of the matrix rank estimation procedures found in the literature for example, the minimum-χ^2 test of Cragg and Donald (1997), or the LDU-based test of Gill and Lewbel (1992) (with a correction of Cragg and Donald (1996)). Under a nonparametric model, the rank of a demand system is estimated by following the central work of Donald (1997).

In most of the statistical work thus far, it has been assumed that prices are constant across consumers, that is, a model contains no variable Z_i. (In this case, there is no distinction between local and global rank tests.) One did so for simplicity and also because most data sets on consumer expenditure, in particular the well-known and commonly used Consumer Expenditures Survey (CEX, in short) data set of the United States, does not contain information on prices. The assumption of constant prices, however, is not realistic. For example, in the case of the United States, the CEX data set covers households across all the United States and prices are clearly different in its various parts. The focus of this work is on extensions of the rank estimation problems to situations where variations in prices are taken into account. We will assume below a nonparametric form of a demand system and provide statistical tests to determine its local ranks. Estimation of local ranks in a particular (semi)-parametric model can be found in Donald, Fortuna and Pipiras (2004a). Estimation of global ranks, being non-trivial, is left for the future work. Another significant part of this work consists of producing a data set which simultaneously contains information on expenditures and prices faced by a consumer in the United States. We do so by matching the CEX data on expenditures of a consumer with the American Chamber of Commerce Researchers Association (ACCRA, in short) data of prices. A data set similar to ours was recently constructed and used by Nicol (2001) in the context of a statistical modelling of demand systems. Our work can also be viewed as a generalization of Donald (1997). Since many ideas and proofs of this paper appear less sophisticated in Donald (1997), we suggest that the reader refers to Donald (1997) for further insight into our work.

In view of our motivation described above, let $(X_i, Z_i) \in \mathbb{R}^n \times \mathbb{R}^m$ be independent variables and $Y_i \in \mathbb{R}^G$ be a response variable explained by (X_i, Z_i). Suppose that the relationship between the variables Y_i and (X_i, Z_i) is given by the nonparametric model

$$Y_i = F(X_i, Z_i) + U_i, \quad i = 1, \dots, N, \tag{NP}$$

where N is the number of observations, $F(x, z) = (F_1(x, z), \ldots, F_G(x, z))'$ is an *unknown* $G \times 1$ vector of functions of x and z, and U_i is a $G \times 1$ noise vector with the variance-covariance matrix

$$\Sigma = EU_iU_i'. \tag{1.1}$$

One of the key assumptions of this work is the non-singularity (invertibility) of the matrix Σ. We will also assume that (X_i, Z_i) are independent for different i's and that

$$E(U_i|X_i, Z_i, X_j, Z_j) = 0. \tag{1.2}$$

These and additional assumptions on the variables X_i, Z_i and U_i, and on the function F are stated in Section 3. To state the problems considered in this paper, we need the following definition generalizing the notion of a local rank $\mathrm{rk}\{F(\cdot, z)\}$ introduced earlier.

Definition 1.1 Define the local rank of a $G \times 1$ vector $F(x, z)$ at z (and related to a $d_1 \times 1$ subvector x^1 of x), denoted by

$$\mathrm{rk}\{F(\cdot, z); x^1\}, \tag{1.3}$$

as the smallest integer L such that, for a $d_1 \times 1$ subvector x^1 of x, a $G \times d_1$ matrix $c(z)$, a $G \times L$ matrix $A(z)$ and a $L \times 1$ vector $H(x, z)$, we have

$$F(x, z) = c(z)x^1 + A(z)H(x, z). \tag{1.4}$$

By $\mathrm{rk}\{F(\cdot, z); 0\}$, also denoted by $\mathrm{rk}\{F(\cdot, z)\}$, we shall mean the smallest L such that the decomposition (1.4) holds without the term $c(z)x^1$.

Observe that the definition of the local rank $\mathrm{rk}\{F(\cdot, z); 0\} = \mathrm{rk}\{F(\cdot, z)\}$ is equivalent to that given in the beginning of the section. Why then introduce the notion of a more general local rank? The answer goes back to the assumption of the non-singularity of the covariance matrix $\Sigma = EU_iU_i'$ which we will use. If $Y_i = f(X_i, Z_i) + \epsilon_i$ is a nonparametric model of a demand system, then the sum of the budget shares Y_{ij}, $j = 1, \ldots, J$, in $Y_i = (Y_{i1}, \ldots, Y_{iJ})$ is always equal to 1. This implies that the covariance matrix of ϵ_i is singular and hence that the results of the paper do not, in principle, apply because they rely on non-singular covariance matrices. The way out is to observe that, because of the sum to 1 condition, $\mathrm{rk}\{f(\cdot, z)\} = \mathrm{rk}\{F(\cdot, z); 1\} + 1$, where $F(\cdot, z)$ is a vector $f(\cdot, z)$ without any of its coordinate functions and $\mathrm{rk}\{F(\cdot, z); 1\}$ is defined by Definition 1.1 with $d_1 = 1$ and $x^1 = 1$. Then, to estimate the local rank of a demand system, drop one share of goods from the analysis, allowing to assume non-singular covariance matrix of disturbances, estimate $\mathrm{rk}\{F(\cdot, z); 1\}$ and add 1 to the result. From this perspective, why then consider the local ranks $\mathrm{rk}\{F(\cdot, z)\}$ and $\mathrm{rk}\{F(\cdot, z); x^1\}$ with $x^1 \neq 1$? We include these ranks in Definition 1.1 because of potential applications to problems other than the rank of a demand system and also because our proofs in the case of $\mathrm{rk}\{F(\cdot, z); 1\}$ and in the general case of $\mathrm{rk}\{F(\cdot, z); x^1\}$ are not very different. In addition, Definition 1.1 follows the framework of Donald (1997) where our general rank with no z is also implicitly defined.

In this work, we focus on and address the following problems related to local rank $\mathrm{rk}\{F(\cdot, z); x^1\}$.

Basic problems. For a fixed z and L, provide statistical tests for the hypothesis testing problem of $H_0 : \mathrm{rk}\{F(\cdot, z); x^1\} \leq L$ against the alternative $H_1 : \mathrm{rk}\{F(\cdot, z); x^1\} > L$. For a fixed z, provide an estimator for the local rank $\mathrm{rk}\{F(\cdot, z); x^1\}$.

Since z is fixed, we will refer to the above statistical tests as *local rank tests* or *rank tests local at* z. The basic idea behind these local rank tests, explained in greater detail in Section 2 below, is to relate the local rank $\mathrm{rk}\{F(\cdot,z);x^1\}$ to the number of zero eigenvalues of some matrix. Then, by testing for the number of zero eigenvalues of this matrix, one can make an inference about the local rank $\mathrm{rk}\{F(\cdot,z);x^1\}$. The difficult parts of this plan are to find the right matrix, to obtain its estimator and, finally, to find and prove the asymptotics of the eigenvalues of the estimator which would allow to distinguish between the two hypothesis. The goal of the paper is to show how these difficulties can be overcome.

The rest of the paper is structured as follows. In Section 2, we explain the basic idea behind the local rank tests and also introduce the related test statistic. In Section 3, we state our assumptions. In Section 4, we establish the asymptotic properties of the test statistic and, based on these properties, we formulate the local rank tests. Section 5 is on the estimation of the local rank itself. Simulation results and applications to demand systems can be found in Sections 6 and 7, respectively. Finally, in Section 8, we draw some conclusions. The proofs of all the results can be found in Appendices A and B. Appendix C contains a result on asymptotics of a second order U-statistics.

2 Preliminaries

The basic idea behind local rank tests for (NP) model lies in the following lemma. See Appendix A for its elementary proof.

Lemma 2.1 *For some fixed z and L, we have* $\mathrm{rk}\{F(\cdot,z);x^1\}\leq L$ *if and only if the matrix*

$$\Gamma_{w,z}=E\gamma(X_i,z)\widetilde{F}(X_i,z)\widetilde{F}(X_i,z)', \tag{2.1}$$

where $\gamma(x,z)>0$ is any real-valued function and

$$\widetilde{F}(x,z)=F(x,z)-E\beta(X_i,z)F(X_i,z)X_i^{1\prime}(E\beta(X_i,z)X_i^1X_i^{1\prime})^{-1}x^1 \tag{2.2}$$

with any real-valued function $\beta(x,z)\neq 0$, has $G-L$ zero eigenvalues, or if and only if the matrix $\Gamma_{w,z}\Sigma^{-1}$ has $G-L$ zero eigenvalues, where Σ is defined in (1.1).

Remark 2.1 Let $\mathrm{rk}\{A\}$ denote the rank of a matrix A. By Lemma 2.1, the condition $\mathrm{rk}\{F(\cdot,z);x^1\}\leq L$ for some z and L, is also equivalent to the condition $\mathrm{rk}\{\Gamma_{w,z}\}\leq L$. In other words, we have

$$\mathrm{rk}\{F(\cdot,z);x^1\}=\mathrm{rk}\{\Gamma_{w,z}\}. \tag{2.3}$$

In this work, by using Lemma 2.1, local rank tests will be based on the eigenvalues of an estimator of the matrix $\Gamma_{w,z}\Sigma^{-1}$. Connection to the rank of the matrix $\Gamma_{w,z}$ allows, however, to view local rank tests in a general framework of rank estimation of symmetric matrices. See Remarks 4.1 and 4.2 below for a further discussion.

Local rank tests for (NP) model will then be based upon the smallest $G-L$ eigenvalues of an estimator of the matrix $\Gamma_{w,z}\Sigma^{-1}$. As can be seen from the proof of Theorem 4.1 below, the matrix Σ^{-1} plays the role of a normalization in order to obtain standardized limit laws. The weights $\gamma(x,z)$ and $\beta(x,z)$ are taken for convenience to allow for easier manipulations. We will take

$$\gamma(x,z) = \frac{p(x,z)^2}{\widetilde{p}(x)}, \quad \beta(x,z) = \frac{p(x,z)}{\widetilde{p}(x)}, \tag{2.4}$$

where $p(x,z)$ and $\widetilde{p}(x)$ are the densities of the vector (X,Z) and the variable X, respectively.

We will define the estimators for the matrices $\Gamma_{w,z}$ and Σ by using kernel functions. Definition of a kernel function, or simply a kernel, is given next. Set $x^b = x_1^{b_1} \dots x_m^{b_m}$ and $|b| = b_1 + \dots + b_m$ for $b = (b_1, \dots, b_m) \in (\mathbb{N} \cup \{0\})^m$ and $x = (x_1, \dots, x_m) \in \mathbb{R}^m$.

Definition 2.1 A function $K : \mathbb{R}^m \to \mathbb{R}$ is a kernel of order $r \in \mathbb{N}$ on $\mathbb{R}^m$ if it has a compact support, is bounded and satisfies the following conditions: (i) $\int_{\mathbb{R}^m} K(x)dx = 1$ and (ii) $\int_{\mathbb{R}^m} x^b K(x)dx = 0$ for any $b \in (\mathbb{N} \cup \{0\})^m$ such that $1 \le |b| < r$.

Kernel functions are used in statistics, as well as in other areas of applied or pure mathematics, because of their localization property which involves a scaled kernel function

$$K_h(\cdot) = h^{-m} K(h^{-1} \cdot), \tag{2.5}$$

where $h > 0$ is called a bandwidth. See Proposition B.1 in Appendix B for a precise statement of the localization property. Two elementary consequences of this property can be expressed as

$$\begin{aligned} E\frac{p(X_i,z)}{\widetilde{p}(X_i)} G(X_i,z) &\approx EG(X_i,Z_i)K_h(z-Z_i), & (2.6) \\ E\widetilde{p}(X_i)G(X_i,X_i) &\approx EG(X_i,X_j)\widetilde{K}_h(X_i-X_j), & (2.7) \end{aligned}$$

where $i \ne j$ (with X_i and X_j assumed independent), G is some function and $\approx$ denotes an approximation as h approaches zero (relation (2.6) follows from Lemma B.11, (c), below).

By using the relations (2.6) and (2.7), we can informally derive the estimator of $\Gamma_{w,z}$ as follows. Observe first that, by (2.4) and since $E(U_i|X_i,Z_i) = 0$ by (1.2),

$$\begin{aligned} E\beta(X_i,z)F(X_i,z)X_i^{1\prime} &= E\frac{p(X_i,z)}{\widetilde{p}(X_i)}F(X_i,z)X_i^{1\prime} \approx EF(X_i,Z_i)X_i^{1\prime}K_h(z-Z_i) \\ &= EY_iX_i^{1\prime}K_h(z-Z_i) \approx \frac{1}{N}\sum_{i=1}^N Y_iX_i^{1\prime}K_h(z-Z_i), & (2.8) \end{aligned}$$

where K is a kernel on $\mathbb{R}^m$, and similarly,

$$E\beta(X_i,z)X_i^1X_i^{1\prime} \approx \frac{1}{N}\sum_{i=1}^N X_i^1X_i^{1\prime}K_h(z-Z_i). \tag{2.9}$$

Therefore, multiplying (2.8) by the inverse of (2.9), we obtain that

$$\begin{aligned} \Pi_1(z)' &:= E\beta(X_i,z)F(X_i,z)X_i^{1\prime}(E\beta(X_i,z)X_i^1X_i^{1\prime})^{-1} \\ &\approx \frac{1}{N}\sum_{i=1}^N Y_iX_i^{1\prime}K_h(z-Z_i)\left(\frac{1}{N}\sum_{i=1}^N X_i^1X_i^{1\prime}K_h(z-Z_i)\right)^{-1} =: \widehat{\Pi}_1(z)'. & (2.10) \end{aligned}$$

Similarly, by using (2.6) again,

$$\begin{aligned}\Gamma_{w,z} &= E\frac{p(X_i,z)^2}{\widetilde{p}(X_i)}(F(X_i,z)-\Pi_1(z)'X_i^1)(F(X_i,z)-\Pi_1(z)'X_i^1)' \\ &\approx Ep(X_i,z)(F(X_i,Z_i)-\Pi_1(z)'X_i^1)(F(X_i,z)-\Pi_1(z)'X_i^1)'K_h(z-Z_i)\end{aligned}$$

and, since $E(U_i|X_i,Z_i)=0$ by (1.2),

$$\Gamma_{w,z}\approx Ep(X_i,z)(Y_i-\Pi_1(z)'X_i^1)(F(X_i,z)-\Pi_1(z)'X_i^1)'K_h(z-Z_i).$$

Taking $j\neq i$, writing $p(X_i,z)$ above as $\widetilde{p}(X_i)(p(X_i,z)/\widetilde{p}(X_i))$ and using relations (2.6) and (2.7), we may get a further approximation of $\Gamma_{w,z}$ as

$$\Gamma_{w,z}\approx E(Y_i-\Pi_1(z)'X_i^1)(F(X_j,Z_j)-\Pi_1(z)'X_j^1)'\widetilde{K}_h(X_i-X_j)K_h(z-Z_i)K_h(z-Z_j), \tag{2.11}$$

where $\widetilde{K}$ is a kernel on $\mathbb{R}^n$. Then, by using $E(U_j|X_i,Z_i,X_j,Z_j)=0$ in (1.2) and by using the approximation (2.10), we obtain that

$$\begin{aligned}\Gamma_{w,z} &\approx E(Y_i-\Pi_1(z)'X_i^1)(Y_j-\Pi_1(z)'X_j^1)'\widetilde{K}_h(X_i-X_j)K_h(z-Z_i)K_h(z-Z_j) \\ &\approx E(Y_i-\widehat{\Pi}_1(z)'X_i^1)(Y_j-\widehat{\Pi}_1(z)'X_j^1)'\widetilde{K}_h(X_i-X_j)K_h(z-Z_i)K_h(z-Z_j). \end{aligned}\tag{2.12}$$

Based on these approximations, we define the estimator of $\Gamma_{w,z}$ as follows.

Definition 2.2 Define the estimator of the matrix $\Gamma_{w,z}$ in (2.1) with (2.2) and (2.4) as

$$\widehat{\Gamma}_{w,z}=\frac{1}{N(N-1)}\sum_{i\neq j}^{N}(Y_i-\widehat{\Pi}_1(z)'X_i^1)(Y_j-\widehat{\Pi}_1(z)'X_j^1)'\widetilde{K}_h(X_i-X_j)K_h(z-Z_i)K_h(z-Z_j), \tag{2.13}$$

where $\widehat{\Pi}_1(z)'$ is given by (2.10).

Remark 2.2 In the case of a nonparametric model with no variable z, Donald (1997) defined the estimator for rank tests as

$$\widehat{\Gamma}_w=\frac{1}{N(N-1)}\sum_{i\neq j}^{N}(Y_i-\widehat{\Pi}_1'X_i^1)(Y_j-\widehat{\Pi}_1'X_j^1)'\widetilde{K}_h(X_i-X_j),$$

where $\widehat{\Pi}_1=(X^{1\prime}X^1)^{-1}X^{1\prime}Y$. The difference between our estimator $\widehat{\Gamma}_{w,z}$ and Donald's estimator $\widehat{\Gamma}_w$ is that we localize at z. Indeed, observe that $\widehat{\Gamma}_{w,z}$ becomes $\widehat{\Gamma}_w$ when we remove localization terms $K_h(z-Z_i)$.

Remark 2.3 According to Definition 2.2 and the argument preceding it, we expect the estimator

$$\widehat{\Gamma}_{w,z}=\frac{1}{N(N-1)}\sum_{i\neq j}^{N}Y_iY_j'\widetilde{K}_h(X_i-X_j)K_h(z-Z_i)K_h(z-Z_j), \tag{2.14}$$

to be used for tests of local rank $\mathrm{rk}\{F(\cdot,z)\}=\mathrm{rk}\{F(\cdot,z);0\}$. The definition (2.14) is simpler than (2.13) because it does not involve the subtracted terms $\widehat{\Pi}_1(z)'X_i^1$. This correction appears in (2.13) to account for the term $c(z)x^1$ in a general definition of a local rank $\mathrm{rk}\{F(\cdot, z); x^1\}$. Interestingly, by expressing the matrix $\widehat{\Pi}_1(z)'$ as

$$\widehat{\Pi}_1(z)' = YDX^{1'}(X^1DX^{1'})^{-1}, \tag{2.15}$$

where Y is a $G \times N$ matrix with columns Y_i, X^1 is a $d_1 \times N$ matrix with columns X_i^1 and $D = \mathrm{diag}\{K_h(z - Z_1), \ldots, K_h(z - Z_N)\}$ is the $N \times N$ diagonal matrix, we see that it can be viewed as a generalized least-squares estimator for the matrix $c(z)$ in the model $Y_i = c(Z_i)X_i^1 + \epsilon_i$. The weight matrix D in (2.15) is used for localization at a fixed value of z.

