BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

LOCAL AND GLOBAL RANK TESTS WITH APPLICATIONS TO DEMAND SYSTEMS

by

NATÉRCIA DA SILVA FORTUNA

M.A., Boston University, 2002
Licenciatura em Economia, Universidade Lusíada, 1994

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2002

UMI Number: 3040703

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

Approved by

First Reader

Ashis K. Gangopadhyay, Ph.D.
Associate Professor of Mathematics
Boston University

Second Reader

Stephen G. Donald, Ph.D.
Professor of Economics
University of Texas, Austin

Acknowledgments

I would like to thank my advisor, Stephen G. Donald, for suggesting the problem, for his constant support and for a careful reading and corrections of the manuscript. I would also like to thank the other reader of the thesis, Ashis K. Gangopadhyay, and the other members of my committee, Mark Glickman, Murad S. Taqqu and Daniel Weiner. Financial support from Fundação para a Ciência e Tecnologia and Faculdade de Economia do Porto is gratefully acknowledged.

I could not have completed this thesis without the help of my parents, Laura and António, my grandparents, Berta, Salvador, Gracinda and António, my sister, Valquíria, my brother-in-law, Camarinha, and my love, Liaga.

LOCAL AND GLOBAL RANK TESTS WITH APPLICATIONS TO DEMAND SYSTEMS

(Order No.)

NATÉRCIA DA SILVA FORTUNA

Boston University Graduate School of Arts and Sciences, 2002

Major Professor: Ashis K. Gangopadhyay, Associate Professor of Mathematics

Abstract

In this thesis, we focus on estimation of local and global ranks of demand systems. Demand systems have been among central objects of study in Economic Theory since as early as the 19th century. They are functional relations $y = f(x, z)$ where y are expenditures of a consumer for some groups of goods, x is consumer's total income and z are prices of these groups of goods faced by the consumer. The interest is in the function f which characterizes the relation. Local and global ranks are its characteristics, essentially the minimum number of functions needed to explain its structure.

We consider two statistical models of demand systems, namely, a semi-parametric factor model and a non-parametric model, and focus on estimation of their local and global ranks. The major departure from the earlier statistical work is that we allow for price variable z to enter into the models, and hence distinction between local and global ranks becomes necessary. The inclusion of prices is meaningful because the data available to researchers cover households across the United States and it is clear that, for example, those living in New York face different prices from those residing in Minneapolis.

Since the two models involve unknown functions, we first introduce and study their estimators. These estimators are then used to provide statistical tests to determine the local ranks of the two models of demand systems. In the case of the semi-parametric factor

model, we apply to our context known rank estimation methods such as the minimum χ^2, asymptotic least squares, or that based on the so-called LDU decomposition. In the case of the non-parametric model, the tests are novel.

We apply our estimators of local ranks to economic data which is constructed by using the CEX expenditure surveys data and the price data published by the ACCRA organization. We also perform some simulations to support conclusions made in applications.

Global rank tests are discussed in length but no formal tests are provided. We explain the difficulties behind the global tests and outline some possible approaches to their construction.

Contents

List of Tables

List of Figures

List of Abbreviations

CEX	Consumer Expenditure Survey
FES	Family Expenditure Survey
ACCRA	American Chamber of Commerce Researchers Association
UMP	Utility Maximization Principle
CMP	Cost Minimization Principle
GPF	Gorman Polar Form
PIGLOG	Price Independent Generalized Logarithmic
QAID	Quadratic Almost Ideal
QES	Quadratic Expenditure
QUAIDS	Quadratic Logarithmic
AIQL	Almost Ideal Quadratic Logarithmic
NLP	Nearly Log Polynomial
NP	Non-parametric model
SPF	Semi-parametric factor model
LDU	Lower-Diagonal-Upper
ALS	Asymptotic Least Squares
SVD	Singular Value Decomposition
GMM	Generalized Method of Moments
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
MSA	Metropolitan Statistical Unit
PSU	Principal Statistical Unit

Notation

Y_i, Z_i, X_i	Vectors of shares, prices and total income, respectively
$p(z)$, $\widetilde{p}(x)$, $p(x,z)$	Densities of Z_i, X_i and (X_i, Z_i), respectively
U_i	Noise variables
N	Sample size
rk, adrk	Rank and adjusted rank
$\widehat{\Sigma}$	Estimator of a covariance matrix of the noise variables
$\widehat{\Theta}(z)$	Estimator for semi-parametric factor model
$\widehat{\Gamma}_{w,z}$	Estimator for local tests in non-parametric model
K, $\widetilde{K}$	Kernel functions
h	Bandwidth
K_h, $\widetilde{K}_h$	Kernel functions scaled by a bandwidth h
$\otimes$, vec, tr	Kronecker product, vector operation and trace of a matrix
I_n, $0_{m\times n}$	$n \times n$ identity matrix and $m \times n$ matrix with 0 entries
$\overset{d}{\leq}$	Stochastic dominance
$\chi^2(k)$	Chi-square distribution with k degrees of freedom
$\mathcal{N}(\mu, \Sigma)$	Gaussian distribution with a location vector μ and a variance-covariance matrix Σ
$\mathcal{Y}_{n\times m}$	$n \times m$ matrix with independent $\mathcal{N}(0,1)$ entries
$\mathcal{Z}_n$	$n \times n$ symmetric matrix with independent $\mathcal{N}(0,1)$ entries on the diagonal and $\mathcal{N}(0,1/2)$ entries off the diagonal

Chapter 1

Introduction

There are many situations in statistics, mathematics and other natural or even social sciences where one wants to define and then to measure the dimension of an object. Well-known examples are the dimension of a geometrical object in a Euclidean space, the rank of a matrix or the dimension of a vector space spanned by a set of functions. The knowledge of a dimension can be used, depending on the context, to understand better the object of interest, to represent it or to eliminate redundant information.