The estimator for the variance-covariance matrix Σ which we will use is defined as follows.

Definition 2.3 Define the estimator of the matrix Σ in (1.1) as

$$\widehat{\Sigma} = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \widehat{F}(X_i, Z_i))(Y_i - \widehat{F}(X_i, Z_i))', \tag{2.16}$$

where

$$\widehat{F}(x, z) = \frac{1}{N}\sum_{i=1}^{N} Y_i\widetilde{K}_h(x - X_i)K_h(z - Z_i)\ \widehat{p}(x, z)^{-1} \tag{2.17}$$

and

$$\widehat{p}(x, z) = \frac{1}{N}\sum_{i=1}^{N}\widetilde{K}_h(x - X_i)K_h(z - Z_i). \tag{2.18}$$

In contrast to the estimator (2.13) which is new, those in (2.16), (2.17) and (2.18) are standard and commonly used estimators for the variance-covariance matrix Σ, an unknown vector of functions $F(x, z)$ and a density function $p(x, z)$, respectively.

3 Assumptions

In this section, we list and briefly discuss the assumptions which will be used for local rank tests.

Assumption 1: *Suppose that $(X_i, Z_i) \in \mathbb{R}^n \times \mathbb{R}^m$, $i = 1, \ldots, N$, are i.i.d. random vectors such that the support of (X_i, Z_i), denoted by $\mathcal{H}_x \times \mathcal{H}_z$, is the Cartesian product of compact intervals and (X_i, Z_i) are continuously distributed with a density $p(x, z)$ which is bounded below by a constant and has an extension to $\mathbb{R}^n \times \mathbb{R}^m$ with $s \geq r$ continuous bounded derivatives.*

Assumption 2: *Suppose that U_i, $i = 1, \ldots, N$, are i.i.d. random vectors, independent of the sequence (X_i, Z_i) and such that $EU_i = 0$ and $EU_iU_i' = \Sigma$, where Σ is a positive definite matrix. Assume also that $EU_j^4 < \infty$.*

Assumption 3: *The function $F : \mathcal{H}_x \times \mathcal{H}_z \to \mathbb{R}^G$ is such that each of its component functions has an extension to $\mathbb{R}^n \times \mathbb{R}^m$ with $s \geq r$ continuous bounded derivatives.*

Assumption 4: *The matrix*

$$Q_1(z) \quad = \quad E\frac{p(X_1, z)}{\widetilde{p}(X_1)}X_1^1X_1^{1'} = \int_{\mathbb{R}^n} x_1^1x_1^{1'}p(x_1, z)dx_1, \tag{3.1}$$

is positive definite (invertible), where $\widetilde{p}(x)$ is the density function of X_i.

ASSUMPTION 5: *The functions $\widetilde{K}$ and K are symmetric kernels on $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, of order r.*

Assumptions 1–4 are in the spirit of those used by Donald (1997). Assumption 1 requires that the density of (X_i, Z_i) is smooth, has a compact support and, moreover, is bounded from below. Assumption 2 imposes an important invertibility restriction on the variance-covariance matrix Σ. In Assumption 3, we suppose that the true nonparametric regression function satisfies smoothness conditions. Assumption 4 requires that the matrix $\Pi_1(z)$ in (2.10) is well-defined. Finally, Assumption 5 states what type of kernels we will use. Some of these assumptions can be slightly weakened, for example, by replacing the independence condition on U_i in Assumption 2 by suitable behavior of U_i given X_i and Z_i. We shall, however, not strive here for utmost generality and leave the assumptions which are easier to work with.

4 Local rank tests

The following result is key to local rank tests for (NP) model. Let $\widehat{\lambda}_1(z) \leq \ldots \leq \widehat{\lambda}_G(z)$ be the eigenvalues of the matrix $\widehat{\Gamma}_{w,z}\widehat{\Sigma}^{-1}$. Since the matrix $\widehat{\Gamma}_{w,z}$ or $\widehat{\Gamma}_{w,z}\widehat{\Sigma}^{-1}$ is symmetric but not necessarily positive definite, its eigenvalues are real but not necessarily positive. Set

$$V(z) = \Big(2\|\widetilde{K}\|_2^2\|K\|_2^4 E\frac{p(X_i,z)^2}{\widetilde{p}(X_i)}\Big)^{-1/2} \tag{4.1}$$

and let

$$\widehat{V}(z) = \Big(2\|\widetilde{K}\|_2^2\|K\|_2^4 N^{-1}\sum_{i=1}^{N}\widehat{p}(X_i,Z_i)K_h(z-Z_i)\Big)^{-1/2}, \tag{4.2}$$

where $\widehat{p}(x,z)$ is given in (2.18). By Lemma B.10 below, under suitable conditions, $\widehat{V}(z)$ is a consistent estimator of $V(z)$. Let also $\mathcal{Z}_k$ be a symmetric $k \times k$ matrix having independent zero mean normal (Gaussian) entries with variance 1 in the diagonal and variance 1/2 off the diagonal, and $\lambda_1(\mathcal{Z}_k) \leq \ldots \leq \lambda_k(\mathcal{Z}_k)$ be the eigenvalues of $\mathcal{Z}_k$ in increasing order.

Theorem 4.1 *Suppose that Assumptions 1–5 of Section 3 hold, and that*

$$Nh^{m+3n/2} \to \infty \quad \text{and} \quad Nh^{m+n/2+2r} \to 0. \tag{4.3}$$

Set $L(z) = \mathrm{rk}\{F(\cdot,z);x^1\}$. Then, for $j = 1,\ldots,G-L(z)$,

$$\widehat{V}(z)Nh^{m+n/2}\widehat{\lambda}_j(z) \xrightarrow{d} \lambda_j(\mathcal{Z}_{G-L(z)}), \tag{4.4}$$

and, for $j = G-L(z)+1,\ldots,G$,

$$\widehat{V}(z)Nh^{m+n/2}\widehat{\lambda}_j(z) \xrightarrow{p} +\infty. \tag{4.5}$$

The proof of Theorem 4.1 is given in Appendix A below. We now state two immediate corollaries of Theorem 4.1 which can be used in local rank tests for (NP) model, namely, to test $H_0 : \mathrm{rk}\{F(\cdot, z); x^1\} \leq L$ against $H_1 : \mathrm{rk}\{F(\cdot, z); x^1\} > L$. To state the first corollary, let

$$\widehat{T}_1(L, z) = \frac{\widehat{V}(z) N h^{m+n/2}}{\sqrt{G-L}} \sum_{j=1}^{G-L} \widehat{\lambda}_j(z). \tag{4.6}$$

Recall that a stochastic dominance $\xi \leq_d \eta$ means that $P(\xi > x) \leq P(\eta > x)$ for all $x \in \mathbb{R}$.

Theorem 4.2 *Under the assumptions of Theorem 4.1, we have that, under the hypothesis* $H_0 : \mathrm{rk}\{F(\cdot, z); x^1\} \leq L$,

$$\widehat{T}_1(L, z) \xrightarrow{d} \frac{1}{\sqrt{G-L}} \sum_{j=1}^{G-L} \lambda_j(\mathcal{Z}_{G-L(z)}) \overset{d}{\leq} \mathcal{N}(0, 1), \tag{4.7}$$

where the stochastic dominance $\leq_d$ *in (4.7) is, in fact,* $=_d$ *for* $L = \mathrm{rk}\{F(\cdot, z); x^1\}$, *and, under the hypothesis* $H_1 : \mathrm{rk}\{F(\cdot, z); x^1\} > L$, $\widehat{T}_1(L, z) \to_p +\infty$.

Theorem 4.2 is proved in Appendix A. Observe that the stochastic dominance result in (4.7) and the divergence of the test statistic $\widehat{T}_1(L, z)$ under the alternative hypothesis can be used to test for the local rank $\mathrm{rk}\{F(\cdot, z); x^1\}$. At a significance level α, the hypothesis $H_0 : \mathrm{rk}\{F(\cdot, z); x^1\} \leq L$ is accepted if $\widehat{T}_1(L, z) \leq \mathcal{N}_\alpha(0, 1)$ where $\mathcal{N}_\alpha(0, 1)$ is the smallest ξ such that $P(\mathcal{N}(0, 1) \geq \xi) = \alpha$.

Another way to test for $\mathrm{rk}\{F(\cdot, z); x^1\}$ is to consider the test statistic defined as the sum of squared eigenvalues, namely,

$$\widehat{T}_2(L, z) = \widehat{V}(z)^2 N^2 h^{2m+n} \sum_{j=1}^{G-L} (\widehat{\lambda}_j(z))^2. \tag{4.8}$$

The following result concerns the asymptotics of $\widehat{T}_2(L, z)$ which can also be used to test for $\mathrm{rk}\{F(\cdot, z); x^1\}$. Its proof can be found in Appendix A. The notation $\chi^2(k)$ below stands for a χ^2-distribution with k degrees of freedom.

Theorem 4.3 *Under the assumptions of Theorem 4.1, we have that, under the hypothesis* $H_0 : \mathrm{rk}\{F(\cdot, z); x^1\} \leq L$,

$$\widehat{T}_2(L, z) \xrightarrow{d} \sum_{j=1}^{G-L} (\lambda_j(\mathcal{Z}_{G-L(z)}))^2 \overset{d}{\leq} \chi^2((G-L)(G-L+1)/2), \tag{4.9}$$

where the stochastic dominance $\leq_d$ *in (4.9) is, in fact,* $=_d$ *for* $L = \mathrm{rk}\{F(\cdot, z); x^1\}$, *and, under the hypothesis* $H_1 : \mathrm{rk}\{F(\cdot, z); x^1\} > L$, $\widehat{T}_2(L, z) \to_p +\infty$.

Theorem 4.2 is in the spirit of Theorem 2 in Donald (1997). To our best knowledge, a result of Theorem 4.3 does not appear elsewhere in connection to rank testing in a nonparametric relationship (however, see Remark 4.1 below).

Remark 4.1 Observe that the test statistic $\widehat{T}_2(L, z)$ can be written as

$$\widehat{T}_2(L, z) = \widehat{V}(z)^2 N^2 h^{2m+n} \sum_{j=1}^{G-L} \widehat{\mu}_j(z),$$

where $\widehat{\mu}_j(z) = (\widehat{\lambda}_j(z))^2$ are the eigenvalues of the matrix $(\widehat{\Gamma}_{w,z}\widehat{\Sigma}^{-1}\widehat{\Gamma}_{w,z})\widehat{\Sigma}^{-1}$. Then, we can show, for example, as in the proof of Theorem 3 in Cragg and Donald (1993) that

$$\widehat{T}_2(L,z) = \widehat{V}(z)^2 N^2 h^{2m+n} \min_{\mathrm{rk}\{\Gamma\}\le L} \mathrm{vec}(\widehat{\Gamma}_{w,z} - \Gamma)'(\widehat{\Sigma}\otimes\widehat{\Sigma})^{-1}\mathrm{vec}(\widehat{\Gamma}_{w,z} - \Gamma), \tag{4.10}$$

where $\mathrm{rk}\{\Gamma\}$ denotes the rank of a matrix Γ, and vec and $\otimes$ stand for the commonly used vec operation and the Kronecker product, respectively. Relation (4.10) shows that $\widehat{T}_2(L,z)$ is a minimum-χ^2 type statistic used to test for the rank $\mathrm{rk}\{\Gamma_{w,z}\}$ of the matrix $\Gamma_{w,z}$ (see Cragg and Donald (1997)). This observation is not surprising because, by Remark 2.1 above, the local rank $\mathrm{rk}\{F(\cdot,z);x^1\}$ is equal to $\mathrm{rk}\{\Gamma_{w,z}\}$. Observe also that the number of degrees of freedom $(G-L)(G-L+1)/2$ in (4.9) is smaller than $(G-L)(G-L)$ used in a minimum$-\chi^2$ test for the rank of a $G\times G$ matrix such as $\Gamma_{w,z}$. This difference in degrees of freedom results from the symmetry restriction on the matrix $\Gamma_{w,z}$. The exact number $(G-L)(G-L+1)/2$ for the degrees of freedom in connection to rank estimation for symmetric matrices also appears in Robin and Smith (2000).

Remark 4.2 Remarks 2.1 and 4.1 indicate that local rank tests can be viewed as tests for the rank of the matrix $\Gamma_{w,z}$. The statistics and econometrics literature offers a number of tests for estimation of rank of an unknown matrix, for example, the LDU-based test in Gill and Lewbel (1992), Cragg and Donald (1996), the minimum-χ^2 test in Cragg and Donald (1997) or the asymptotic least-squares test in Robin and Smith (2000). A key assumption in all these tests is the asymptotic normality of some estimator of an unknown matrix. In our case, it is not very difficult to see that the asymptotic normality of $\widehat{\Gamma}_{w,z}$ can be shown by suitably adapting the proof of Theorem 4.1. However, we cannot readily apply most of the rank tests available in the literature because they are formulated for matrices with no restrictions. The matrices $\Gamma_{w,z}$ and $\widehat{\Gamma}_{w,z}$, on the other hand, are necessarily symmetric. Symmetry imposes restrictions on matrices involved and consequently alters rank test results (see Remark 4.1 above). These issues are addressed in greater depth by Donald, Fortuna and Pipiras (2004*b*). Observe also that none of the currently available methods for estimation of rank in a matrix correspond to the first test statistic $\widehat{T}_1(L,z)$ in (4.6). This statistic was not considered in a general context because eigenvalues are defined for square (in particular, symmetric) matrices only.

Remark 4.3 Observe from Definition 2.2 and the discussion preceding it that the bandwidths h corresponding to X_i and Z_i play somewhat different roles. The bandwidth corresponding to Z_i allows to localize the mean $\Gamma_{w,z}$ at a fixed point z. The bandwidth corresponding to X_i allows to express the mean $\Gamma_{w,z}$ in a convenient way as a U-statistic by localizing X_i at X_j (U-statistic is defined in Appendix C below). Hence, particularly in practice, one may want to distinguish between the bandwidths corresponding to X_i and Z_i, namely, to consider the test statistic

$$\widehat{\Gamma}_{w,z} = \frac{1}{N(N-1)}\sum_{i\neq j}^{N}(Y_i - \widehat{\Pi}_1(z)'X_i^1)(Y_j - \widehat{\Pi}_1(z)'X_j^1)'\widetilde{K}_{h_1}(X_i - X_j)K_{h_2}(z-Z_i)K_{h_2}(z-Z_j), \tag{4.11}$$

where $h_1, h_2 > 0$ (compare with Definition 2.2). One may show that, under suitable conditions, the eigenvalues $\widehat{\lambda}_j(z)$ of the matrix $\widehat{\Gamma}_{w,z}\widehat{\Sigma}^{-1}$ (where $\widehat{\Gamma}_{w,z}$ is defined in (4.11)) satisfy the limit results analogous to those in Theorems 4.1, 4.2 and 4.3. The only difference is that the factor $Nh^{m+n/2}$ in the normalization of (4.4), (4.5), (4.6) and (4.8) should now be replaced by

$$Nh_2^m h_1^{n/2}. \tag{4.12}$$

In our simulation study and applications (see Sections 6 and 7 below), we will consider the test statistic (4.11) and use the normalization (4.12).

5 Estimation of local rank

In this section, we use the local rank tests to estimate the true rank $\mathrm{rk}\{F(\cdot, z); x^1\}$ in (NP) model. Two methods available in the statistical literature can be used in order to determine the true rank, namely, the sequential testing procedure and the model selection criteria. We will focus here on the sequential testing procedure only because the model selection criteria has been found to perform poorly for small samples in a related problem (see Cragg and Donald (1997)).

Let $\widehat{T}_1(L, z)$ be the test statistic (4.6) used for local rank tests in (NP) model. The sequential testing is based on the following procedure: first, for increasing integer values $L = 1, \ldots, G$, and by using the statistic $\widehat{T}_1(L, z)$, test the hypothesis $H_0 : \mathrm{rk}\{F(\cdot, z); x^1\} \leq L$ against the alternative $H_1 : \mathrm{rk}\{F(\cdot, z); x^1\} > L$ at a given level of significance α, that is, determine whether

$$\widehat{T}_1(L, z) \leq \mathcal{N}_\alpha(0, 1), \tag{5.1}$$

where $\mathcal{N}_\alpha(0, 1)$ is the minimum ξ such that $P(\mathcal{N}(0, 1) > \xi) = \alpha$; second, stop at the first value of L which does not reject the hypothesis H_0, that is, when (5.1) holds. Denote this value of L by $\widehat{L}(z)$. In view of Theorem 4.1, $\widehat{L}(z)$ will not be a consistent estimator of $\mathrm{rk}\{F(\cdot, z); x^1\}$ because, as N increases and h becomes small, $\widehat{L}(z)$ will overestimate $\mathrm{rk}\{F(\cdot, z); x^1\}$ with probability $\alpha > 0$ (which is a fixed confidence level). The idea then, proposed by Pötscher (1983) in the context of determiniming the order of an autoregressive moving average (ARMA) model and by Bauer, Pötscher and Hackl (1988) in the context of model selection is to make α depend on N and h, that is, $\alpha = \alpha(N, h)$, and let $\alpha(N, h) \to 0$ as $N \to \infty$ and $h \to 0$. In this way, one can obtain a consistent estimator $\widehat{L}(z)$ of $\mathrm{rk}\{F(\cdot, z); x^1\}$. Let

$$\widehat{L}_1(z) = \min\{L : \hat{T}_1(L, z) < \mathcal{N}_{\alpha(N,h)}(0, 1)\},$$

where $\mathcal{N}_{\alpha(N,h)}(0, 1)$ is the smallest ξ such that $P(\mathcal{N}(0, 1) \geq \xi) = \alpha(N, h)$, be the minimum L which does not reject the null hypothesis $H_0 : \mathrm{rk}\{F(\cdot, z); x^1\} \leq L$ at a significance level $\alpha(N, h)$. The following result shows that $\widehat{L}_1(z)$ is a consistent estimator of $\mathrm{rk}\{F(\cdot, z); x^1\}$, provided the specified conditions on the significance levels $\alpha(N, h)$ hold. See Appendix A for a proof of this result.