One area where the notion of a dimension arises naturally and has important implications, is the theory of demand systems in economics. Imagine a typical consumer or a household purchasing goods over a period of time. Since goods are of great variety, consider them in groups of similar goods, like food, health, transportation, clothing, recreation and others. For a particular group of goods, a consumer then allocates a share of her/his income. A demand system explains how these expenditures for groups of goods, called budget shares of goods, are related to total expenditures (income) of a consumer. More precisely, it is a functional relation $y = f(x, z)$ where y is a vector of budget shares of goods, z is a vector of prices of goods faced by a consumer and x is the total income of a consumer. We will assume hereafter that there are J groups of goods so that the dimension of a vector y or a vector f is J. The variable z can be not only a vector of prices but also a vector of demographic or household characteristics, like the age of a consumer or the total number of children, or other non-income characteristics which affect preferences of a consumer or a household. Demand systems have been extensively studied in theory, starting with Engel's work as early as the end of the 19th century, and they have been also widely used in applications.

We will focus on a particular property of a demand system, namely, its dimension which is called a rank of a demand system. The notion of a rank was introduced by Gorman [41] and developed further by Lewbel [63]. It has attracted a growing interest among researchers in economics and in statistics. To understand the rank of a demand system, consider its functional form $y = f(x, z)$. Since f is a $J \times 1$ vector, its coordinates are J functions $f_1(x, z), \ldots, f_J(x, z)$. The local (at z) rank of a demand system is then defined as a dimension of the space spanned by the functions $f_j(x, z)$, $j = 1, \ldots, J$, for a fixed z. The global rank of a demand system is defined as the maximum over z of all local ranks of a demand system.

While the definition of rank may seem easy from a mathematical point of view, it has important implications for Economic Theory. For example, Lewbel [63] has shown that functional structure and aggregation of demand systems are implications of their rank. Functional structure refers to an explicit form of the function f in a demand system relation, for example, $f(x, z) = a(z) + b(z)x + c(z) \ln x$. Such forms are deduced from a rank of demand systems by using general, utility function based, Economic Theory. Aggregation properties of a demand system determine how one can go from a demand system for an individual consumer to a demand system for the market as whole.

Parallel to understanding its implications for Economic Theory, the rank of a demand system has been also extensively studied on statistical and empirical levels. Since this constitutes an important conceptual departure from the deterministic situation discussed above, we need to make some preliminary observations. A statistical model for a demand system can be 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 total income and the prices faced by (or demographic characteristics of) the ith consumer, and ϵ_i is the noise term. The rank of a stochastic demand system is then defined in the same way as in the deterministic situation by using the coordinate functions of a vector $f(x, z)$.

Two types of problems have been related to statistical inference concerning the rank of demand systems. In the first type of problems, the function f in a demand system relation is supposed to have a parametric or a semi-parametric form, for example, $f(x, z) = \sum_{r=1}^{d} \theta_r(z)(\ln x)^{r-1} = \theta(z)V(x)$, where θ is a $J \times d$ matrix and $V(x) = (1, \ln x, \ldots, (\ln x)^{d-1})$ is a $d \times 1$ vector. The specific form of the function f depends on one's beliefs on what best represents a real life situation, or it can be chosen from a large library of functions already used to model demand systems independently of the theory of their ranks. Under a (semi-)parametric model, one can typically express the rank of a demand system as a rank of some matrix. For example, in the last example $f(x, z) = \theta(z)V(x)$, the (local) rank of a demand system is the rank of a matrix $\theta(z)$. The problem then is that of determining the rank of a matrix estimated from the observations. This problem was first addressed by Hsu [53] and by Anderson [7, 8] under normality assumptions. Their work was extended to more general situations by Gill and Lewbel [38], and Cragg and Donald [20] leading to what is now known as the LDU based test and the minimum-χ^2 test for the rank of a matrix. In the second type of problems, one starts with the non-parametric demand system relation and then makes an inference about its rank directly. The pioneering and most successful work in this direction has been accomplished by Donald [28]. Donald developed statistical tests to determine whether a $(J-1) \times 1$ vector $F(x, z)$, which is obtained from the $J \times 1$ vector $f(x, z)$ by eliminating one of its elements, can be factored as $F(x, z) = c(z) + A(z)H(x, z)$, where $c(z)$ is a $(J-1) \times 1$ vector, $A(z)$ is a $(J-1) \times L$ matrix and $H(x, z)$ is a $L \times 1$ vector. The rank of a demand system $y = f(x, z)$ is then determined by finding the smallest L for which the above factorization holds, and adding to it 1. (One drops a share of goods from a demand system in order to have a non-singular variance-covariance matrix of a random noise.)