Theorem 5.1 *With the above notation and under the assumptions of Theorem 4.1, we have* $\widehat{L}_1(z) \to_p \mathrm{rk}\{F(\cdot, z); x^1\}$ *as long as* $\alpha(N, h) \to 0$ *and* $(-\ln \alpha(N, h))^{1/2}/Nh^{m+n/2} \to 0$.

When using the test statistic $\widehat{T}_2(L, z)$, we need to consider

$$\widehat{L}_2(z) = \min\{L : \widehat{T}_2(L, z) < \chi^2_{\alpha(N,h)}((G - L)(G - L + 1)/2)\},$$

where $\chi^2_{\alpha(N,h)}((G-L)(G-L+1)/2)$ is the minimum ξ such that $P(\chi^2((G-L)(G-L+1)/2) > \xi) = \alpha(N, h)$. The following result is analogous to Theorem 5.1 above. Its proof can be found in Appendix A as well.

Theorem 5.2 *With the above notation and under the assumptions of Theorem 4.1, we have* $\widehat{L}_2(z) \to_p \mathrm{rk}\{F(\cdot, z); x^1\}$ *as long as* $\alpha(N, h) \to 0$ *and* $-\ln \alpha(N, h)/N^2h^{2m+n} \to 0$.

6 Simulation study

In this section, we use Monte Carlo simulations to examine size and power properties of local rank tests and properties of local rank estimators obtained through sequential testing. We also compare local rank estimation to other rank estimation procedures, e.g. the situation when the variable z (being part of the nonparametric model) is ignored. For simplicity, we focus henceforth only on local ranks $\mathrm{rk}\{F(\cdot,z)\} = \mathrm{rk}\{F(\cdot,z);0\}$ (see Definition 1.1).

The experimental setup is as follows. We consider $Y_i = \delta F(X_i, Z_i) + U_i$, $i = 1,\dots,N$, where

$$F(x,z) = \begin{pmatrix} \frac{5+z}{1+x^4}\sin(2.5+x^4) \\ \\ \frac{1}{3}(2-z)(z+\frac{1}{2})\arctan(5(x-1-\gamma z)) \\ \\ z(z+\frac{1}{2})(1+z^2) \end{pmatrix}. \tag{6.1}$$

We suppose that (X_i, Z_i), $i = 1,\dots,N$, are independent identically distributed $\mathbb{R}^2$ random vectors: X_i and Z_i are distributed uniformly on $[0,2]$ and $[-1,1]$, respectively, and X_i and Z_i are independent. We consider two sample sizes $N = 750$ and $N = 1500$, and two signal-to-noise ratios $\delta = 1$ and $\delta = 1/2$. The noise variables U_i are normally distributed as $\mathcal{N}(0, I_3)$, where I_3 is a 3×3 identity matrix. We suppose $\gamma = 1$ though we shall also discuss briefly the case $\gamma = 2$. We shall estimate local ranks $\mathrm{rk}\{F(\cdot,z)\}$ at $z = 1/2$, $z = 0$ and $z = -1/2$ corresponding to $\mathrm{rk}\{F(\cdot,1/2)\} = 3$, $\mathrm{rk}\{F(\cdot,0)\} = 2$ and $\mathrm{rk}\{F(\cdot,-1/2)\} = 1$, respectively. Size and power computations will be based on local rank tests for these values of z. The number of replications in Monte Carlo simulations is 1000 throughout. For kernel smoothing, we use a popular Epanechnikov kernel.

Remark 6.1 Our experimental setup is motivated by the following considerations. On one hand, we seek functions which lead to a desired local rank. For example, the functions in (6.1) are chosen in such a way that $\mathrm{rk}\{F(\cdot,1/2)\} = 3$. On the other hand, we do not want to consider noise perturbations which are too small leading to nearly deterministic shapes of observed functions, or which are too large in which case some observed functions become nearly indistinguishable. Such balance can be achieved by examining the plots of the observed functions of interest. In our choice of the function F and the noise variables, we have followed precisely this graphical approach.

Remark 6.2 In our simulation study, we consider two sample sizes $N = 750$ and $N = 1500$. Observe that these values of N are, in particular, larger than those considered in a simulation study of Donald (1997), namely, $N = 200$ and $N = 1000$. Our choice of larger N, however, should not be surprising. Setting $Z_i \equiv z$, the test statistic $\widehat{T}_1(L,z)$ reduces (up to a multiplicative constant $\|K\|_2^{-2}$) to the corresponding test statistic $\widehat{T}_p(L)$ considered by Donald (1997). Hence, local rank tests for $\mathrm{rk}\{F(\cdot,z)\}$ can be essentially thought as rank tests of Donald (1997) applied to the observed data at the cross section around a fixed value of z. For example, if $N = 750$, Z_i is uniformly distributed on $[-1,1]$ and $h = 0.3$ is a bandwidth, then there are on average $2N(2/h)^{-1} = 225$ available observations at the cross section of size $2h$ at a fixed value of z. Local rank tests with $N = 750$ and $h = 0.3$ then essentially correspond to Donald's rank tests with $N = 225$.

We shall focus in this section on local rank tests based on the test statistic $\widehat{T}_1(L,z)$. Results for the test statistic $\widehat{T}_2(L,z)$ are discussed at the end of this section and related tables can found at the end of this work.

Table 1 presents size computations of local rank tests in our simulation study, based on the test statistic $\widehat{T}_1(L,z)$. For example, for local rank 2 (at $z = 1/2$), the sizes are computed as actual rejection frequencies by using the respective asymptotic 5 percent critical values for the local rank test of $H_0 : \mathrm{rk}\{F(\cdot, 1/2)\} \leq 2$ in 1000 Monte Carlo replications with the specified values of N, bandwidths h_x and h_z (see Remark 4.3 above) and signal-to-noise ratio δ. We consider all possible combinations of bandwidths $h_x = 0.1, 0.3, 0.5$ and $h_z = 0.1, 0.3, 0.5$. These combinations cover a wide range of values of h_x and h_z. Moreover, the interval $[0.1, 0.5]$ corresponding to their smallest and largest values, contains most of h_x and h_z which where obtained through the generalized and the usual cross validations in a number of Monte Carlo simulations. For example, when $N = 750$ and $\delta = 1$, the generalized cross validation chose the values $(h_x, h_z) = (0.2, 0.4)$ and $(0.2, 0.3)$ in 30 Monte Carlo simulations.

Size of tests using $\widehat{T}_1$								
Local rank			1 ($z=-1/2$)			2 ($z=0$)		
N	δ	$h_x \setminus h_z$	0.1	0.3	0.5	0.1	0.3	0.5
750	1	0.1	6.1	5.1	9.6	0.1	1.0	2.3
		0.3	3.3	4.7	14.4	0.1	0.5	1.5
		0.5	2.4	4.1	16.8	0.1	0.5	1.9
		min-χ^2	7.0	8.3	24.6	2.0	2.7	6.4
	1/2	0.1	4.8	3.6	6.4	0.1	0.4	0.2
		0.3	2.5	3.7	5.4	0.0	0.1	0.3
		0.5	1.9	2.3	5.7	0.0	0.2	1.2
		min-χ^2	6.9	6.4	9.1	0.5	0.8	1.7
1500	1	0.1	3.1	3.5	18.3	0.8	1.0	3.3
		0.3	2.0	2.6	30.6	0.3	0.6	5.0
		0.5	1.6	2.7	36.8	0.3	0.5	6.5
		min-χ^2	5.3	5.1	45.0	3.2	5.2	14.5
	1/2	0.1	2.9	2.6	6.9	0.2	0.3	0.5
		0.3	2.1	1.9	7.3	0.2	0.1	0.5
		0.5	1.6	1.5	9.4	0.1	0.0	0.7
		min-χ^2	5.2	3.8	13.3	1.3	1.7	4.9

Table 1: Size of local rank tests using $\widehat{T}_1$

To compare a nonparametric approach to a parametric one, we also present in Table 1 sizes computed from a fixed parametric model. More precisely, we fit to the data a semiparametric factor model $Y_i = A(Z_i)H(X_i) + \epsilon_i$, where

$$H(x) = \begin{pmatrix} 1 \\ x \\ (x-1)^2 \end{pmatrix} \tag{6.2}$$

and $A(z)$ is unknown. The local rank $\mathrm{rk}\{F(\cdot, z)\}$ of the system $F(x,z) = A(z)H(x)$ can be shown to be the rank of the matrix $A(z)$. We therefore estimate the matrix $A(z)$ by using kernel smoothing methods and then test for its rank by using a minimum-χ^2 test for the rank of a matrix (Cragg and Donald (1997)). For more information on rank estimation in a semiparametric factor model, see Donald, Fortuna and Pipiras (2004a).

A few observations can be drawn from Table 1. The results indicate that local rank tests are likely to be undersized. Undersizing is most pronounced for local rank 2 ($z = 0$). Moreover, observe that the size appears to increase as h_z becomes larger, and it appears to get optimal as h_x decreases. The latter observation suggests that we should use h_x corresponding to undersmoothing. Interestingly, the same finding was also reported by Donald (1997) in the context of nonparametric rank testing without z. What h_z should be used is less clear, especially for rank 1. For rank 2, on the other hand, a larger h_z appears to be optimal. Then, we should perhaps use h_z corresponding to oversmoothing. Comparing nonparametric and parametric approaches, we see that local rank tests have significantly greater sizes for the parametric model. This discrepancy possibly results from the fact that the first function in (6.1), for example, is not easily approximated by a second order polynomial.

Power of tests using $\widehat{T}_1$											
True local rank			$L_0 = 2$			$L_0 = 3$					
Local rank test			$L = 1$			$L = 1$			$L = 2$		
N	δ	$h_x \setminus h_z$	0.1	0.3	0.5	0.1	0.3	0.5	0.1	0.3	0.5
750	1	0.1	17.8	42.4	52.4	89.7	100.0	100.0	51.7	93.3	98.7
		0.3	23.4	55.9	58.3	97.1	100.0	100.0	67.7	98.5	99.9
		0.5	25.6	61.7	49.6	99.1	100.0	100.0	74.4	98.8	99.7
		min-χ^2	23.4	59.9	56.5	97.7	100.0	100.0	80.1	99.4	99.7
	1/2	0.1	6.9	13.7	12.1	27.8	85.1	99.0	18.2	42.4	63.8
		0.3	9.4	14.5	17.7	46.9	96.2	100.0	26.5	57.5	72.6
		0.5	7.0	16.8	15.7	49.2	97.6	100.0	28.9	63.9	75.7
		min-χ^2	7.6	17.9	20.7	47.3	97.1	100.0	29.8	69.6	82.6
1500	1	0.1	33.7	81.8	83.2	99.9	100.0	100.0	78.8	99.7	100.0
		0.3	50.9	94.0	88.6	100.0	100.0	100.0	92.3	100.0	100.0
		0.5	52.8	94.4	81.8	100.0	100.0	100.0	93.6	100.0	100.0
		min-χ^2	49.6	94.3	83.7	100.0	100.0	100.0	96.0	100.0	100.0
	1/2	0.1	11.1	22.4	24.0	64.5	99.8	100.0	31.2	73.1	87.9
		0.3	16.3	35.5	33.8	84.7	100.0	100.0	41.4	85.9	95.1
		0.5	16.1	38.1	30.6	88.4	100.0	100.0	45.6	88.3	95.4
		min-χ^2	16.0	39.2	35.0	86.0	100.0	100.0	51.2	90.3	96.4

Table 2: Power of local rank tests using $\widehat{T}_1$

Power computations for local rank tests are presented in Table 2. These are size adjusted powers computed as follows. Consider for example the column $L = 2$ ($L_0 = 3$) in Table 2. The powers in this column are computed as actual rejection frequencies for the local rank test of $H_0 : \mathrm{rk}\{F(\cdot, 1/2)\} \leq 2$ (here, the true local rank $\mathrm{rk}\{F(\cdot, 1/2)\} = 3$ and hence the alternative $H_1 : \mathrm{rk}\{F(\cdot, 1/2)\} > 2$ is true). Size adjustment enters into computations through the critical values used for local rank tests, e.g. for $H_0 : \mathrm{rk}\{F(\cdot, 1/2)\} \leq 2$. These critical values are taken as to make the sizes of the corresponding local rank tests of Table 1 equal to 5 percent. For this column in Table 2, for example, the chosen critical values make the actual size of local rank tests $H_0 : \mathrm{rk}\{F(\cdot, 0)\} \leq 2$ equal to 5 percent.

The results of Table 2 suggest that power of local rank tests increases as h_x and h_z become larger. Observe, however, that the loss in power is not very significant as long as h_z is not too small. When the variable z is ignored, Donald (1997) has also observed that there is little loss in power when h_x decreases. Together with the results of Table 1, this observation suggests that using smaller h_x corresponding to undersmoothing and larger h_z corresponding to oversmoothing, may be more optimal for local rank tests. Observe also that powers obtained by fitting the semiparametric factor model are comparable to those of a nonparametric approach. This finding surprised us since we expected a great loss in power for the parametric approach, similarly to the results found in Donald (1997). In our understanding, this difference from Donald (1997) is just the result of our model choice (6.1).

Empirical distribution of $\widehat{L}$ using $\widehat{T}_1$ ($N = 750$, $\alpha = 0.05$)											
True rank			$L_0 = 1$			$L_0 = 2$			$L_0 = 3$		
δ	h_x	h_z	$\widehat{L} = 1$	$\widehat{L} = 2$	$\widehat{L} = 3$	$\widehat{L} = 1$	$\widehat{L} = 2$	$\widehat{L} = 3$	$\widehat{L} = 1$	$\widehat{L} = 2$	$\widehat{L} = 3$
1	0.1	0.1	0.939	0.060	0.001	0.809	0.190	0.001	0.095	0.647	0.258
		0.3	0.949	0.051	0.000	0.571	0.419	0.010	0.000	0.143	0.857
		0.5	0.904	0.093	0.003	0.351	0.626	0.023	0.000	0.030	0.970
	0.3	0.1	0.967	0.033	0.000	0.810	0.189	0.001	0.041	0.622	0.337
		0.3	0.953	0.046	0.001	0.457	0.538	0.005	0.000	0.079	0.921
		0.5	0.856	0.140	0.004	0.222	0.763	0.015	0.000	0.006	0.994
	0.5	0.1	0.976	0.024	0.000	0.806	0.193	0.001	0.020	0.659	0.321
		0.3	0.959	0.041	0.000	0.432	0.563	0.005	0.000	0.070	0.930
		0.5	0.832	0.164	0.004	0.208	0.773	0.019	0.000	0.004	0.996
	min-χ^2	0.1	0.930	0.065	0.005	0.694	0.286	0.020	0.016	0.280	0.704
		0.3	0.917	0.078	0.005	0.331	0.642	0.027	0.000	0.015	0.985
		0.5	0.754	0.227	0.019	0.162	0.774	0.064	0.000	0.002	0.998
1/2	0.1	0.1	0.898	0.047	0.001	0.911	0.068	0.001	0.718	0.263	0.015
		0.3	0.964	0.036	0.000	0.898	0.098	0.004	0.174	0.698	0.128
		0.5	0.936	0.064	0.000	0.859	0.139	0.002	0.007	0.683	0.310
	0.3	0.1	0.953	0.025	0.000	0.922	0.061	0.000	0.633	0.347	0.018
		0.3	0.963	0.037	0.000	0.889	0.110	0.001	0.054	0.726	0.220
		0.5	0.946	0.053	0.001	0.810	0.187	0.003	0.000	0.580	0.420
	0.5	0.1	0.957	0.019	0.000	0.942	0.053	0.000	0.613	0.377	0.010
		0.3	0.977	0.023	0.000	0.872	0.126	0.002	0.037	0.751	0.212
		0.5	0.943	0.055	0.002	0.830	0.168	0.002	0.000	0.593	0.407
	min-χ^2	0.1	0.911	0.064	0.005	0.889	0.103	0.005	0.464	0.409	0.127
		0.3	0.936	0.060	0.004	0.796	0.196	0.008	0.023	0.471	0.506
		0.5	0.909	0.085	0.006	0.717	0.266	0.017	0.000	0.289	0.711

Table 3: Empirical distribution of $\widehat{L}$ using $\widehat{T}_1$

For the sake of completeness, we provide in Tables 3 and 4 empirical distributions of local rank estimators obtained through sequential testing at a constant significance level $\alpha = 0.05$, for local ranks $\mathrm{rk}\{F(\cdot, -1/2)\}$ (true rank $L_0 = 1$), $\mathrm{rk}\{F(\cdot, 0)\}$ (true rank $L_0 = 2$) and $\mathrm{rk}\{F(\cdot, 1/2)\}$ (true rank $L_0 = 3$). It is quite remarkable that local rank tests perform so well when $N = 1500, \delta = 1$, and do fairly well even when $N = 1500, \delta = 1/2$ and $N = 750, \delta = 1$. The results appear poor for $N = 750, \delta = 1/2$ but we should not expect them better because there is just too much noise in the data. Let us also note that values for the empirical distribution of $\widehat{L}$ in Tables 3 and 4 sometimes do not sum to 1, e.g. when $N = 750, L_0 = 1, \delta = 1/2, h_x = 0.1$ and $h_z = 1$ in Table 3. In these cases, the estimator of $\widehat{L}$ takes also the value 0 (all three functions are indistinguishable from 0).

Empirical distribution of $\widehat{L}$ using $\widehat{T}_1$ ($N = 1500$, $\alpha = 0.05$)											
True rank			$L_0 = 1$			$L_0 = 2$			$L_0 = 3$		
δ	h_x	h_z	$\widehat{L} = 1$	$\widehat{L} = 2$	$\widehat{L} = 3$	$\widehat{L} = 1$	$\widehat{L} = 2$	$\widehat{L} = 3$	$\widehat{L} = 1$	$\widehat{L} = 2$	$\widehat{L} = 3$
1	0.1	0.1	0.969	0.031	0.000	0.704	0.288	0.008	0.001	0.366	0.633
		0.3	0.965	0.034	0.001	0.233	0.757	0.010	0.000	0.005	0.995
		0.5	0.817	0.177	0.006	0.046	0.921	0.033	0.000	0.000	1.000
	0.3	0.1	0.980	0.020	0.000	0.604	0.393	0.003	0.000	0.232	0.768
		0.3	0.974	0.026	0.000	0.086	0.908	0.006	0.000	0.001	0.999
		0.5	0.694	0.304	0.002	0.012	0.938	0.050	0.000	0.000	1.000
	0.5	0.1	0.984	0.016	0.000	0.605	0.392	0.003	0.000	0.247	0.753
		0.3	0.973	0.027	0.000	0.076	0.919	0.005	0.000	0.000	1.000
		0.5	0.632	0.366	0.002	0.013	0.922	0.065	0.000	0.000	1.000
	min-χ^2	0.1	0.947	0.049	0.004	0.500	0.468	0.032	0.000	0.061	0.939
		0.3	0.949	0.048	0.003	0.057	0.891	0.052	0.000	0.000	1.000
		0.5	0.550	0.421	0.029	0.009	0.846	0.145	0.000	0.000	1.000
1/2	0.1	0.1	0.970	0.029	0.000	0.919	0.079	0.002	0.413	0.528	0.059
		0.3	0.974	0.026	0.000	0.819	0.178	0.003	0.005	0.544	0.451
		0.5	0.931	0.069	0.000	0.716	0.279	0.005	0.000	0.290	0.710
	0.3	0.1	0.979	0.021	0.000	0.910	0.088	0.002	0.238	0.666	0.096
		0.3	0.981	0.019	0.000	0.770	0.229	0.001	0.002	0.385	0.613
		0.5	0.927	0.073	0.000	0.582	0.413	0.005	0.000	0.143	0.857
	0.5	0.1	0.984	0.016	0.000	0.913	0.086	0.001	0.194	0.714	0.092
		0.3	0.985	0.014	0.001	0.779	0.221	0.000	0.000	0.403	0.597
		0.5	0.906	0.094	0.000	0.578	0.415	0.007	0.000	0.149	0.851
	min-χ^2	0.1	0.948	0.048	0.004	0.836	0.151	0.013	0.138	0.513	0.349
		0.3	0.962	0.036	0.002	0.644	0.339	0.017	0.000	0.146	0.854
		0.5	0.867	0.125	0.008	0.472	0.479	0.049	0.000	0.036	0.964

Table 4: Empirical distribution of $\widehat{L}$ using $\widehat{T}_1$

Finally, in Table 5, we present empirical distribution of estimated rank $\widehat{L}$ when the variable z is ignored altogether. In other words, though still generating the variables (Y_i, X_i, Z_i) as above, we now focus only on the data $(Y_i, X_i), i = 1, \ldots, N$, and estimate its rank either by Donald's (1997) nonparametric approach or by a minimum-χ^2 test as the rank of a regression coefficient matrix after fitting a quadratic parametric model. It is interesting to compare the results of Table 5 to those of Tables 3 and 4. First, by letting h_z become very large, for any z, the test statistic $\widehat{T}_1(L, z)$ approaches the rank test statistic $\widehat{T}_p(L)$ considered by Donald (1997). Thus, if we included much larger values of h_z in Tables 3 and 4, we should get the results similar to those in Table 5. Table 5 thus sheds light on the estimation of local ranks when h_z becomes large.

Second, observe also that the results of Table 5 for $\widehat{L} = 3$ are a little smaller but still comparable to those in Tables 3 and 4 for the true rank $L_0 = 3, \widehat{L} = 3$ (with $h_z = 0.3$ or 0.5). Since $L_0 = 3 = \mathrm{rk}\{F(\cdot, 1/2)\} = \max_z \mathrm{rk}\{F(\cdot, 1/2)\}$ for the function system (6.1), one might think that the rank estimated ignoring the value z, provides a good estimator for the global rank $\max_z \mathrm{rk}\{F(\cdot, z)\}$. Our guess is that, depending on the model, this will not always be true. A simple example is the function system (6.1) with $\gamma = 2$. Since the argument $5(x - 1 - 2z)$ of the second function in (6.1) can take now a much larger range of values, the second coordinate of the data Y_i will appear much more like a white noise as a function of X_i (Z_i being ignored). Hence, the rank estimated ignoring z will not concentrate at 3 as in Table 5 but at the value 2. We have studied the data of this example through simulations. The obtained results confirmed our guess.

Empirical distribution of $\widehat{L}$ ignoring z					
N	δ	h_x	$\widehat{L} = 1$	$\widehat{L} = 2$	$\widehat{L} = 3$
750	1	0.1	0.000	0.189	0.811
		0.3	0.000	0.103	0.897
		0.5	0.000	0.106	0.894
		min-χ^2	0.000	0.047	0.953
	2	0.1	0.007	0.825	0.168
		0.3	0.001	0.777	0.222
		0.5	0.001	0.788	0.211
		min-χ^2	0.000	0.662	0.337
1500	1	0.1	0.000	0.012	0.988
		0.3	0.000	0.002	0.998
		0.5	0.000	0.002	0.998
		min-χ^2	0.000	0.000	1.000
	2	0.1	0.000	0.604	0.396
		0.3	0.000	0.454	0.546
		0.5	0.000	0.461	0.539
		min-χ^2	0.000	0.319	0.681

Table 5: Empirical distribution of $\widehat{L}$ ignoring z ($\alpha = 0.05$)

We have focused thus far on the simulation results obtained through the test statistic $\widehat{T}_1(L, z)$. In Tables 10–13 below, we also report analogous results based on the alternative test statistic $\widehat{T}_2(L, z)$. The results of Tables 3–4 for $\widehat{T}_1(L, z)$ and Tables 10–11 for $\widehat{T}_2(L, z)$ suggest that the statistic $\widehat{T}_2(L, z)$ slightly overestimates (respectively, underestimates) the local rank as compared to the statistic $\widehat{T}_1(L, z)$ when the true rank is not full, that is, the true rank $L_0 = 1$ or $L_0 = 2$ in the tables (respectively, the true rank is full, that is the true rank $L_0 = 3$ in the tables). This observation translates into the fact that the tests based on $\widehat{T}_2(L, z)$ have better size properties than those based on $\widehat{T}_1(L, z)$ (compare Tables 1 and 12). Focus now on the power properties of the statistics $\widehat{T}_1(L, z)$ and $\widehat{T}_2(L, z)$ in Tables 2 and 13, respectively. It can be seen from these tables that the power of tests based on $\widehat{T}_2(L, z)$ is typically worse than that based on $\widehat{T}_1(L, z)$. Observe, however, that the powers are still comparable in most cases of practical interest. (There is a significant difference in power for $L_0 = 3$, $L = 2$ and

$\delta = 1$ when h_x or h_z is the smallest, and for $L_0 = 3$, $L = 2$ when the signal-to-noise ratio $\delta = 1/2$ is the smallest.) Since the tests based on $\widehat{T}_2(L, z)$ have better size properties, this suggests that using the statistic $\widehat{T}_2(L, z)$ may often be more reliable than using the statistic $\widehat{T}_1(L, z)$.

7 Application to demand system

In this section, we estimate local rank in a demand system. The data set which we use contains information on expenditures and prices faced by a number of consumers across the United States. Expenditures are taken from the US CEX micro data of the first quarter of 2000.[1] More specifically, we first extract from the CEX data set only those households which contain married couples, whose tenure status is renter household or homeowner with or without mortgage, and whose age of the head is between 25 and 60. We also drop from our analysis those households whose total income was lower than \$3,000 or higher than \$75,000. (In addition, we consider households in the so-called metropolitan statistical areas because we can associate prices only to these households; see below.) Such selection, similarly used by Nicol (2001), Donald (1997), Lewbel (1991) and others, allows to have a somewhat homogeneous sample of households. With each of the selected households, we also retain the variables of interest to our study, namely, some location variables (for matching with prices) and expenditure variables, grouped into 6 categories of goods: food, health care, transportation, household, apparel (clothing) and miscellaneous goods. The total number of households which met the above criteria was 897 (out of 7860 in the CEX data set).

The CEX data set contains no information on prices. We draw prices from the ACCRA data set[2] which provides a composite price index and prices indices for 6 different categories of goods (grocery items, housing, utilities, transportation, health care and miscellaneous goods and services) for various cities across the US. We are able to associate these prices to household selected from the CEX data set by using some location variables in the CEX data set as matching variables, and also some confidential information kindly provided by the Bureau of Labor Statistics. Details on matching procedure can be obtained from the author upon request.

Though ACCRA prices are available for a few categories of goods and could be assigned for each type of expenditures considered in the demand system, we shall use only the composite price index in our study. We do so to avoid the so-called empty-space phenomenon (see, for example, pp. 59-60 in Pagan and Ullah (1999) or pp. 92-93 in Silverman (1986)): for a high-dimensional vectors Z_i, a great number of observations is needed in order to localize at a fixed value z. We expect that the simplest one-dimensional case becomes a guide to more general situations of multidimensional vectors Z_i which require larger data sets. These can be constructed, for example, as in Nicol (2001), by considering the CEX and the ACCRA data for multiple quarters and using a CPI data to account for inflation in multiple quarters.

After performing the above steps, we produce the data set of the expenditure shares Y_i, the logarithm of the income (total expenditures) X_i and the prices Z_i, $i = 1, \ldots, N$, for $N = 901$ households across the US. The shares Y_i are for 6 categories of goods and the prices Z_i are one-dimensional. For

[1] U.S. Dep. of Labor, Bureau of Labor Statistics. Consumer Expenditure Survey, 1999: Interview Survey and Detailed Expenditure Files [Computer file]. Washington, DC: U.S. Dept. of Labor, Bureau of Labor Statistics [producer], 2001. Ann Arbor, MI: Inter-University Consortium for Political and Social Research [distributor], 2001.

[2] ACCRA Cost of Living Index, Data for First Quarter 2000, ACCRA, July 2000, **33**(1). For more information, see http://www.accra.org

notational simplicity, we have divided the price indices Z_i by 100 so that $z = 100$ in the ACCRA data now corresponds to $z = 1$. The logarithm of the income X_i, rather than the income itself, was used by Donald (1997), Hausman et al (1995) and others.

Our goal is to illustrate how local rank of the constructed data set can be estimated at several values of the prices z. We cannot apply rank estimation tests of Section 4 directly to the data Y_i, X_i and Z_i because the different shares in Y_i add up to 1 and hence the perturbation terms U_i in the nonparametric model (NP) have a singular variance-covariance matrix Σ (the key Assumption 2 of Section 3 is therefore violated). As indicated in the introduction, a way out of this difficulty is to eliminate one share from the analysis. When one share is removed, the rest of the shares do not add up to 1 and it becomes reasonable to suppose that the variance-covariance matrix of the perturbation terms is non-singular. Moreover, by Lemma B.12 in Appendix B below, the local rank of a full demand system can be estimated from the local rank (related to $x^1 = 1$) of a reduced system by adding 1. To estimate local rank of the full demand system, we hence eliminate one share from the analysis, estimate the local rank (related to $x^1 = 1$) in the reduced system by using tests of Section 4 and then add 1 to ranks in the obtained results. It can be shown theoretically and is easily observed in practice that the local rank tests of Section 4 are invariant to which share is eliminated from the analysis.

Tables 6–8 present local rank estimation results for the full data set at three different values of prices $z = 1$, $z = 0.95$ and $z = 1.2$. These values were motivated by the fact that prices associated with household in the constructed data set ranged from 0.911 to 1.251. The smoothing parameters h_x and h_z take one of the values $h_x = 0.1$, 0.3 and 0.55, and $h_z = 0.05$, 0.09 and 0.15. These choices were suggested in part by the fact that $h_x = 0.55$ and $h_z = 0.09$ were the optimal smoothing parameters obtained by the generalized cross validation procedure for the data set consisting of all expenditure shares. The entries in Tables 6–7 are the p-values for the local rank tests of Section 4 applied to the constructed data set. We do not report the p-values for $L = 4$ and $L = 5$ because there are no results where the rank $L \leq 3$ is rejected (except in one extreme case mentioned above).

Rank estimation at $z = 1$										
Local rank test		$L = 1$			$L = 2$			$L = 3$		
Statistic	$h_x \setminus h_z$	0.05	0.09	0.15	0.05	0.09	0.15	0.05	0.09	0.15
$\widehat{T}_1$	0.1	0.000	0.000	0.000	0.327	0.001	0.000	0.948	0.957	0.885
	0.3	0.000	0.000	0.000	0.067	0.000	0.000	0.955	0.953	0.896
	0.55	0.000	0.000	0.000	0.051	0.000	0.000	0.980	0.986	0.971
$\widehat{T}_2$	0.1	0.000	0.000	0.000	0.055	0.000	0.000	0.653	0.493	0.763
	0.3	0.000	0.000	0.000	0.000	0.000	0.000	0.689	0.770	0.791
	0.55	0.000	0.000	0.000	0.000	0.000	0.000	0.614	0.407	0.333

Table 6: P-values in rank estimation at $z = 1$

The results of Tables 6 and 7 suggest that the local ranks of the full demand system is 3 at $z = 1$ and $z = 0.95$. This conclusion is reached for almost all considered values h_x, h_z, both tests statistics and any reasonably small significance level α.[3] Note also from Tables 6 and 7 that the p-values

[3] When the smallest considered bandwidth $h_x = 0.1$ and $h_z = 0.05$ are used at $z = 0.95$, observe from Table 7 that the hypothesis $H_0 : L \leq 3$ is rejected at the significance level $\alpha = 0.05$. In fact, in this only case, the local rank equal to 6 was estimated by using the test statistic $\widehat{T}_2(L, z)$. We do not fully understand this behavior of the test statistic $\widehat{T}_2(L, z)$.

are smaller for the statistic $\widehat{T}_2(L, z)$ which is consistent with our findings in the simulation study of Section 6. The results of Table 8, on the other hand, suggest that the local rank is 2 at $z = 1.2$. This conclusion is more evident when the test statistic $\widehat{T}_1(L, z)$ is used and, in particular, it is true at the significance level $\alpha = 0.05$ for both statistics when the smoothing parameter values $h_x = 0.55$ and $h_z = 0.09$ obtained by the generalized cross validation are used. We have also tried estimating local rank at other higher values of z and for data sets of households with other characteristics. We have found in all of these experiments that estimates of local ranks tend to become smaller as z increases.

Rank estimation at $z = 0.95$										
Local rank test		$L = 1$			$L = 2$			$L = 3$		
Statistic	$h_x \setminus h_z$	0.05	0.09	0.15	0.05	0.09	0.15	0.05	0.09	0.15
	0.1	0.000	0.000	0.000	0.493	0.103	0.000	0.993	0.983	0.942
$\widehat{T}_1$	0.3	0.000	0.000	0.000	0.136	0.005	0.000	0.997	0.988	0.949
	0.55	0.000	0.000	0.000	0.051	0.000	0.000	0.994	0.985	0.969
	0.1	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.376	0.672
$\widehat{T}_2$	0.3	0.000	0.000	0.000	0.000	0.000	0.000	0.137	0.479	0.808
	0.55	0.000	0.000	0.000	0.000	0.000	0.000	0.312	0.547	0.583

Table 7: P-values in rank estimation at $z = 0.95$

In Table 9, we also present rank estimation results when the price variable Z_i is ignored. The statistic $\widehat{T}_1$ in Table 9 refers to that used by Donald (1997) and the statistic $\widehat{T}_2$ is defined analogously to the statistic $\widehat{T}_2(L, z)$ used in this work. The results of Table 9 strongly suggest that the rank of a demand system ignoring the price variable Z_i is 3. This conclusion should not be surprising as rank 3 has been found in many other demand systems by various authors, e.g. Donald (1997), Lewbel (1991) and others.

Rank estimation at $z = 1.2$										
Local rank test		$L = 1$			$L = 2$			$L = 3$		
Statistic	$h_x \setminus h_z$	0.05	0.09	0.15	0.05	0.09	0.15	0.05	0.1	0.19
	0.1	0.242	0.040	0.026	0.556	0.276	0.250	0.752	0.586	0.599
$\widehat{T}_1$	0.3	0.213	0.068	0.030	0.505	0.591	0.355	0.767	0.913	0.918
	0.55	0.152	0.018	0.006	0.402	0.459	0.282	0.711	0.891	0.928
	0.1	0.991	0.566	0.588	0.998	0.808	0.938	0.989	0.725	0.985
$\widehat{T}_2$	0.3	0.982	0.137	0.059	0.990	0.764	0.257	0.984	0.809	0.871
	0.55	0.975	0.022	0.001	0.982	0.637	0.098	0.990	0.872	0.870

Table 8: P-values in rank estimation at $z = 1.2$

Several observations can be made from the above rank estimation results. Interestingly, different local ranks may be estimated at distinct values of prices z. This motivates the study of global rank tests and finding causes for the observed phenomenon which we intend to pursue in a future work. The role and significance of the rank estimation ignoring prices should also be further clarified.