In most of the statistical work thus far, it has been assumed that either prices are constant across consumers or that demographics are not important (or observed). Neither of these assumptions is realistic. For example, prices for those living in New York are higher than for those residing in Minneapolis. The focus of the thesis will be on extensions of the problems described above to the situation where variations in prices (or also demographic characteristics) are taken into account. More precisely, we will provide statistical tests to determine the local ranks in two types of demand systems, namely, demand systems given by the semi-parametric factor relation $Y_i = \theta(Z_i)V(X_i) + \epsilon_i$, $i = 1, \dots, N$, where $\theta(z)$ is a $J \times d$ unknown matrix and $V(x)$ is a $d \times 1$ known vector, and demands systems given by a non-parametric relation $Y_i = f(X_i, Z_i) + \epsilon_i$, $i = 1, \dots, N$, where $f(x, z)$ is a $J \times 1$ unknown vector. We will also explain the difficulties behind estimation of corresponding global ranks and outline some possible approaches to solution.

To avoid singularity of a variance-covariance matrix of noise variables ϵ_i (due to the summing up to 1 condition of the shares), we will drop one share of goods from our analysis and consider instead the corresponding reduced systems, namely, a semi-parametric factor (SPF) model $Y_i = \Theta(Z_i)V(X_i) + U_i$, $i = 1, \dots, N$, where $\Theta(z)$ is a $G \times d$ unknown matrix, $V(x)$ is a $d \times 1$ known vector and U_i are noise variables, and a non-parametric (NP) model $Y_i = F(X_i, Z_i) + U_i$, $i = 1, \dots, N$, where $F(x, z)$ is a $G \times 1$ unknown vector ($G = J - 1$ when these models are applied to demand systems). The noise variables U_i are now supposed to have a non-singular variance-covariance matrix. Local ranks for semi-parametric factor and non-parametric demand systems will be deduced from the local rank of (SPF) model and from, what we call, the adjusted rank of (NP) model. The methods to estimate local (adjusted) ranks in the two models will be referred to as local tests and those to estimate global (adjusted) ranks will be referred to as global tests.

The local rank of (SPF) model is, in fact, the rank of the matrix $\Theta(z)$. Since the matrix $\Theta(z)$ is unknown, we will first construct its estimator, which will be kernel based. We will then go over the known tests for the rank of a matrix, in particular, the LDU based test and the minimum-χ^2 test, applied to the matrix $\Theta(z)$ and its estimator $\widehat{\Theta}(z)$. One of the key assumptions in these tests is the asymptotic normality of the estimator $\widehat{\Theta}(z)$. We will show that $\widehat{\Theta}(z)$ is indeed asymptotically normal and, moreover, as part of our study, we will establish its consistency with corresponding convergence rates. In addition, we will draw some connections between the minimum-χ^2 test and eigenvalues of some random matrices. This will shed light on the minimum-χ^2 test statistic and will allow us to use different techniques in reestablishing its limit laws.

Local tests in (NP) model will be developed by following ideas of Donald mentioned above. Specifically, we will use the key observation that the (local) rank of (NP) model is L if and only if the matrix

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

where $\widetilde{F}(x, z) = F(x, z)E\beta(X_i, z) - EF(X_i, z)\beta(X_i, z)$ with some suitably chosen real-valued functions $\gamma(x, z) > 0$ and $\beta(x, z) \neq 0$, has $G - L$ zero eigenvalues. The local tests will then be based upon the asymptotics of the smallest eigenvalues of a kernel based estimator of $\Gamma_{w,z}$. To establish these asymptotic laws, we will use the so-called Fujikoshi expansions for eigenvalues along with techniques from the theory of U-statistics. Our results therefore extend the results of Donald to the case where coefficient matrices vary with covariates, so that the distinction between local and global tests becomes necessary.

Turning to global tests, we will introduce some global test statistics, explain the difficulties in establishing their limit laws and outline some possible approaches to solution. This will lay a path and point to directions for the future research.

The final part of the thesis will be devoted to applications of the introduced estimators in economic data and to their simulation study. In applications, we will use the Interview Survey Public Use Tapes of the Consumer Expenditure Surveys data (CEX data, in short) published by the Bureau of Labor Statistics in the United States and the Inner-City Price Indices data for the United States published by the American Chamber of Commerce Researchers Association (ACCRA data, in short). The latter dataset only now comes to the attention of econometricians working on demand systems. It was first used by Nicol

[76] to account for price variations in an observed data of demand systems.

The rest of the thesis is structured as follows. In Chapter 2, we give a short overview of demand systems, their ranks and related statistical work. In Chapter 3, we introduce semi-parametric factor (SPF) and non-parametric (NP) models, formulate related problems and draw connections to ranks of demand systems. In Chapter 4, we introduce some kernel based estimators that are used later, and establish some of their properties. Chapters 5 and 6 are on local and global (adjusted) ranks for semi-parametric factor (SPF) and non-parametric (NP) models. Applications and a simulation study of the introduced estimators can be found in Chapter 7.

Chapter 2

Demand Systems

As described in the introduction in Chapter 1, our motivation behind the statistical problems considered in this thesis lies in the theory of demand systems and their ranks. In this chapter, we will give a short overview of demand systems, their ranks and related statistical work. By doing so, we want to familiarize a more casual reader with this interesting milieu of Economic Theory and Statistics. We also want to show where our statistical models and problems fit in the earlier work on ranks of demand systems. The chapter is structured as follows. In Section 2.1, we define a demand system and describe its connections to the classical Economic Theory. Section 2.2 is on the rank of a demand system and Section 2.3 concerns its implications. Finally, in Section 2.4, we describe earlier statistical work related to ranks.

2.1 What is a demand system?

We begin by defining a demand system.

Definition 2.1.1 *(Demand system) A demand system of a collection of goods is a relation in which the amount or the quantity of each of the goods that a consumer is willing and able to purchase in a specified period of time, is determined as a function of the prices of all goods, the consumer's total expenditure and possibly other determinants such as prices of related goods, tastes, demographic variables (e.g. the age of a consumer, the number of children) and others.*

In other words, supposing that the number of goods in a demand system is J, denoting the prices of J goods by the column vector $p = (p_1, \ldots, p_J)'$, the quantity of J goods purchased by the column vector $q = (q_1, \ldots, q_J)'$, the total consumption expenditure (income or cost, in short) by x and other determinants by w, a demand system can be mathematically expressed as a functional relation

$$q = \begin{pmatrix} q_1 \\ \vdots \\ q_J \end{pmatrix} = \begin{pmatrix} g_1(x,p,w) \\ \vdots \\ g_J(x,p,w) \end{pmatrix} = g(x,p,w), \tag{2.1}$$

where g is some function. In the sequel, we will often denote the vector (p, w) by z and hence the demand system (2.1) by

$$q = g(x, z). \tag{2.2}$$

An equivalent way to write a demand system (2.1) is in a budget share form. A budget share y_j, $j = 1, \ldots, J$, for a good j is defined as

$$y_j = \frac{q_j p_j}{x}, \tag{2.3}$$

that is, as the expenditures on the jth good divided by the total expenditures (hence, the term "budget shares"). By using (2.1), the vector $y = (y_1, \ldots, y_J)'$ of budget shares can be expressed as

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_J \end{pmatrix} = \begin{pmatrix} f_1(x,p,w) \\ \vdots \\ f_J(x,p,w) \end{pmatrix} = f(x,p,w), \tag{2.4}$$

where $f_j(x,p,w) = g_j(x,p,w)p_j/x$, $j = 1, \ldots, J$, or, by denoting (x, w) by z, as $y = f(x, z)$. Conversely, any relation (2.4) can be expressed as in (2.1) with $g_j(x,p,w) = f_j(x,p,w)x/p_j$.

Relation (2.4) is called a representation of *a demand system in a budget share form.* In most of the cases, we will use and work with demand systems in a budget share form.

Remark 2.1.1 Note that the budget shares y_j, $j = 1, \dots, J$, sum up to 1. This observation, seemingly simple minded, has nontrivial implications in some statistical problems related to demand systems (in particular, the problem of estimation of rank studied in this thesis).

Remark 2.1.2 In applications, J goods in Definition 2.1.1 are taken as J groups of similar goods, for example, food, clothing, transportation, health and others. Such grouping of goods allows, for instance, to avoid problems associated with availability of data or infrequencies of purchases. An interesting introduction and survey on statistical issues arising in applications of consumer demand systems can be found in Lewbel [65]. See also Pollak and Wales [80]. Some of these issues are also discussed in this chapter below.

Demand systems arise and are extensively studied in the classical Economic Theory (more precisely, consumer demand theory). Let u be a utility function of a rational consumer. In Economic Theory, a market demand system (demand system, in short) is defined as the rule g that assigns the optimal consumption vector q in the utility maximization problem to each price-income situation (x, p), namely,

$$q = g(x,p) = \underset{q \geq 0}{\operatorname{argmax}}\ u(q) \quad \text{subject to} \quad pq \leq x. \tag{2.5}$$

(Market demands are also called integrable or Walrasian or Marshallian or ordinary demands.) If a demand system $q = g(x,p)$ satisfies the relation (2.5), one says that it is derived through the utility maximization problem (UMP, in short) or, simply, through the utility maximization.

Terminology. In other words, we distinguish between *any demand systems* as in Definition 2.1.1 and *demand systems derived through the utility maximization* as in relation (2.5). The latter demand systems are of special interest in Economic Theory. The need to introduce and work with any demand systems as in Definition 2.1.1 lies in statistical applications where there is no point, *a priori*, to assume a UMP derived structure on a demand system. In the sequel, by a "demand system" we will mean any demand system as in Definition 2.1.1 unless it is clear or mentioned in the text that we are in the context of Economic Theory and hence deal only with demand systems derived through UMP.