Rank estimation ignoring z				
Statistic	h_x	$L=1$	$L=2$	$L=3$
$\widehat{T}_1$	0.1	0.000	0.000	0.552
	0.3	0.000	0.000	0.918
	0.55	0.000	0.000	0.967
$\widehat{T}_2$	0.1	0.000	0.000	0.936
	0.3	0.000	0.000	0.601
	0.55	0.000	0.000	0.598

Table 9: P-values in rank estimation ignoring z

8 Conclusions

In the present work, we provided consistent tests to determine the local rank in nonparametric models. Two tests statistics were considered: one defined as a sum of the eigenvalues and the other defined as a sum of the squared eigenvalues of a kernel-based estimator of a matrix. The asymptotics of these statistics were based on the asymptotics of the eigenvalues which was established by using Fujikoshi expansion and U-statistics techniques. Simulation study showed that the local rank tests perform fairly well and that the two tests statistics have slightly different small sample properties. Our results extend those of Donald (1997) to the case where coefficient matrices vary with covariates so that distinction between local and global ranks becomes necessary. We applied our rank estimation methods to determine local ranks in a demand system constructed by combining the CEX and the ACCRA data sets. Results obtained in applications to demand systems show importance of studying global ranks. Estimation of global ranks will be addressed in a future work.

A Proofs of principal results

Proof of Lemma 2.1: Let us show first that $\mathrm{rk}\{F(\cdot,z);x^1\} \le L$ implies that the matrix $\Gamma_{w,z}$ has $G-L$ zero eigenvalues. By Definition 1.1, we have $\mathrm{rk}\{F(\cdot,z);x^1\} \le L$ if and only if (1.4) is verified. Relation (1.4) implies that $\beta(x,z)F(x,z)x^{1'} = \beta(x,z)c(z)x^1x^{1'} + A(z)\beta(x,z)H(x,z)x^{1'}$ and, in particular, by substituting X_i for x and taking the expectation, that

$$E\beta(X_i,z)F(X_i,z)X_i^{1'} = E\beta(X_i,z)c(z)X_i^1X_i^{1'} + A(z)E\beta(X_i,z)H(X_i,z)X_i^{1'}.$$

Multiplying this relation by $(E\beta(X_i,z)X_i^1X_i^{1'})^{-1}x^1$ and subtracting it from (1.4), we obtain that

$$\widetilde{F}(x,z) = A(z)\Big(H(x,z) - E\beta(X_i,z)H(X_i,z)X_i^{1'}(E\beta(X_i,z)X_i^1X_i^{1'})^{-1}x^1\Big).$$

It follows that there are $G-L$ linearly independent vectors $c_j(z)$, $j=1,\ldots,G-L$, such that $c_j(z)'\widetilde{F}(x,z)=0$. This is equivalent to $c_j(z)'\gamma(x,z)^{1/2}\widetilde{F}(x,z)=0$ and

$$E\left(c_j(z)'\gamma(X_i,z)^{1/2}\widetilde{F}(X_i,z)\right)^2 = c_j(z)'\Gamma_{w,z}c_j(z) = 0,$$

for $j=1,\ldots,G-L$. The last relation holds if and only if the matrix $\Gamma_{w,z}$ has $G-L$ zero eigenvalues. One can, in fact, go back in the arguments above which establishes the first "if and only if" part of the lemma. The second "if and only if" part is easy to show. □

Proof of Theorem 4.1: The proof of the convergence (4.4) uses ideas of the proof of Lemma 2 in Section 2.2 of Donald (1997). To simplify notation, we set $\widetilde{K}_{ij} = \widetilde{K}_h(X_i - X_j)$, $K_{z,i} = K_h(z - Z_i)$, and

$$\overline{\Pi}_1(z)' = \Big(\frac{1}{N}\sum_{i=1}^{N} F(X_i, z)X_i^{1\prime} K_h(z - Z_i)\Big)\widehat{Q}_1(z)^{-1}, \quad \widehat{Q}_1(z) = \frac{1}{N}\sum_{i=1}^{N} X_i^1 X_i^{1\prime} K_h(z - Z_i),$$

$$\overline{U}(z)' = \Big(\frac{1}{N}\sum_{i=1}^{N} U_i X_i^{1\prime} K_h(z - Z_i)\Big)\widehat{Q}_1(z)^{-1}, \quad \overline{\Delta F}(z)' = \Big(\frac{1}{N}\sum_{i=1}^{N} \Delta F(X_i, Z_i, z)X_i^{1\prime} K_h(z - Z_i)\Big)\widehat{Q}_1(z)^{-1},$$

where $\Delta F(x_i, z_i, z) = F(x_i, z_i) - F(x_i, z)$. By using Definition 2.2 of $\widehat{\Gamma}_{w,z}$, and by writing $Y_i = F(X_i, Z_i) + U_i = F(X_i, z) + \Delta F(X_i, Z_i, z) + U_i$ and $\widehat{\Pi}_1(z)$ in (2.10) as $\widehat{\Pi}_1(z) = \overline{\Pi}_1(z) + \overline{\Delta F}(z) + \overline{U}(z)$, we can express the matrix $\widehat{\Gamma}_{w,z}$ as

$$\widehat{\Gamma}_{w,z} = \mathcal{A}_1 + \delta\mathcal{A}_2 + \delta^2\mathcal{A}_3 = A_1 + \delta(A_2 + A_2') + \delta^2(A_3 + A_3' + A_4), \tag{A.1}$$

where $\delta^{-1} = \sqrt{Nh^{m+n/2}}$, the first order term A_1 is

$$A_1 = A_{1,1} - \overline{\Pi}_1(z)' A_{1,2} - A_{1,2}'\overline{\Pi}_1(z) + \overline{\Pi}_1(z)' A_{1,3}\overline{\Pi}_1(z),$$

where

$$A_{1,1} = \frac{1}{N(N-1)}\sum_{i\neq j}^{N} F(X_i, z)F(X_j, z)'\widetilde{K}_{ij}K_{z,i}K_{z,j},$$

$$A_{1,2} = \frac{1}{N(N-1)}\sum_{i\neq j}^{N} X_i^1 F(X_j, z)'\widetilde{K}_{ij}K_{z,i}K_{z,j}, \quad A_{1,3} = \frac{1}{N(N-1)}\sum_{i\neq j}^{N} X_i^1 X_j^{1\prime}\widetilde{K}_{ij}K_{z,i}K_{z,j},$$

the second order term A_2 is

$$A_2 = \delta^{-1}(A_{2,1} + A_{2,3}) - \delta^{-1}(A_{2,2} + A_{2,4})\overline{\Pi}_1(z) + \delta^{-1}(\overline{\Delta F}(z) + \overline{U}(z))'(A_{1,3}\overline{\Pi}_1(z) - A_{1,2}),$$

where

$$A_{2,1} = \frac{1}{N(N-1)}\sum_{i\neq j}^{N} \Delta F(X_i, Z_i, z)F(X_j, z)'\widetilde{K}_{ij}K_{z,i}K_{z,j}, \quad A_{2,2} = \frac{1}{N(N-1)}\sum_{i\neq j}^{N} \Delta F(X_i, Z_i, z)X_j^{1\prime}\widetilde{K}_{ij}K_{z,i}K_{z,j},$$

$$A_{2,3} = \frac{1}{N(N-1)}\sum_{i\neq j}^{N} U_i F(X_j, z)'\widetilde{K}_{ij}K_{z,i}K_{z,j}, \quad A_{2,4} = \frac{1}{N(N-1)}\sum_{i\neq j}^{N} U_i X_j^{1\prime}\widetilde{K}_{ij}K_{z,i}K_{z,j},$$

the third order term A_3 is

$$A_3 = \delta^{-2}(A_{3,1} + A_{3,2}) + \delta^{-2}((\overline{\Delta F}(z) + \overline{U}(z))' A_{1,3} - A_{2,2} - A_{2,4})(\overline{\Delta F}(z) + \overline{U}(z)),$$

where

$$A_{3,1} = \frac{1}{2N(N-1)}\sum_{i\neq j}^{N} \Delta F(X_i, Z_i, z)\Delta F(X_j, Z_j, z)'\widetilde{K}_{ij}K_{z,i}K_{z,j}, \quad A_{3,2} = \frac{1}{N(N-1)}\sum_{i\neq j}^{N} \Delta F(X_i, Z_i, z)U_j'\widetilde{K}_{ij}K_{z,i}K_{z,j},$$

and

$$A_4 = \frac{\delta^{-2}}{N(N-1)}\sum_{i\neq j}^{N} U_i U_j'\widetilde{K}_{ij}K_{z,i}K_{z,j}.$$

We are interested in the eigenvalues of the matrix $\widehat{\Gamma}_{w,z}\widehat{\Sigma}^{-1}$. These are also the eigenvalues of the matrix $J'\widehat{\Gamma}_{w,z}J(J'\widehat{\Sigma}J)^{-1}$, where J is any orthogonal matrix (that is, $J^{-1} = J'$). The idea then is to take a special J which would allow for easier manipulations later. In order to choose such J, observe first that, by Lemma B.1 below, the matrix $A_1\Sigma^{-1}$ has $G - L(z)$ zero eigenvalues and the remaining ones are strictly positive with probability approaching 1. Since we need to show convergence in distribution, we may suppose without loss of generality that all the eigenvalues of $A_1\Sigma^{-1}$ are positive. Hence, there is an orthogonal matrix $J = J(N, z)$ such that the matrix

$$J'A_1\Sigma^{-1}J = J'A_1J(J'\Sigma J)^{-1} \tag{A.2}$$

is diagonal with the eigenvalues of $A_1\Sigma^{-1}$ on the diagonal. Since Σ is positive definite, there is an orthogonal matrix J_0 such that $J_0'\Sigma J_0 = C$, where C is a diagonal matrix. We will suppose without loss of generality that $C = I$ and hence that $J_0'\Sigma J_0 = I$. Since there is an orthogonal matrix J_1 such that $J_0J_1 = J$, we have

$$J'\Sigma J = J_1'J_0'\Sigma J_0J_1 = J_1'J_1 = I. \tag{A.3}$$

Relations (A.2) and (A.3), and the discussion above imply that the matrix $J'A_1J$ is diagonal with $G - L(z)$ zeros on the diagonal and the remaining elements on the diagonal strictly positive (with probability approaching 1). One can then arrange the matrix J as $J = (J_1\ J_2)$, where J_1 is a $G \times L(z)$ submatrix and J_2 is a $G \times (G - L(z))$ submatrix, in such a way that $J_2'A_1J_2 = 0$. Since J_2 consists of eigenvectors corresponding to zero eigenvalues of A_1, it follows from Lemma B.2 below that $A_2J_2 = 0$ and hence that $J'A_2J$ has its last $G - L(z)$ columns identically zero. Similarly, the last $G - L(z)$ rows of $J'A_2'J$ are identically zero as well. Finally, observe also that, by using (A.3), the effect of J's on the term A_4 is such that $E(J'U_iU_iJ) = I$.

By using $\widehat{\Sigma} = \Sigma + \delta B$ with $B = o_p(1)$ in Lemma B.9 below, $A_i = O_p(1)$, $i = 1, \ldots, 4$, in Lemmas B.3–B.6 below and the discussion above, $\delta^{-2}\widehat{\lambda}_j(z)$ is equal to δ^{-2} times the jth smallest eigenvalue of the matrix

$$\begin{aligned} J'\widehat{\Gamma}_{w,z}J(J'\widehat{\Sigma}J)^{-1} &= J'\widehat{\Gamma}_{w,z}J(J'\Sigma J + \delta J'BJ)^{-1} = J'\widehat{\Gamma}_{w,z}J(I + \delta J'BJ)^{-1} \\ &= J'(\mathcal{A}_1 + \delta\mathcal{A}_2 + \delta^2\mathcal{A}_3)J(I - \delta J'BJ + \delta^2 J'B^2J - \ldots) \\ &= D_1 + \delta D_2 + \delta^2 D_3 + O_p(\delta^3), \end{aligned}$$

where $D_1 = J'\mathcal{A}_1J = J'A_1J$ is diagonal, $D_2 = J'(\mathcal{A}_2 - \mathcal{A}_1B)J = O_p(1)$ and $D_3 = J'(\mathcal{A}_3 - \mathcal{A}_2B + \mathcal{A}_1B^2)J = O_p(1)$. By applying Lemma 1 in Fujikoshi (1977), we can conclude that $\widehat{\lambda}_j(z)$, $j = 1, \ldots, G - L(z)$, are also the eigenvalues of the matrix

$$0I + \delta\widetilde{D}_2 + \delta^2\widetilde{D}_3 + O_p(\delta^3), \tag{A.4}$$

where the matrices $\widetilde{D}_2$ and $\widetilde{D}_3$ are described in greater detail below.

The matrix $\widetilde{D}_2$ in (A.4) is a $(G - L(z)) \times (G - L(z))$ matrix made of the last $G - L(z)$ rows and the last $G - L(z)$ columns of the matrix $D_2 = J'\mathcal{A}_2J - J'\mathcal{A}_1BJ$. Recall from (A.1) and the discussion above that $J'\mathcal{A}_2J$ is a sum of two matrices $J'A_2J$ and $J'A_2'J$, the matrix $J'A_2J$ with its last $G - L(z)$ columns zero and the matrix $J'A_2'J$ with its last $G - L(z)$ rows zero. Hence, the $(G - L(z)) \times (G - L(z))$ matrix corresponding to $J'\mathcal{A}_2J$ is identically zero. Turning to the second term $J'\mathcal{A}_1BJ = J'\mathcal{A}_1J(J'BJ)$ in the matrix D_2, since $J'\mathcal{A}_1J$ is diagonal with its last $G - L(z)$ rows zero, we obtain that the $(G - L(z)) \times (G - L(z))$ matrix corresponding to $J'\mathcal{A}_1BJ$ is identically zero as well. Then, $\widetilde{D}_2 = 0$ and hence $\widehat{\lambda}_j(z)$, $j = 1, \ldots, G - L(z)$, are also the eigenvalues of the matrix $\delta^2\widetilde{D}_3 + O_p(\delta^3)$ or $\delta^{-2}\widehat{\lambda}_j(z) = Nh^{m+n/2}\widehat{\lambda}_j(z)$, $j = 1, \ldots, G - L(z)$, are the eigenvalues of the matrix

$$\widetilde{D}_3 + o_p(1). \tag{A.5}$$

According to Lemma 1 in Fujikoshi (1977), the matrix $\widetilde{D}_3$ in (A.5) (or (A.4)) is a sum of two matrices $\widetilde{D}_{3,1}$ and $\widetilde{D}_{3,2}$. The first term $\widetilde{D}_{3,1}$ is made of the last $G - L(z)$ rows and the last $G - L(z)$ columns of the matrix D_3. The second term $\widetilde{D}_{3,2}$ involves the sum of some submatrices of the last $G - L(z)$ rows and the last $G - L(z)$ columns of the matrix D_2. By using the facts that $A_2 = o_p(1)$, $B = o_p(1)$ and a special structure of the matrix $J'A_1J$, one can conclude that $\widetilde{D}_{3,2} = o_p(1)$. As for the matrix $\widetilde{D}_{3,1}$, by using $A_3 = o_p(1)$, we obtain that it consists of the last $G - L(z)$ rows and the last $G - L(z)$ columns of the matrix

$$J'A_4J + o_p(1) = \frac{\delta^{-2}}{N(N-1)} \sum_{i \neq j}^{N} (J'U_i)(J'U_j)'\widetilde{K}_{ij}K_{z,i}K_{z,j} + o_p(1).$$

Hence, it follows that

$$\widetilde{D}_3 = \frac{\delta^{-2}}{N(N-1)} \sum_{i \neq j}^{N} \widetilde{U}_i\widetilde{U}_j'\widetilde{K}_{ij}K_{z,i}K_{z,j} + o_p(1),$$

where a $(G - L(z)) \times 1$ vector $\widetilde{U}_j$ satisfies $E\widetilde{U}_j\widetilde{U}_j' = I$. By Lemma B.8 below, we have

$$\widehat{V}(z)\widetilde{D}_3 \xrightarrow{d} \mathcal{Z}_{G-L_0(z)}. \tag{A.6}$$

The convergence (4.4) then follows from (A.5) and (A.6) by the continuous mapping theorem.