Example 2.1.1 Suppose that a consumer's preference ordering over market goods q_1 and non-market goods q_2 can be represented by a Cobb-Douglas utility function $u(q_1, q_2) = q_1^{\alpha} q_2^{1-\alpha}$ for some $\alpha \in (0,1)$. For this particular case the UMP can be written as

$$\underset{(q_1,q_2)>(0,0)}{\operatorname{argmax}} \quad q_1^{\alpha} q_2^{1-\alpha} \quad \text{subject to} \quad p_1 q_1 + p_2 q_2 = x. \tag{2.6}$$

Given variables $q = (q_1, q_2)$ and λ, we can define the Lagrangian function for (2.6) as

$$\mathcal{L}(q, \lambda) = q_1^{\alpha} q_2^{1-\alpha} - \lambda(p_1 q_1 + p_2 q_2 - x). \tag{2.7}$$

The Kühn-Tücker conditions for (2.7) are

$$\frac{\partial \mathcal{L}}{\partial q_1} = \alpha q_1^{\alpha-1} q_2^{1-\alpha} - \lambda p_1 = 0, \tag{2.8}$$

$$\frac{\partial \mathcal{L}}{\partial q_2} = (1-\alpha) q_1^{\alpha} q_2^{-\alpha} - \lambda p_2 = 0, \tag{2.9}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = p_1 q_1 + p_2 q_2 - x = 0, \tag{2.10}$$

for some $\lambda \geq 0$. Dividing condition (2.8) by (2.9) and using the budget constraint (2.10), we obtain the following Walrasian demand functions

$$g_1(x,p) = \frac{\alpha x}{p_1} \quad \text{and} \quad g_2(x,p) = \frac{(1-\alpha)x}{p_2}. \qquad \square \tag{2.11}$$

Another way of capturing the consumer's problem of choosing the most preferred consumption bundle given prices and income is through a problem dual to UMP, known in the economic literature as cost or expenditure minimization problem (CMP, in short). Instead of computing the maximal level of utility for given income and prices as in UMP, one first finds the minimum cost or expenditure, denoted by $c(u,p)$, necessary to attain a specific utility level u at given prices p, namely,

$$c(u,p) = \min_{q\geq 0} pq \quad \text{subject to} \quad u(q) \geq u. \tag{2.12}$$

(The coordinate functions of the vector q minimizing (2.12) are called Hicksian or compensated demand functions and are denoted by $h_j(u,p)$, $j = 1,\ldots,J$.) Since for a utility maximizing consumer $x = c(u,p)$, one can next invert this relation to obtain the so-called indirect utility function $u = v(x,p)$. The connection to the demand system (2.5) is then expressed through Roy's identity (Roy [91]) by using the indirect utility function as follows: for all $j = 1,\ldots,J$

$$q_j = g_j(x,p) = -\frac{\partial v(x,p)/\partial p_j}{\partial v(x,p)/\partial x}. \tag{2.13}$$

One may obviously get the demand system in a budget share form from (2.13) by using relation (2.3). An alternative way, available in the economic literature, is to start with the cost function $c(u,p)$ itself, take its logarithmic derivative with respect to prices $\partial \ln c(u,p)/\partial \ln p_j$ and then substitute for u the indirect utility function $u = v(x,p)$, that is,

$$y_j = f_j(x,p) = \left.\frac{\partial \ln c(u,p)}{\partial \ln p_j}\right|_{u=v(x,p)}. \tag{2.14}$$

A key result underlying the proof of (2.14) is the so-called Shephard's lemma (see Shephard [96]). For more information on demand systems derived through UMP and CMP, see for example Deaton and Muellbauer [24].

Example 2.1.2 The CMP applied to Example 2.1.1 can be stated as

$$\min_{(q_1,q_2)>(0,0)} p_1q_1 + p_2q_2 \quad \text{subject to} \quad q_1^{\alpha}q_2^{1-\alpha} = u. \tag{2.15}$$

The Lagrangian function and the Kühn-Tücker conditions corresponding to (2.15) are then

$$\mathcal{L}(q,\lambda) = p_1q_1 + p_2q_2 - \lambda(q_1^{\alpha}q_2^{1-\alpha} - u)$$

and

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial q_1} &= p_1 - \lambda\alpha q_1^{\alpha-1}q_2^{1-\alpha} = 0,\\
\frac{\partial \mathcal{L}}{\partial q_2} &= p_2 - \lambda(1-\alpha)q_1^{\alpha}q_2^{-\alpha} = 0,\\
\frac{\partial \mathcal{L}}{\partial \lambda} &= q_1^{\alpha}q_2^{1-\alpha} - u = 0,
\end{aligned}$$

respectively. Hence, the solution vector (q_1, q_2) for (2.15) are the Hicksian demand functions

$$h_1(u,p) = u\left(\frac{\alpha p_2}{(1-\alpha)p_1}\right)^{1-\alpha} \quad \text{and} \quad h_2(u,p) = u\left(\frac{(1-\alpha)p_1}{\alpha p_2}\right)^{\alpha}$$

and the corresponding cost function $c(u,p) = p_1h_1(u,p) + p_2h_2(u,p)$ is

$$c(u,p) = \alpha^{-\alpha}(1-\alpha)^{\alpha-1}up_1^{\alpha}p_2^{1-\alpha}.$$

The demand system in a budget share form can now be obtained by using (2.14), namely,

$$\begin{aligned}
y_1 &= \frac{\partial \ln c(u,p)}{\partial \ln p_1} = \alpha\\
y_2 &= \frac{\partial \ln c(u,p)}{\partial \ln p_2} = 1-\alpha.
\end{aligned}$$

(Observe that this result also follows from (2.11) by using relation (2.3).) In other words, the expenditure on each good is a constant fraction of income for any price vector p. A share of α goes for the market goods and a share of $(1-\alpha)$ goes for the non-market goods. □

Many of the studied demand systems are derived through the utility maximization as in Example 2.1.1. Well-known examples, some of which also appear below in this thesis, are the class of *demand systems of the Gorman polar form* (GPF)

$$q_j = a_j(p) + b_j(p)x, \tag{2.16}$$

where a_j and b_j are suitable functions of prices, the class of *quadratic expenditure demand systems* (QES) of the form

$$q_j = c_j(p)x^2 + d_j(p)x + e_j(p), \tag{2.17}$$

for some suitable functions c_j, d_j and e_j of prices, the class of *demand systems of the price independent generalized logarithmic* (PIGLOG) form

$$y_j = a_j(p) + b_j(p)\ln x, \tag{2.18}$$

the class of *the quadratic almost ideal* (QAID) *demand systems* of the form

$$y_j = c_j(p) + d_j(p)\ln x + e_j(p)(\ln x)^2 \tag{2.19}$$

and many others. (Observe that unlike (2.16) and (2.17), demand systems (2.18) and (2.19) are expressed in the budget share form.) Functional forms for demand systems, like those in (2.16)–(2.19), are interesting because they can be used as models in statistical applications. They can be also used, for example, to test whether consumers choose demands to maximize a utility function (see Pollak and Wales [79] and Banks, Blundell and Lewbel [10]).

Remark 2.1.3 It is important to note that the functions of prices in (2.16)–(2.19) have special structures because the corresponding demand systems are derived through UMP. For example, one can show that the functions a_j and b_j in (2.18) are given by

$$a_j(p) = \frac{\partial\alpha(p)}{\partial p_j}\frac{\ln\beta(p)}{\alpha(p)} + \frac{\partial\beta(p)}{\partial p_j}\frac{1}{\beta(p)}, \quad b_j(p) = -\frac{\partial\alpha(p)}{\partial p_j}\frac{1}{\alpha(p)},$$

where $\alpha(p)$ and $\beta(p)$ are homogeneous functions of orders 0 and 1, respectively. On the other hand, from applications perspective, demand systems (2.16)–(2.19) are interesting even when corresponding functions of prices are arbitrary. For this reason, we will sometimes continue to refer to relations (2.16), (2.17), (2.18) and (2.19) as GPF, QES, PIGLOG and QAID demand systems, respectively, even when the functions of prices in these demand equations are arbitrary.

Finally, let us mention that most of the demand systems in Economic Theory as well as in applications are analyzed for a fixed level of prices p, in which case they are known as *Engel curves.* An example of Engel curve, which by now is already classical, was introduced by Working [102] and developed by Leser [58] who postulated a linear relation between the budget share of each good and the logarithm of total expenditure, that is,

$$y_j = a_j + b_j \ln x.$$

(It is a PIGLOG demand system for fixed prices p.) This relation, known also as the Working-Leser specification, was widely used in the work by Deaton, Muellbauer and others. Relations (2.16), (2.17) and (2.19), when p is fixed, are other examples of Engel curves.

Remark 2.1.4 To avoid confusion, since a demand system can be expressed in the form of budget shares as in (2.4) or in the form of quantity of goods demanded as in (2.1), one always needs to keep in mind which of the two forms is meant when speaking about Engel curves. For example, the Engel curves of a demand system $y_j = a_j$ (with some constants a_j) in the budget share form are functions of a constant level while the Engel curves of the same demand equation but in the form (2.1) are lines passing through the origin and having possibly different slopes.

2.2 Rank of a demand system

In this thesis, we focus on a characteristic of demand systems called a rank of a demand system. We will consider two types of ranks, namely, local and global ranks of demand systems. These ranks, generalized by Lewbel [63] from Gorman's [41] work, are defined as follows. We work with demand systems in a budget share form $y = f(x, z)$ as given by (2.4).

Definition 2.2.1 *(Rank of demand system) Let z be fixed. The (local at z) rank of any demand system*

$$y = f(x, z) = \begin{pmatrix} f_1(x, z) \\ \vdots \\ f_J(x, z) \end{pmatrix} \tag{2.20}$$

is the dimension of the function space spanned, for fixed z, by the coordinate functions $f_j(x, z)$, $j = 1, \ldots, J$, that is, by the Engel curves of the demand system (2.20). The (global) rank of a demand system (2.20) is the supremum of its local ranks over all possible values of z.