The convergence (4.5) holds, since by the continuous mapping theorem, $\widehat{\lambda}_j(z) \to \lambda_j(z)$ in probability, where $0 \le \lambda_1(z) \le \ldots \le \lambda_G(z)$ are the eigenvalues of the matrix $\Gamma_{w,z}\Sigma^{-1}$ and, by Lemma 2.1, $\lambda_j(z) > 0$ for $j = G - L(z) + 1, \ldots, G$. $\square$

Proof of Theorem 4.2: The convergence in (4.7) follows from (4.4) in Theorem 4.1. In order to show the stochastic dominance in (4.7), we use the proof of Theorems 1 and 2 in Donald (1997). By the Poincaré separation theorem (see Magnus and Neudecker (1999), p. 209, or Rao (1973), p. 65), we have $\lambda_i(\mathcal{Z}_{G-L(z)}) \le \lambda_i(B'\mathcal{Z}_{G-L(z)}B)$ for $i = 1, \ldots, G-L$, where $L(z) = \mathrm{rk}\{F(\cdot, z); x^1\}$ and B is any $(G - L(z)) \times (G - L)$ matrix such that $B'B = I_{G-L}$. Now take $B = (0_{(G-L)\times(L-L(z))} \quad I_{G-L})'$ so that $B'B = I_{G-L}$. Observe that $B'\mathcal{Z}_{G-L(z)}B =_d \mathcal{Z}_{G-L}$ and hence

$$\frac{1}{\sqrt{G-L}} \sum_{j=1}^{G-L} \lambda_j(\mathcal{Z}_{G-L(z)}) \overset{d}{\le} \frac{1}{\sqrt{G-L}} \sum_{j=1}^{G-L} \lambda_j(\mathcal{Z}_{G-L}) = \frac{1}{\sqrt{G-L}} \mathrm{tr}(\mathcal{Z}_{G-L}) \overset{d}{=} \mathcal{N}(0,1).$$

The convergence under the hypothesis H_1 follows from (4.5) of Theorem 4.1. $\square$

Proof of Theorem 4.3: The convergence in (4.9) follows from (4.4) in Theorem 4.1. To prove the stochastic dominance in (4.9), observe first that

$$\sum_{j=1}^{G-L} (\lambda_j(\mathcal{Z}_{G-L(z)}))^2 = \sum_{j=1}^{G-L} \lambda_j(\mathcal{Z}^2_{G-L(z)}), \tag{A.7}$$

where $\lambda_j(\mathcal{Z}^2_{G-L(z)})$, $j = 1, \ldots, G - L(z)$, denote the eigenvalues of $\mathcal{Z}^2_{G-L(z)}$ in the increasing order. Letting $B = (0_{(G-L)\times(L-L(z))} \quad I_{G-L})'$ and arguing as in the proof of Theorem 4.2, we can conclude that

$$\sum_{j=1}^{G-L} \lambda_j(\mathcal{Z}^2_{G-L(z)}) \overset{d}{\le} \sum_{j=1}^{G-L} \lambda_j((B'\mathcal{Z}_{G-L(z)}B)(B'\mathcal{Z}_{G-L(z)}B)), \tag{A.8}$$

Since $B'\mathcal{Z}_{G-L(z)}B =_d \mathcal{Z}_{G-L}$, it follows from (A.7) and (A.8) that

$$\sum_{j=1}^{G-L} (\lambda_j(\mathcal{Z}_{G-L(z)}))^2 \overset{d}{\le} \sum_{j=1}^{G-L} \lambda_j(\mathcal{Z}^2_{G-L}) = \mathrm{tr}\{\mathcal{Z}^2_{G-L}\} = \mathrm{vec}(\mathcal{Z}_{G-L})'\mathrm{vec}(\mathcal{Z}_{G-L}) \overset{d}{=} \chi^2\big((G-L)(G-L+1)/2\big),$$

since $\mathcal{Z}_{G-L}$ is a symmetric matrix consisting of independent (below the diagonal) zero mean normal random variables with variance 1 on the diagonal and variance 1/2 off the diagonal (use the fact $2(\mathcal{N}(0,1/2))^2 =_d \mathcal{N}(0,1)^2$). $\square$

Proof of Theorem 5.1: The proof is similar to that of Theorem 3 in Donald (1997) or Theorem 5.2 in Robin and Smith (2000). Let $\mathcal{A}_L$ denote the event that the null hypothesis $H_0 : \mathrm{rk}\{F(\cdot, z); x^1\} \le L$ is rejected by using the statistic $\widehat{T}_1(L, z)$ at the significance level $\alpha = \alpha(N, h)$. Then, we have

$$P(\widehat{L}_1(z) = L) = P\left(\mathcal{A}_1 \cap \ldots \cap \mathcal{A}_{L-1} \cap \mathcal{A}^c_L\right), \tag{A.9}$$

where $\mathcal{A}^c_L$ denotes the complement of $\mathcal{A}_L$. Let $\mathcal{N}_{\alpha(N,h)}(0,1)$ be the minimum ξ such that $P(\mathcal{N}(0,1) \ge \xi) = \alpha(N,h)$. Obviously, $\mathcal{N}_{\alpha(N,h)}(0,1) \to \infty$ if $\alpha(N,h) \to 0$. It is also an easy exercise to see that $\mathcal{N}_{\alpha(N,h)}(0,1)/Nh^{m+n/2} \to 0$ if $(-\ln\alpha(N,h))^{1/2}/Nh^{m+n/2} \to 0$. Then, for any $L < \mathrm{rk}\{F(\cdot, z); x^1\}$, we obtain from (A.9) that

$$P(\widehat{L}_2(z) = L) \le P(\mathcal{A}^c_L) = 1 - P(\widehat{T}_1(L,z) > \mathcal{N}_{\alpha(N,h)}(0,1)) = 1 - P(\widehat{T}_1(L,z)/Nh^{m+n/2} > \mathcal{N}_{\alpha(N,h)}(0,1)/Nh^{m+n/2}) \to 0, \tag{A.10}$$

by using $\widehat{T}_1(L,z)/Nh^{m+n/2} \to Const > 0$ and $\mathcal{N}_{\alpha(N,h)}(0,1)/Nh^{m+n/2} \to \infty$. Observe also that, by setting $L(z) = \mathrm{rk}\{F(\cdot, z); x^1\}$, we have

$$P(\widehat{L}_2(z) > L(z)) \le P(\mathcal{A}_{L(z)}) \quad = \quad P(\widehat{T}_1(L(z), z) > \mathcal{N}_{\alpha(N,h)}(0,1)) \to 0, \tag{A.11}$$

by using Theorem 4.2 and since $\mathcal{N}_{\alpha(N,h)}(0,1) \to 0$. The convergence in (A.10) and (A.11) show that $P(\widehat{L}_2(z) = L(z)) \to 1$. $\square$

Proof of Theorem 5.2: The proof is similar to that of Theorem 5.1 above. Introduce $\chi^2_{\alpha(N,h)}((G-L)(G-L+1)/2)$ as the minimum ξ such that $P(\chi^2((G-L)(G-L+1)/2) \ge \xi) = \alpha(N,h)$. Observe that, by Theorem 5.8 in Pötscher (1983), we have $\chi^2_{\alpha(N,h)}((G-L)(G-L+1)/2) \to \infty$ if $\alpha(N,h) \to 0$ and $\chi^2_{\alpha(N,h)}((G-L)(G-L+1)/2) \to 0$ if $-\ln\alpha(N,h)/N^2h^{2m+n} \to 0$. $\square$

B Intermediate results

We first prove two elementary results used in Theorem 4.1 in Appendix A.

Lemma B.1 *The matrix A_1 (or the matrix $A_1\Sigma^{-1}$) in (A.1) has $G-L(z)$ zero eigenvalues and the remaining ones are positive with probability approaching* 1.

Proof: Observe that

$$A_1 = \frac{1}{N(N-1)}\sum_{i\neq j}^{N}(F(X_i,z)-\overline{\Pi}_1(z)'X_i^1)(F(X_j,z)-\overline{\Pi}_1(z)'X_j^1)'\widetilde{K}_{ij}K_{z,i}K_{z,j}. \tag{B.1}$$

By using (1.4), we have $\overline{\Pi}_1(z) = c(z) + A(z)\overline{H}_1(z)$, where $\overline{H}_1(z) = \left(N^{-1}\sum_{i=1}^N H(X_i,z)X_i^{1\prime}K_{z,i}\right)\widehat{Q}_1(z)^{-1}$ and $\widehat{Q}_1(z)$ is defined before (A.1). By using (1.4) again, we then deduce that $F(X_i,z)-\overline{\Pi}_1(z)'X_i^1 = A(z)(H(X_i,z)-\overline{H}_1(z)X_i^1)$. By substituting this relation into (B.1) we further obtain that $A_1 = A(z)H_1A(z)'$, where $H_1 = (N(N-1))^{-1}\sum_{i\neq j}^N(H(X_i,z)-\overline{H}_1(z)X_i^1)(H(X_j,z)-\overline{H}_1(z)X_j^1)'\widetilde{K}_{ij}K_{z,i}K_{z,j}$. Since $A(z)$ is a $G\times L(z)$ matrix, there are $G-L(z)$ linearly independent vectors $c_j(z)$ such that $c_j(z)A(z) = 0$. Then, $A_1c_j(z)' = A(z)H_1A(z)'c_j(z)' = 0$ for $j=1,\ldots,G-L(z)$, which shows that A_1 has $G-L(z)$ zero eigenvalues. The remaining eigenvalues are positive with probability approaching 1 because $A_1 \to_p \Gamma_{w,z}$ (Lemma B.3 below) and the matrix $\Gamma_{w,z}$ has $G-L(z)$ zero eigenvalues with the remaining ones strictly positive (Lemma 2.1 above). □

Lemma B.2 *The eigenvectors corresponding to $G-L(z)$ zero eigenvalues of the matrix A_1 in Lemma B.1 are also eigenvectors for the matrix A_2 in (A.1) corresponding to a zero eigenvalue.*

Proof: Let c be an eigenvector corresponding to a zero eigenvalue of the matrix A_1. Then, with the notation of the proof of Lemma B.1 above $cA(z) = 0$. Observe now that A_2 can be expressed as $A_2 = (\delta N(N-1))^{-1}\sum_{i\neq j}^N\left(\Delta F(X_i,Z_i,z)+U_i-\overline{\Delta F}(z)-\overline{U}(z)\right)\left(F(X_j,z)-\overline{\Pi}_1(z)'X_j^1\right)'\widetilde{K}_{ij}K_{z,i}K_{z,j}$. Then, as in the proof of Lemma B.1, $A_2 = (\delta N(N-1))^{-1}\sum_{i\neq j}^N\left(\Delta F(X_i,Z_i,z)+U_i-\overline{\Delta F}(z)-\overline{U}(z)\right)\left(H(X_j,z)-\overline{H}_1(z)X_j^1\right)'\widetilde{K}_{ij}K_{z,i}K_{z,j}\,A(z)'$. Since $cA(z)=0$, it follows that $A_2c' = 0$. □

The next four lemmas concern the orders of the terms A_1, A_2, A_3 and A_4 in the decomposition (A.1). Their proofs often use the notion of a second order U-statistic whose definition we recall in Appendix C, together with a useful result on their asymptotic behavior.

Lemma B.3 *Under the assumptions of Theorem 4.1, we have $A_1 = \Gamma_{w,z} + o_p(1)$.*

Proof: By using Lemma B.7 below, it is enough to show that

$$A_{1,1} = E\frac{p(X_i,z)^2}{\widetilde{p}(X_i)}F(X_i,z)F(X_i,z)' + o_p(1), \tag{B.2}$$

$$A_{1,2} = E\frac{p(X_i,z)^2}{\widetilde{p}(X_i)}F(X_i,z)' + o_p(1), \tag{B.3}$$

$$A_{1,3} = E\frac{p(X_i,z)^2}{\widetilde{p}(X_i)} + o_p(1), \tag{B.4}$$

where $A_{1,1}$, $A_{1,2}$ and $A_{1,3}$ are defined after (A.1). We will prove only relation (B.2) in the case $G=1$ since the proofs of (B.3) and (B.4) are similar, and the case $G\geq 2$ follows by considering matrices component-wise. Observe that $A_{1,1}$ can be expressed as a second order U-statistic (C.1) with $W_i = (X_i,Z_i)$ and $a_N(W_i,W_j) = F(X_i,z)F(X_j,z)\widetilde{K}_{ij}K_{z,i}K_{z,j}$. By using the assumptions of Theorem 4.1 and applying Lemma B.11, (a), below, we have

$$Ea_N(W_i,W_j) = E\frac{p(X_i,z)^2}{\widetilde{p}(X_i)}F(X_i,z)F(X_i,z) + o(1),$$

$$E\left(E(a_N(W_i,W_j)|W_i)^2\right) = O(h^{-m}) = o(N), \quad Ea_N(W_i,W_j)^2 = O(h^{-2m-n}) = o(N^2),$$

since $Nh^m\to\infty$ and $Nh^{m+n/2}\to\infty$. The relation (B.2) then follows from Lemma C.1 below. □

Lemma B.4 *Under the assumptions of Theorem 4.1, we have* $A_2 = o_p(1)$.

PROOF: We will argue that

$$A_{2,i} = O_p\left(h^r + \sqrt{\frac{h^2}{Nh^m}} + \sqrt{\frac{h^2}{N^2h^{2m+n}}}\right), \quad i = 1, 2, \tag{B.5}$$

$$A_{2,i} = O_p\left(\frac{1}{\sqrt{Nh^m}} + \frac{1}{\sqrt{N^2h^{2m+n}}}\right), \quad i = 3, 4, \tag{B.6}$$

where $A_{2,i}$, $i = 1, 2, 3, 4$, are defined after (A.1). Then, by using the definition of A_2, Lemma B.7 below and the relations (B.3) and (B.4), the order of A_2 can be shown to be $O_p(\sqrt{Nh^{m+n/2+2r}} + \sqrt{h^{n/2}} + 1/\sqrt{Nh^{m+n/2}} = o_p(1)$ since $Nh^{m+n/2+2r} \to 0$ and $Nh^{m+n/2} \to \infty$. Consider first the relation (B.5) with $i = 2$ and suppose for simplicity that $G = 1$ and $d = 1$. Observe that $A_{2,2}$ is a second order U-statistic (C.1) with $W_i = (X_i, Z_i)$ and $a_N(W_i, W_j) = 2^{-1}\left(\Delta F(X_i, Z_i, z)X_j^1 + \Delta F(X_j, Z_j, z)X_i^1\right)\widetilde{K}_{ij}K_{z,i}K_{z,j} =: a_{N,1}(W_i, W_j) + a_{N,2}(W_i, W_j)$. By using the assumptions of Theorem 4.1 and applying Lemma B.11, (a), below, we get $Ea_N(W_i, W_j) = O(h^r)$ and

$$E\left(E(a_N(W_i, W_j)|W_i)^2\right) \le 2E\left(E(a_{N,1}(W_i, W_j)|W_i)^2\right) + 2E\left(E(a_{N,2}(W_i, W_j)|W_i)^2\right) = O(h^{2-m} + h^{2r-m}) = O(h^{2-m}),$$

$$Ea_N(W_i, W_j)^2 \le 2Ea_{N,1}(W_i, W_j)^2 + 2Ea_{N,2}(W_i, W_j)^2 = O\left(\frac{h^2}{h^{2m+n}}\right).$$

Relation (B.5) with $i = 2$ then follows from Lemma C.1 below. The proof of (B.5) with $i = 1$ is similar. In the case of (B.6) with $i = 4$, supposing for simplicity that $G = 1$ and $d = 1$, $A_{2,4}$ is a second order U-statistic with $W_i = (Y_i, X_i, Z_i)$ and $a_N(W_i, W_j) = 2^{-1}(U_iX_j^1 + U_jX_i^1)\widetilde{K}_{ij}K_{z,i}K_{z,j} =: U_ia_{N,1}(W_i, W_j) + U_ja_{N,2}(W_i, W_j)$. By using the assumptions of Theorem 4.1 and applying Lemma B.11, (a), below again, we have $Ea_N(W_i, W_j) = 0$, $E\left(E(a_N(W_i, W_j)|W_i)^2\right) = EU_i^2\left(E(a_{N,1}(W_i, W_j)|W_i)^2\right) = O(h^{-m})$ and $Ea_N(W_i, W_j)^2 = Ea_{N,1}(W_i, W_j)^2 + Ea_{N,2}(W_i, W_j)^2 = O(h^{-2m-n})$. The conclusion follows from Lemma C.1 below. The proof of (B.6) with $i = 3$ is similar. □

Lemma B.5 *Under the assumptions of Theorem 4.1, we have* $A_3 = o_p(1)$.

PROOF: We will argue that

$$A_{3,1} = O_p\left(h^{2r} + \sqrt{\frac{h^{2r+2}}{Nh^m}} + \sqrt{\frac{h^4}{N^2h^{2m+n}}}\right), \tag{B.7}$$

$$A_{3,2} = O_p\left(\sqrt{\frac{h^{2r}}{Nh^m}} + \sqrt{\frac{h^2}{N^2h^{2m+n}}}\right), \tag{B.8}$$

where $A_{3,1}$ and $A_{3,2}$ are defined after (A.1). Then, by the definition of A_3, Lemma B.7 below and the relations (B.4), (B.5) and (B.6), we can deduce that $A_3 = O_p\left(\sqrt{Nh^{m+n/2+2r}} + 1/\sqrt{Nh^m} + h + h^{n/2}\right) = o_p(1)$ since $Nh^{m+n/2+2r} \to 0$ and $Nh^m \to \infty$. Supposing for simplicity that $G = 1$, $A_{3,1}$ is a second order U-statistic (C.1) with $W_i = (X_i, Z_i)$ and $a_N(W_i, W_j) = \Delta F(X_i, Z_i, z)\Delta F(X_j, Z_j, z)\widetilde{K}_{ij}K_{z,i}K_{z,j}$. Then, by using the assumptions of Theorem 4.1 and applying Lemma B.11, (a), below, we have $Ea_N(W_i, W_j) = O(h^{2r})$, $E\left(E(a_N(W_i, W_j)|W_i)^2\right) = O(h^{2r+2-m})$ and $Ea_N(W_i, W_j)^2 = O(h^{4-2m-n})$. Relation (B.7) then follows by using Lemma C.1. As for $A_{3,2}$, it is a second order U-statistic (C.1) with $W_i = (Y_i, X_i, Z_i)$ and $a_N(W_i, W_j) = 2^{-1}(\Delta F(X_i, Z_i, z)U_j + \Delta F(X_j, Z_j, z)U_i)\widetilde{K}_{ij}K_{z,i}K_{z,j}$. By the assumptions of Theorem 4.1, $Ea_N(W_i, W_j) = 0$ and, by using Lemma B.11, (a), below, we can show that $E\left(E(a_N(W_i, W_j)|W_i)^2\right) = O(h^{2r-m})$ and $Ea_N(W_i, W_j)^2 = O(h^{2-2m-n})$. Relation (B.8) follows from Lemma C.1 below. □

Lemma B.6 *Under the assumptions of Theorem 4.1, we have* $A_4 = O_p(1)$.

PROOF: Arguing as in the proof of Lemma B.8 below, we may show that A_4 is asymptotically normal and hence, $A_4 = O_p(1)$. □

The next result was used a number of times in the proofs of Lemmas B.3–B.6 above.