Equivalently, a demand system has local rank $R(z)$ if there are $R(z)$, but not less, goods such that the Engel curve of any good can be expressed as a linear combination of the Engel curves of those $R(z)$ goods. For example, a demand system with the Working-Leser specification of Engel curves $y_j = a_j + b_j \ln x$ has rank 2 unless, by using the fact

$\sum_j y_j = 1$, we have $y_j \equiv a_j$.

Notation. We will denote the local at z rank of a demand system (2.20) by

$$\mathrm{rk}\{f(\cdot,z)\}. \tag{2.21}$$

Let $R(z) = \mathrm{rk}\{f(\cdot,z)\}$ be the local rank of a demand system and $R = \sup_z R(z)$ be its global rank. Observe that, by Definition 2.2.1 and a discussion following it, $R(z)$ and R are positive integers between 1 and J, and that $R(z) \le R$. Observe also that, by Definition 2.2.1, for fixed z,

$$y_j = f_j(x,z) = \sum_{k=1}^{R(z)} a_{jk}(z)h_k(x,z), \quad j = 1,\dots,J, \tag{2.22}$$

where $h_k(x,z)$ are some functions and a_{jk} are some weights, or in a matrix form,

$$y = f(x,z) = \sum_{k=1}^{R(z)} a_k(z)h_k(x,z) = a(z)h(x,z), \tag{2.23}$$

where

$$a_k(z) = \begin{pmatrix} a_{1k}(z) \\ \vdots \\ a_{Jk}(z) \end{pmatrix}, \quad h(x,z) = \begin{pmatrix} h_1(x,z) \\ \vdots \\ h_{R(z)}(x,z) \end{pmatrix},$$

and $a(z) = (a_1(z),\dots,a_{R(z)}(z))$. By Definition 2.2.1, relation (2.22) and (2.23) still hold when $R(z)$ is replaced by R and where the matrix a and the vector h are defined in a similar way as for (2.22) and (2.23). Conversely, a demand system $y = f(x,z)$ can be always written in the form (2.22) or (2.23) with $L \le J$ in the upper limit of the corresponding sums. The local rank of $y = f(x,z)$ is then the smallest such L for which (2.22) or (2.23) hold for fixed z. The global rank is the smallest such L for which these equations hold for all values of z. Let us state these facts as a lemma for later reference.

Lemma 2.2.1 *Let $f(x,z)$ be a $J \times 1$ vector of functions of x and z. Then, the local at z rank $\mathrm{rk}\{f(\cdot,z)\}$ is the smallest L such that, for fixed z,*

$$f(x,z) = a(z)h(x,z), \tag{2.24}$$

where $a(z)$ is a $J \times L$ matrix of functions of z and $h(x,z)$ is a $L \times 1$ vector of functions of x and z. The global rank $\sup_z \mathrm{rk}\{f(\cdot,z)\}$ is the smallest L such that (2.24) holds for all z and x.

Remark 2.2.1 Suppose that the rank of a demand system (2.20) is $R < J$. Then, the discussion above implies that the demand system can be expressed in terms of R functions of prices and total income as compared to J such prices in the original representation (2.20). Hence, the rank lower than J indicates that there is redundancy in the representation of a demand system and, as a consequence, that the demand system can be expressed by using fewer "parameters". Other implications of rank are discussed in Section 2.3 below.

While Definition 2.2.1 of ranks is stated in terms of the demand system function $f(x,z)$, the rank can be also characterized in terms of the cost function $c(u,z)$ in (2.12) when the demand system is derived through UMP. The proof of this characterization result can be found in Lewbel [63].

Theorem 2.2.1 *An integrable demand system has (global) rank R if and only if R is the smallest integer such that the cost function $c(u,z)$ is of the form*

$$c(u,z) = B(u, \phi_1(z), \ldots, \phi_R(z)), \tag{2.25}$$

where u is a utility level, B is some function and $\phi_1(z), \ldots, \phi_R(z)$ are some homogeneous functions.

Theorem 2.2.1 is useful in Economic Theory because cost functions are often used to derive a demand system equation. In other words, the theorem shows that for a demand system to be of rank R, it has to be derived from a cost function expressed in terms of R, but not less, so-called price indices $\phi_1(z), \dots, \phi_R(z)$. Let us also mention that a result similar to Theorem 2.2.1 holds also when the cost function $c(u, z)$ and the utility level u in (2.25) are replaced by the indirect utility function $v(x, z)$ and the income x, respectively.

We now illustrate the notion of a rank through an example.

Example 2.2.1 Consider the PIGLOG demand system in (2.18) defined by

$$y_j = a_j(p) + b_j(p) \ln x, \quad j = 1, \dots, J.$$

We can write it in a matrix notation as

$$y = \begin{pmatrix} a_1(p) & b_1(p) \\ \vdots & \vdots \\ a_J(p) & b_J(p) \end{pmatrix} \begin{pmatrix} 1 \\ \ln x \end{pmatrix} =: \theta(p)V(x), \tag{2.26}$$

where the matrix $\theta(p)$ is $J \times 2$ and the vector $V(x)$ is 2×1. By Definition 2.2.1 and the discussion following it, the local rank of a PIGLOG demand system is the rank of the matrix $\theta(p)$ for fixed p. Hence, the rank of the demand system is either 1 or 2. If one assumes in addition that the PIGLOG demand system is derived through UMP, then one can show that the rank is necessarily 2. One way to argue this is to observe that the PIGLOG demand system can be derived from the cost function $c(u, p) = e^{u/\alpha(p)}\beta(p)$ which involves two functions of prices, hence, by Theorem 2.2.1, leading to a demand system of rank 2. □

2.3 Implications of rank

Although definition of rank may seem easy from a mathematical point of view, it has many interesting and important implications in Economic Theory and applications. In this section, we will describe some of these implications in greater detail.