Lemma B.7 *Under the assumptions of Theorem 4.1 and with the notation before (A.1),*

$$\widehat{Q}_1(z) = Q_1(z) + o_p(1), \quad \overline{\Pi}_1(z)' = \left(EF(X_i, z) X_i^{1\prime} \frac{p(X_i, z)}{\widetilde{p}(X_i)} \right) Q_1(z)^{-1} + o_p(1) \tag{B.9}$$

and

$$\overline{\Delta F}(z) = O_p\left(h^r + \frac{h}{\sqrt{Nh^m}} \right), \quad \overline{U}(z) = O_p\left(\frac{1}{\sqrt{Nh^m}} \right). \tag{B.10}$$

Proof: We suppose for simplicity that $G = 1$ and $d = 1$. To prove the first relation in (B.9), observe that $E(\widehat{Q}_1(z) - Q_1(z))^2 = E\widehat{Q}_1(z)^2 - 2E\widehat{Q}_1(z)Q_1(z) + Q_1(z)^2$. By using the definition of $\widehat{Q}_1(z)$ and Lemma B.11, (c), below, we have $E\widehat{Q}_1(z) = Q_1(z) + o(1)$ and

$$E\widehat{Q}_1(z)^2 = \frac{1}{N} E((X_i^1)^2 K_{z,i})^2 + \frac{N-1}{N} (E(X_i^1)^2 K_{z,i})^2 = O\left(\frac{1}{Nh^m} \right) + Q_1(z)^2.$$

Hence, $E(\widehat{Q}_1(z) - Q_1(z))^2 = o(1)$ which yields the first relation in (B.9). To prove the second relation in (B.9), it is enough to show that $E(\overline{F}_1(z) - F_1(z))^2 \to 0$, where $\overline{F}_1(z) = \overline{\Pi}(z)\widehat{Q}_1(z)$ and $F_1(z) = EF(X_i, z) X_i^1 p(X_i, z)/\widetilde{p}(X_i)$. This can be done by writing $E(\overline{F}_1(z) - F_1(z))^2 = E\overline{F}_1(z)^2 - 2E\overline{F}_1(z)F_1(z) + F_1(z)^2$ and using Lemma B.11, (c), below to conclude that $E\overline{F}_1(z) = F_1(z) + o(1)$ and $E\overline{F}_1(z)^2 = E(F(X_i, z)X_i^1 K_{z,i})^2/N + (N-1)(EF(X_i, z)X_i^1 K_{z,i})^2/N = O(1/Nh^m) + F_1(z)^2$.

Relations in (B.10) can be proved in a similar way. For example, to show the first relation, it is enough to show that $E\overline{\Delta F}_1(z)^2 = O_p(h^{2r} + h^r/Nh^m)$, where $\overline{\Delta F}_1(z) = \overline{\Delta F}(z)\widehat{Q}_1(z)$. This follows by writing $E\overline{\Delta F}_1(z)^2 = E(\Delta F(X_i, Z_i, z)K_{z,i})^2/N + (N-1)(E\Delta F(X_i, Z_i, z)K_{z,i})^2/N$ and applying the Lemma B.11, (c), below to obtain $E(\Delta F(X_i, Z_i, z)K_{z,i})^2 = O(h^{2-m})$ and $E\Delta F(X_i, Z_i, z)K_{z,i} = O(h^r)$. □

We now prove an asymptotic normality result (A.6) used in the proof of Theorem 4.1.

Lemma B.8 *Under the assumptions and with the notation of Theorem 4.1 and its proof, we have*

$$\widehat{V}(z) \frac{h^{m+n/2}}{N} \sum_{i \neq j}^{N} \widetilde{U}_i \widetilde{U}_j' \widetilde{K}_{ij} K_{z,i} K_{z,j} \xrightarrow{d} \mathcal{Z}_{G-L(z)}. \tag{B.11}$$

Proof: Set $t = G - L(z)$, $\widetilde{U}_i = (\widetilde{U}_{i1}, \ldots, \widetilde{U}_{it})'$ and

$$A_{p,q}(N) = \frac{h^{m+n/2}}{N} \sum_{i \neq j}^{N} \widetilde{U}_{ip} \widetilde{U}_{jq} \widetilde{K}_{ij} K_{z,i} K_{z,j}, \quad p, q = 1, \ldots, t,$$

so that the left-hand side of (B.11) can be expressed as $\widehat{V}(z)(A_{p,q}(N))_{p,q=1,\ldots,t}$. We will show first that, for fixed p and q,

$$A_{p,q}(N) \xrightarrow{d} \mathcal{N}(0, \sigma_{p,q}^2 V(z)^{-2}), \tag{B.12}$$

where $V(z)$ is defined in the beginning of Section 4 and $\sigma_{p,q}^2 = 1$, if $p = q$, and $1/2$, if $p \neq q$. By using Lemma B.10 below, the convergence (B.11) then holds component-wise.

To show (B.12), we follow the proof of Theorem 4.5 in White and Hong (1999) (see also Lemma B.2 in Donald (1997)). Since $\widetilde{U}_i$ can be expressed in terms of $W_i = (Y_i, X_i, Z_i)$, we can write $A_{p,q}(N) = \sum_{i \neq j}^{N} \widetilde{a}_N(W_i, W_j) = \sum_{i<j} a_N(W_i, W_j)$, where $\widetilde{a}_N(W_i, W_j) = h^{m+n/2} N^{-1} \widetilde{U}_{ip} \widetilde{U}_{jq} \widetilde{K}_{ij} K_{z,i} K_{z,j}$ and $a_N(W_i, W_j) = \widetilde{a}_N(W_i, W_j) + \widetilde{a}_N(W_j, W_i)$. Observe that, for $i < j$, $E(a_N(W_i, W_j)|W_i) = 0$. Hence, by Proposition 3.2 in de Jong (1987), convergence (B.12) holds if (1) $\mathrm{Var}(A_{p,q}(N)) \to \sigma_{p,q}^2$, and (2) $G_{N,i} = o(\mathrm{Var}(A_{p,q}(N))^2) = o(1)$ for $i = 1, 2$ and 4, where

$$G_{N,1} = \sum_{1 \le i < j \le N} E a_{ij}^4, \quad G_{N,2} = \sum_{1 \le i < j < k \le N} \left(E a_{ij}^2 a_{ik}^2 + E a_{ji}^2 a_{jk}^2 + E a_{ki}^2 a_{kj}^2 \right),$$

$$G_{N,4} = \sum_{1 \le i < j < k < l \le N} \left(E a_{ij} a_{ik} a_{lj} a_{lk} + E a_{ij} a_{il} a_{kj} a_{kl} + E a_{ik} a_{il} a_{jk} a_{jl} \right) \quad \text{with } a_{ij} = a_N(W_i, W_j).$$

To show part (1), observe that, by using Lemma B.11, (b), below,

$$\mathrm{Var}(A_{p,q}(N)) = 2\sigma_{p,q}^2 \frac{h^{2m+n}}{N^2}(N-1)NE(\widetilde{K}_{ij}K_{z,i}K_{z,j})^2 = \sigma_{p,q}^2 2\|\widetilde{K}\|_2^2\|K\|_2^4 E\frac{p(X_i,z)^2}{\widetilde{p}(X_i)} + o(1) = \sigma_{p,q}^2 V(z)^{-2} + o(1).$$

As for part (2), by using the Lemma B.11, (d), below,

$$G_{N,1} \le \mathrm{Const}\frac{h^{4m+2n}}{N^4}\sum_{i\neq j} E\widetilde{K}_{ij}^4 K_{z,i}^4 K_{z,j}^4 \le \mathrm{Const}\frac{h^{4m+2n}}{N^2} E\widetilde{K}_{ij}^4 K_{z,i}^4 K_{z,j}^4 = O\left(\frac{1}{N^2h^{2m+n}}\right) = o(1),$$

$$G_{N,2} \le \mathrm{Const}\frac{h^{4m+2n}}{N} E\widetilde{K}_{ij}^2\widetilde{K}_{il}^2 K_{z,i}^4 K_{z,j}^2 K_{z,l}^2 = O\left(\frac{1}{Nh^m}\right) = o(1),$$

$$G_{N,4} \le \mathrm{Const} h^{4m+2n} E\widetilde{K}_{ij}\widetilde{K}_{ik}\widetilde{K}_{lj}\widetilde{K}_{lk} K_{z,i}^2 K_{z,j}^2 K_{z,k}^2 K_{z,l}^2 = O(h^{2n}) = o(1).$$

Arguing similarly as above, we may show that, for any $c_j \in \mathbb{R}$, $p_j, q_j \in \{1,\dots,t\}$, a linear combination $\sum_{j=1}^d c_j A_{p_j,q_j}(N)$ is asymptotically normal with the limiting variance $\sigma(\boldsymbol{p},\boldsymbol{q})^2$ characterized by

$$\mathrm{Var}\left(\sum_{j=1}^d c_j A_{p_j,q_j}(N)\right) \to \sigma(\boldsymbol{p},\boldsymbol{q})^2.$$

Since $EA_{p,q}(N)A_{p',q'}(N) = 0$ for different pairs (p,q) and (p',q'), we conclude that $\sigma(\boldsymbol{p},\boldsymbol{q})^2 = \sigma_{p_1,q_1}^2 + \ldots + \sigma_{p_d,q_d}^2$. Together with the convergence (B.12), this shows that (B.11) holds. □

The next two results were used in the proof of Theorem 4.1 to replace the variance-covariance matrix Σ and a normalizing constant $V(z)$ by their estimators $\widehat{\Sigma}$ and $\widehat{V}(z)$, respectively. (See (4.1) and (4.2) for definitions of $V(z)$ and $\widehat{V}(z)$, respectively.)

Lemma B.9 *Under the assumptions of Theorem 4.1, we have* $\widehat{\Sigma} = \Sigma + \delta B$ *with* $B = o_p(1)$.

PROOF: As shown in the proof of Lemma 2 in Donald (1997), pp. 126–127,

$$\widehat{\Sigma} = \Sigma + O_p\left(\frac{1}{\sqrt{N}} + \frac{1}{Nh^{m+n}} + h^{2r}\right).$$

By using the assumptions of Theorem 4.1 and since $\delta^{-1} = \sqrt{Nh^{m+n/2}}$, we obtain that $\widehat{\Sigma} = \Sigma + \delta B$ with $B = o_p(1)$. □

Lemma B.10 *Under the assumptions of Theorem 4.1, we have* $\widehat{V}(z) = V(z) + o_p(1)$.

PROOF: In view of (4.1) and (4.2), it is enough to show that $N^{-1}\sum_{i=1}^N \widehat{p}(X_i,Z_i)K_h(z-Z_i)$ converges in probability to $Ep(X_i,z)^2/\widetilde{p}(X_i)$. Setting $K_{ij} = K_h(Z_i - Z_j)$ and using (2.18), we can write the sum above as

$$\frac{1}{N^2}\sum_{i,j}\widetilde{K}_{ij}K_{z,i}K_{ij} = \frac{\widetilde{K}(0)K(0)}{Nh^{m+n}}\frac{1}{N}\sum_i K_{z,i} + \frac{1}{N^2}\sum_{i\neq j}^N \widetilde{K}_{ij}K_{z,i}K_{ij} =: I_1 + I_2.$$

Arguing as in the proof of Lemma B.7, we can show that $N^{-1}\sum_i K_{z,i} = O_p(1)$. Since $Nh^{m+n}\to\infty$, it follows that $I_1 = o_p(1)$. Arguing as in the proof of Lemma B.3, we obtain that $I_2 = E\widetilde{K}_{ij}K_{z,i}K_{ij} + o_p(1)$. By using Lemma B.11, ($b$), below we have $E\widetilde{K}_{ij}K_{z,i}K_{ij} = Ep(X_i,z)^2/\widetilde{p}(X_i) + o(1)$ which concludes this proof. □

The next result, used a number of times earlier, is a direct consequence of a localization property of kernel functions stated in Proposition B.1 below. We use our earlier notation $\widetilde{K}_{ij} = \widetilde{K}_h(X_i - X_j)$, $K_{z,i} = \widetilde{K}_h(z - Z_i)$ and $K_{ij} = K_h(Z_i - Z_j)$.

Lemma B.11 *Suppose that $G, H : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ are two deterministic functions with continuous bounded derivatives up to order r, K and $\widetilde{K}$ are kernels functions of order r and (X_i, Z_i) are i.i.d. random vectors satisfying, Assumption 1 of Section 3. Then, (a) for $i \neq j, EG(X_i, Z_i, z)H(X_j, Z_j, z)\widetilde{K}_{ij}K_{z,i}K_{z,j} = EG(X_i, z, z)H(X_i, z, z)p(X_i, z)^2/\widetilde{p}(X_i) + O(h^r)$, $E(G(X_i, Z_i, z)H(X_j, Z_j, z)\widetilde{K}_{ij}K_{z,i}K_{z,j})^2 = \|K\|_2^2 h^{-2m-n}EG(X_i, z, z)^2 H(X_i, z, z)^2 p(X_i, z)^2/\widetilde{p}(X_i) + O(h^{-2m-n+2})$, and $E(E(G(X_i, Z_i, z)H(X_j, Z_j, z)\widetilde{K}_{ij}K_{z,i}K_{z,j}|X_i, Z_i)^2) = O(h^{-m+2r+2})$ if $G(x_i, z, z) \equiv 0$ and $H(x_i, z, z) \equiv 0$, $O(h^{-m+2r})$ if $H(x_i, z, z) \equiv 0$, and $O(h^{-m})$, otherwise; (b) for $i \neq j, E\widetilde{K}_{ij}K_{z,i}K_{ij} = Ep(X_i, z)^2/\widetilde{p}(X_i) + O(h^r)$, $E(\widetilde{K}_{ij}K_{z,i}K_{ij})^2 = O(h^{-2m-n})$, $E(E(\widetilde{K}_{ij}K_{z,i}K_{ij}|X_i, Z_i)^2) = O(h^{-m})$, and $E(E(\widetilde{K}_{ij}K_{z,j}K_{ij}|X_i, Z_i)^2) = O(h^{-m})$; (c) $EG(X_i, Z_i, z)K_{z,i} = EG(X_i, z, z)p(X_i, z)/\widetilde{p}(X_i) + O(h^r)$, $E(G(X_i, Z_i, z)K_{z,i})^2 = EG(X_i, z, z)^2 p(X_i, z)/\widetilde{p}(X_i) + O(h^2)$; (d) for $i < j < k < l$, $E(\widetilde{K}_{ij}K_{z,i}K_{z,j})^4 = O(h^{-6m-3n})$, $E(\widetilde{K}_{ij}^2\widetilde{K}_{il}^2K_{z,i}^4K_{z,j}^2K_{z,l}^2) = O(h^{-5m-2n})$ and $E\widetilde{K}_{ij}\widetilde{K}_{ik}\widetilde{K}_{lj}\widetilde{K}_{lk}K_{z,i}^2K_{z,j}^2K_{z,k}^2K_{z,l}^2 = O(h^{-4m})$.*

Proof: The results of the lemma are consequences of Proposition B.1 below and the assumptions of the lemma. To show the first result of (a), observe that its left-hand side is

$$\int \left(\int G(x_i, z_i, z)H(x_j, z_j, z)p(x_i, z_i)p(x_j, z_j)\widetilde{K}_h(x_i - x_j)K_h(z - z_j)K_h(z - z_i)dx_j dz_j dz_i \right) dx_i$$

$$= \int G(x_i, z, z)H(x_i, z, z)p(x_i, z)^2 dx_i + O(h^r),$$

which is also its right-hand side. The second result of (a) follows from the first one since $\widetilde{K}_h(x)^2 = h^{-n}\|\widetilde{K}\|_2^2\widetilde{K}_{2,h}(x)$ and $K_h(z)^2 = h^{-m}\|K\|_2^2K_{2,h}(z)$, where $\widetilde{K}_{2,h}(x) = h^{-n}\widetilde{K}(x/h)^2/\|\widetilde{K}\|_2^2$ and $K_{2,h}(z) = h^{-m}K(z/h)^2/\|K\|_2^2$ are kernel functions of order 2. The third result of (a) can be obtained by observing that its left-hand side is

$$h^{-m}\|K\|_2^2 \int G(x_i, z_i, z)p(x_i, z_i)K_{2,h}(z - z_i) \left(\int H(x_j, z_j, z)p(x_j, z_j)K_h(x_i - x_j)K_h(z - z_j)dx_j dz_j \right)^2 dx_i dz_i.$$

When $H(x_i, z, z) \equiv 0$, for example, the inner integral squared above is $O(h^{2r})$ and hence the full integral is $O(h^{2r-m}) \int G(x_i, z_i, z)p(x_i, z_i)K_{2,h}(z - z_i) = O(h^{2r-m+2})$ if $G(x_i, z, z) \equiv 0$, and $O(h^{2r-m})$, otherwise. The first three results of part (b) can be shown similarly as in part (a). The last result of (b) can be proved by observing that its left-hand side is

$$\int p(x_i, z_i) \left(\int p(x_j, z_j)\widetilde{K}_h(x_i - x_j)K_h(z - z_j)K_h(z_i - z_j)dx_j dz_j \right)^2 dx_i dz_i$$

and

$$\int p(x_j, z_j)\widetilde{K}_h(x_i - x_j)K_h(z - z_j)K_h(z_i - z_j)dx_j dz_j = O\left(\int p(x_j, z_j)|\widetilde{K}_h(x_i - x_j)|h^{-2m}1_{\{|z-z_j| \le Ch\}}1_{\{|z_i-z_j| \le Ch\}}dx_j dz_j \right)$$

$$= O\left(\int p(x_j, z_j)|\widetilde{K}_h(x_i - x_j)|h^{-m}1_{\{|z-z_j| \le Ch\}}dx_j dz_j \right) h^{-m}1_{\{|z_i-z| \le Ch\}} = O(1)h^{-m}1_{\{|z_i-z| \le Ch\}}.$$

The results of parts (c) and (d) can be proved in a similar, in fact, much simpler way. □

The following localization property of kernel functions can be easily proved by using Taylor expansions and the definition of the order of a kernel function. We omit its prove for shortness sake.

Proposition B.1 *Let K be a kernel on $\mathbb{R}^m$ of order $r \in \mathbb{N}$. Suppose that a function $g : \mathbb{R}^m \to \mathbb{R}$ is $r-$times continuously differentiable in a neighborhood of $z_0 \in \mathbb{R}^m$. Then, as $h \to 0$,*

$$\int_{\mathbb{R}^m} g(z)K_h(z - z_0)dz = g(z_0) + O(h^r). \tag{B.13}$$

Moreover, if the function g has its r-order derivatives bounded on $\mathbb{R}^m$, then the term $O(h^r)$ in (B.13) does not depend on z_0.

The next lemma, implicit in Donald (1997), was used in Section 7 to argue that local rank of a demand system can be estimated from the local rank of a reduced demand system.

Lemma B.12 *Let $f(x,z) = (f_1(x,z),\dots,f_J(x,z))'$ be a $J\times 1$ vector of functions such that $\sum_{j=1}^J f_j(x,z)=1$. Then,*

$$\mathrm{rk}\{f(\cdot,z)\} = \mathrm{rk}\{F(\cdot,z);1\}+1, \tag{B.14}$$

where $F(x,z)$ is a $(J-1)\times 1$ vector obtained by removing an arbitrary coordinate function $f_j(x,z)$ from the vector $f(x,z)$ and the local rank $\mathrm{rk}\{F(\cdot,z);1\}$ is defined in Definition 1.1.

PROOF: Suppose without loss of generality that the coordinate function $f_1(x,z)$ is eliminated. Set $L(z)=\mathrm{rk}\{f(\cdot,z)\}$ and let $F^{(1)}(x,z)$ denote the vector $f(x,z)$ with the coordinate function $f_1(x,z)$ eliminated. By Definition 1.1,

$$f(x,z) = a(z)h(x,z), \tag{B.15}$$

where $a(z)=(a_{kl}(z))$ is a $J\times L(z)$ matrix and $h(x,z)=(h_l(x,z))$ is a $L(z)\times 1$ vector. Since the J shares add up to 1, we obtain from (B.15) that

$$1 = \left(\sum_{k=1}^J a_{k1}(z)\right)h_1(x,z)+\dots+\left(\sum_{k=1}^J a_{kL(z)}(z)\right)h_{L(z)}(x,z).$$

Suppose, for example, that $\sum_{k=1}^J a_{k1}(z)\neq 0$. Then, we have

$$h_1(x,z) = \left(\sum_{k=1}^J a_{k1}(z)\right)^{-1} - \left(\sum_{k=1}^J a_{k1}(z)\right)^{-1}\left(\sum_{k=1}^J a_{k2}(z)\right)h_2(x,z) - \ldots - \left(\sum_{k=1}^J a_{k1}(z)\right)^{-1}\left(\sum_{k=1}^J a_{kL(z)}(z)\right)h_{L(z)}(x,z)$$

Substituting this expression into (B.15), we conclude that

$$F^{(1)}(x,z) = c(z) + A(z)H(x,z), \tag{B.16}$$

where $A(z)$ is a $(J-1)\times(L(z)-1)$ matrix, $H(x,z)$ is a $(L(z)-1)\times 1$ vector and $c(z)$ is a $(J-1)\times 1$ vector. In view of Definition 1.1, (B.16) implies that

$$\mathrm{rk}\{F^{(1)}(\cdot,z);1\}\le L(z)-1. \tag{B.17}$$

To show the converse, observe that, by using (1.4), the elements $f_2(x,z),\dots,f_J(x,z)$ of $F^{(1)}(x,z)$ can be expressed as linear combinations of $\mathrm{rk}\{F^{(1)}(\cdot,z);1\}+1$ functions of x and z. Since $f_1(x,z)=1-f_2(x,z)-\ldots-f_J(x,z)$, the function $f_1(x,z)$ can be also expressed as a linear combination of these $\mathrm{rk}\{F^{(1)}(\cdot,z);1\}+1$ functions. In view of Definition 1.1, we obtain that

$$L(z) = \mathrm{rk}\{f(\cdot,z)\}\le \mathrm{rk}\{F^{(1)}(\cdot,z);1\}+1. \tag{B.18}$$

The conclusion follows from (B.17) and (B.18). □

C Asymptotics for second order U-statistics

The notion of a second order U-statistic was used numerous time above.

Definition C.1 Let W_i, $i=1,\dots,N$, be i.i.d. random vectors in $\mathbb{R}^d$ and $a_N:\mathbb{R}^d\times\mathbb{R}^d\mapsto\mathbb{R}$ be a symmetric kernel (that is, $a_N(x,y)=a_N(y,x)$). A second order U-statistic for the sequence $\{W_i\}$ is defined as

$$U_N = \frac{2}{(N-1)N}\sum_{1\le i<j\le N} a_N(W_i,W_j). \tag{C.1}$$

The following useful result concerns the limit behavior of a second order U-statistic. Although it easily follows from the proof of Lemma 3.1 in Powell, Stock and Stoker (1989), the result is often easier to use and yields stronger results than a direct application of Lemma 3.1 in Powell et al. (1989) itself.

Lemma C.1 *Let U_N be a second order U-statistic defined by (C.1). Then,*

$$U_N = Ea_N(W_i,W_j) + O_p\left(\sqrt{\frac{E(E(a_N(W_i,W_j)|W_i)^2)}{N}}+\sqrt{\frac{Ea_N(W_i,W_j)^2}{N^2}}\right). \tag{C.2}$$

Proof: Let

$$\widehat{U}_N = Ea_N(W_i,W_j) + \frac{2}{N}\sum_{i=1}^{N}(E(a_N(W_i,W_j)|W_i) - Ea_N(W_i,W_j)) \tag{C.3}$$

be the so-called projection of the U-statistic U_N (see Serfling (1980) or Powell et al. (1989)). Then, as in the proof of Lemma 3.1 in Powell et al. (1989),

$$E(U_N - \widehat{U}_N)^2 = \frac{2}{(N-1)N}\sum_{1\le i<j\le N} Eb_N(W_i,W_j)^2,$$

where $b_N(W_i,W_j) = a_N(W_i,W_j) - E(a_N(W_i,W_j)|W_i) - E(a_N(W_i,W_j)|W_j) + Ea_N(W_i,W_j)$. Since $Eb_N(W_i,W_j)^2 = O(Ea_N(W_i,W_j)^2)$, we obtain that $E(U_N - \widehat{U}_N)^2 = O(Ea_N(W_i,W_j)^2/N^2)$ or

$$U_N - \widehat{U}_N = O_p\left(\sqrt{\frac{Ea_N(W_i,W_j)^2}{N^2}}\right). \tag{C.4}$$

By the independence of $E(a_N(W_i,W_j)|W_i)$ for different i's and by using the formula $E(\xi - E\xi)^2 \le E\xi^2$, we have

$$\begin{aligned} E\Big(\frac{2}{N}\sum_{i=1}^{N}(E(a_N(W_i,W_j)|W_i) - Ea_N(W_i,W_j))\Big)^2 &= \frac{4}{N}E\Big(E(a_N(W_i,W_j)|W_i) - Ea_N(W_i,W_j)\Big)^2 \\ &\le \frac{4E(E(a_N(W_i,W_j)|W_i)^2)}{N}. \end{aligned}$$

The result (C.2) then follows from (C.3) and (C.4). □

References

Bauer, P., Pötscher, B. M. & Hackl, P. (1988), 'Model selection by multiple test procedures', *Statistics* **19**(1), 39–44.

Cragg, J. G. & Donald, S. G. (1993), 'Testing identifiability and specification in instrumental variable models', *Econometric Theory* **9**(2), 222–240.

Cragg, J. G. & Donald, S. G. (1996), 'On the asymptotic properties of LDU-based tests of the rank of a matrix', *Journal of the American Statistical Association* **91**(435), 1301–1309.

Cragg, J. G. & Donald, S. G. (1997), 'Inferring the rank of a matrix', *Journal of Econometrics* **76**(1-2), 223–250.

de Jong, P. (1987), 'A central limit theorem for generalized quadratic forms', *Probability Theory and Related Fields* **75**(2), 261–277.

Donald, S. G. (1997), 'Inference concerning the number of factors in a multivariate nonparametric relationship', *Econometrica* **65**(1), 103–131.

Donald, S. G., Fortuna, N. & Pipiras, V. (2004*a*), Local rank tests for smooth semiparametric factor models, In preparation.

Donald, S. G., Fortuna, N. & Pipiras, V. (2004*b*), Rank estimation in symmetric matrices with some applications, In preparation.

Fujikoshi, Y. (1977), Asymptotic expansions for the distributions of some multivariate tests, *in* 'Multivariate analysis, IV (Proc. Fourth Internat. Sympos., Dayton, Ohio, 1975)', North-Holland, Amsterdam, pp. 55–71.

Gill, L. & Lewbel, A. (1992), 'Testing the rank and definiteness of estimated matrices with applications to factor, state-space and ARMA models', *Journal of the American Statistical Association* **87**(419), 766–776.

Gorman, W. M. (1981), Some Engel curves, *in* A. Deaton, ed., 'Essays in the theory and measurement of consumer behaviour: in honour of Sir Richard Stone', Cambridge University Press, Cambridge, pp. 7–29.

Hausman, J. A., Newey, W. K. & Powell, J. L. (1995), 'Nonlinear errors in variables: estimation of some Engel curves', *Journal of Econometrics* **65**(1), 205–233.

Lewbel, A. (1989), 'A demand system rank theorem', *Econometrica* **57**(3), 701–705.

Lewbel, A. (1991), 'The rank of demand systems: theory and nonparametric estimation', *Econometrica* **59**(3), 711–730.

Lewbel, A. & Perraudin, W. (1995), 'A theorem on portfolio separation with general preferences', *Journal of Economic Theory* **65**(2), 624–626.

Magnus, J. R. & Neudecker, H. (1999), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley & Sons Ltd., Chichester. Revised reprint of the 1988 original.

Nicol, C. J. (2001), 'The rank and model specification of demand systems: an empirical analysis using United States microdata', *Canadian Journal of Economics* **34**(1), 259–289.

Pagan, A. & Ullah, A. (1999), *Nonparametric Econometrics*, Cambridge University Press, Cambridge.

Pötscher, B. M. (1983), 'Order estimation in ARMA-models by Lagrangian multiplier tests', *The Annals of Statistics* **11**(3), 872–885.

Powell, J. L., Stock, J. H. & Stoker, T. M. (1989), 'Semiparametric estimation of index coefficients', *Econometrica* **57**(6), 1403–1430.

Rao, C. R. (1973), *Linear Statistical Inference and its Applications*, Second edn, John Wiley & Sons, New York-London-Sydney. Wiley Series in Probability and Mathematical Statistics.

Robin, J.-M. & Smith, R. J. (2000), 'Tests of rank', *Econometric Theory* **16**(2), 151–175.

Russell, T. & Farris, F. (1993), 'The geometric structure of some systems of demand equations', *Journal of Mathematical Economics* **22**(4), 309–325.

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons Inc., New York. Wiley Series in Probability and Mathematical Statistics.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, New York.

White, H. & Hong, Y. (1999), M-testing using finite and infinite dimensional parameter estimators, Discussion paper 93-01R, Department of Economics, University of California, San Diego.

Natércia da Silva Fortuna
Faculdade de Economia do Porto
Rua Dr. Roberto Frias
4200–464 Porto, Portugal
nfortuna@fep.up.pt

Empirical distribution of $\widehat{L}$ using $\widehat{T}_2$ ($N = 750$, $\alpha = 0.05$)											
True rank			$L_0 = 1$			$L_0 = 2$			$L_0 = 3$		
δ	h_x	h_z	$\widehat{L} = 1$	$\widehat{L} = 2$	$\widehat{L} = 3$	$\widehat{L} = 1$	$\widehat{L} = 2$	$\widehat{L} = 3$	$\widehat{L} = 1$	$\widehat{L} = 2$	$\widehat{L} = 3$
1	0.1	0.1	0.826	0.079	0.095	0.675	0.222	0.103	0.096	0.682	0.222
		0.3	0.921	0.059	0.020	0.508	0.454	0.038	0.000	0.189	0.811
		0.5	0.877	0.113	0.010	0.322	0.649	0.029	0.000	0.039	0.961
	0.3	0.1	0.935	0.055	0.010	0.739	0.248	0.013	0.047	0.672	0.281
		0.3	0.949	0.050	0.001	0.377	0.612	0.011	0.000	0.091	0.909
		0.5	0.811	0.184	0.005	0.174	0.816	0.010	0.000	0.008	0.992
	0.5	0.1	0.953	0.046	0.001	0.751	0.248	0.001	0.032	0.703	0.265
		0.3	0.937	0.063	0.000	0.351	0.645	0.004	0.000	0.092	0.908
		0.5	0.780	0.217	0.003	0.166	0.825	0.009	0.000	0.008	0.992
1/2	0.1	0.1	0.816	0.063	0.100	0.792	0.094	0.108	0.638	0.280	0.081
		0.3	0.936	0.048	0.016	0.858	0.124	0.018	0.189	0.717	0.094
		0.5	0.923	0.067	0.010	0.834	0.154	0.012	0.006	0.738	0.256
	0.3	0.1	0.931	0.047	0.010	0.896	0.090	0.006	0.584	0.390	0.025
		0.3	0.957	0.043	0.000	0.852	0.146	0.002	0.064	0.770	0.166
		0.5	0.935	0.064	0.001	0.747	0.252	0.001	0.000	0.642	0.358
	0.5	0.1	0.940	0.046	0.000	0.928	0.066	0.003	0.554	0.441	0.005
		0.3	0.962	0.038	0.000	0.834	0.164	0.002	0.056	0.783	0.161
		0.5	0.906	0.094	0.000	0.791	0.207	0.002	0.000	0.650	0.350

Table 10: Empirical distribution of $\widehat{L}$ using $\widehat{T}_2$

Empirical distribution of $\widehat{L}$ using $\widehat{T}_2$ ($N = 1500$, $\alpha = 0.05$)											
True rank			$L_0 = 1$			$L_0 = 2$			$L_0 = 3$		
δ	h_x	h_z	$\widehat{L} = 1$	$\widehat{L} = 2$	$\widehat{L} = 3$	$\widehat{L} = 1$	$\widehat{L} = 2$	$\widehat{L} = 3$	$\widehat{L} = 1$	$\widehat{L} = 2$	$\widehat{L} = 3$
1	0.1	0.1	0.924	0.048	0.028	0.627	0.327	0.046	0.003	0.412	0.585
		0.3	0.951	0.039	0.010	0.153	0.807	0.040	0.000	0.005	0.995
		0.5	0.799	0.185	0.016	0.033	0.939	0.028	0.000	0.000	1.000
	0.3	0.1	0.968	0.031	0.001	0.541	0.450	0.009	0.000	0.288	0.712
		0.3	0.957	0.043	0.000	0.060	0.937	0.003	0.000	0.001	0.999
		0.5	0.622	0.376	0.002	0.008	0.959	0.033	0.000	0.000	1.000
	0.5	0.1	0.966	0.034	0.000	0.522	0.474	0.004	0.000	0.286	0.714
		0.3	0.944	0.056	0.000	0.047	0.949	0.004	0.000	0.001	0.999
		0.5	0.547	0.451	0.002	0.007	0.946	0.047	0.000	0.000	1.000
1/2	0.1	0.1	0.928	0.042	0.030	0.877	0.096	0.027	0.428	0.516	0.056
		0.3	0.953	0.036	0.011	0.789	0.192	0.019	0.007	0.608	0.385
		0.5	0.898	0.084	0.018	0.662	0.320	0.018	0.000	0.363	0.637
	0.3	0.1	0.967	0.033	0.000	0.879	0.118	0.003	0.257	0.672	0.071
		0.3	0.962	0.038	0.000	0.695	0.303	0.002	0.000	0.463	0.537
		0.5	0.886	0.114	0.000	0.521	0.475	0.004	0.000	0.193	0.807
	0.5	0.1	0.970	0.030	0.000	0.883	0.117	0.000	0.198	0.739	0.063
		0.3	0.963	0.037	0.000	0.685	0.315	0.000	0.000	0.475	0.525
		0.5	0.867	0.133	0.000	0.519	0.476	0.005	0.000	0.194	0.806

Table 11: Empirical distribution of $\widehat{L}$ using $\widehat{T}_2$

Size of tests using $\widehat{T}_2$								
Local rank			1 ($z=-1/2$)			2 ($z=0$)		
N	δ	$h_x \setminus h_z$	0.1	0.3	0.5	0.1	0.3	0.5
750	1	0.1	17.4	7.9	12.3	10.3	3.8	2.9
		0.3	6.5	5.1	18.9	1.3	1.1	1.0
		0.5	4.7	6.3	22.0	0.1	0.4	0.9
	1/2	0.1	16.3	6.4	7.7	10.8	1.8	1.2
		0.3	5.7	4.3	6.5	0.6	0.2	0.1
		0.5	4.6	3.8	9.4	0.3	0.2	0.2
1500	1	0.1	7.6	4.9	20.1	4.6	4.0	2.8
		0.3	3.2	4.3	37.8	0.9	0.3	3.3
		0.5	3.4	5.6	45.3	0.4	0.4	4.7
	1/2	0.1	7.2	4.7	10.2	2.7	1.9	1.8
		0.3	3.3	3.8	11.4	0.3	0.2	0.4
		0.5	3.0	3.7	13.3	0.0	0.0	0.5

Table 12: Size of local rank tests using $\widehat{T}_2$

Power of tests using $\widehat{T}_2$												
True local rank			$L_0=2$			$L_0=3$						
Local rank test			$L=1$			$L=1$			$L=2$			
N	δ	$h_x \setminus h_z$	0.1	0.3	0.5	0.1	0.3	0.5	0.1	0.3	0.5
750	1	0.1	14.4	44.6	53.5	80.9	100.0	100.0	12.4	80.9	96.2
		0.3	23.9	62.3	58.1	94.6	100.0	100.0	28.5	92.1	99.6
		0.5	25.2	60.8	50.5	96.9	100.0	100.0	31.8	93.9	99.6
	1/2	0.1	5.2	11.6	12.9	21.5	77.9	98.9	1.7	7.9	25.1
		0.3	8.3	15.6	22.9	38.2	93.8	100.0	2.9	21.0	41.8
		0.5	7.0	20.1	14.3	45.4	95.6	100.0	1.9	23.6	43.5
1500	1	0.1	34.4	85.6	87.9	99.7	100.0	100.0	55.3	99.5	100.0
		0.3	50.7	94.4	87.3	100.0	100.0	100.0	74.3	99.9	100.0
		0.5	54.4	94.8	80.6	100.0	100.0	100.0	76.4	100.0	100.0
	1/2	0.1	10.3	22.0	22.9	53.5	99.4	100.0	3.6	37.7	65.1
		0.3	14.3	33.7	32.2	77.7	100.0	100.0	9.6	60.7	85.8
		0.5	16.0	36.6	22.9	85.5	100.0	100.0	10.1	62.9	88.1

Table 13: Power of local rank tests using $\widehat{T}_2$