One implication concerns restrictions on the ranks of particular classes of demand systems derived through UMP. Consider, for example, the class of the so-called exactly aggregable demand systems defined as follows.

Definition 2.3.1 *(Exactly aggregable demand systems) A demand system is called exactly aggregable if it has the representation*

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_J \end{pmatrix} = \sum_{k=1}^{d} \begin{pmatrix} \theta_{1k}(z) \\ \vdots \\ \theta_{Jk}(z) \end{pmatrix} V_k(x) = \theta(z)V(x), \tag{2.27}$$

where $\theta(z) = (\theta_{jk}(z))$ is a $J \times d$ matrix of functions of prices (and demographic variables), and $V(x) = (V_k(x))$ is a $d \times 1$ vector of functions of total income.

As shown in Gorman [41], if an exactly aggregable demand system is derived through UMP, then its rank is necessarily less than or equal to 3. (Russell and Farris [92] have a curious mathematical connection of this result to Lie groups.) Gorman's result is interesting, for example, for the following reason. Since exactly aggregable demand systems have nice theoretical properties related to aggregation and representative consumer, and nest many well-known examples of demand systems, they are widely used in applications. Therefore, it is interesting to see whether the properties of demand systems observed in practice are consistent with those derived in Economic Theory. In other words, if an exactly aggregable demand system model is used in applications and if it is found to be of rank greater than 3, then one may question whether consumers indeed maximize a utility function as is assumed in Economic Theory. (This issue is further discussed in Section 2.4 below.)

Remark 2.3.1 Exactly aggregable demand systems (2.27) are also of special interest in the context of this thesis. As can be seen from Section 3.1 below, one of the statistical models used in this thesis, in fact, coincides with the class of exactly aggregable demand systems.

Remark 2.3.2 Another example of a class of demand systems with restrictions on their ranks is the class of deflated income demand system introduced by Lewbel [61]. *Deflated income demand systems* are defined by

$$y = c(z)h\left(\frac{x}{t(z)}\right) \tag{2.28}$$

where $c(z)$ is a $J \times d$ matrix of functions of prices and demographic variables, $t(z)$ is a real-valued function and $h(x/t(z))$ is a $d \times 1$ vector of functions of the deflated income $x/t(z)$. Lewbel [61] proved that deflated income demand systems can be written as transformations of exactly aggregable systems. As a consequence, Lewbel was able to conclude that deflated income demand systems are necessarily of rank less or equal than 4.

Another implication is that the rank provides a convenient characteristic of demand systems in Economic Theory according to which demand systems can be classified, discussed and analyzed further. For example, a demand system has *rank 1* if and only if its Engel curves for any fixed price regime are expressed as

$$y_j = a_j h_1(z),$$

where $h_1(z)$ is some real-valued function which does not depend on j. Since the budget shares sum to one for all z, the function $h_1(z)$ is necessarily independent of z. Then, a demand system has rank 1 if and only if its Engel curves are constant functions. Another way to express this fact is to say that the demand functions are derived from so-called homothetic preferences, that is, consumers with different incomes facing the same prices will demand goods in the same proportions.

Similarly, a demand system (when derived through UMP) can be shown to be of *rank 2* if and only if it is generalized linear (GL, in short) as introduced by Muellbauer [74, 75].

A demand system is GL if it can be expressed as

$$y_j = t(x,z)a_j(z) + b_j(z), \tag{2.29}$$

where $\sum_j a_j(z) = 0$, $\sum_j b_j(z) = 1$ and t is a suitable function. Examples of GL demand systems (2.29) are the GPF demand system in (2.16), the PIGLOG demand system in (2.18) and also, not previously mentioned, so-called almost ideal (AIDS), translog, quasi-homothetic, linear expenditures and fractional demand systems. GL systems are also of interest for other reasons than being the only demand systems of rank 2. Muellbauer [74] showed that the GL is a necessary and sufficient condition for aggregate demands to be consistent with representative agent models. Freixas and Mas-Collel [35] showed that GL (defined in their paper as no torsion condition) is a necessary condition for aggregate demands to exhibit the weak axiom of revealed preferences (WARP).

An example of a demand system of *rank 3* is a demand system quadratic in total expenditures (QES). QES demand systems were introduced by Pollak and Wales [78] and Howe, Pollak and Wales [52] in a first attempt, to account for quadratic expenditures effects in demand systems. Another example is the quadratic logarithmic demand system (QUAIDS) of Banks, Blundell and Lewbel [10] which extends AIDS demand system (see Deaton and Muellbauer [24]) by adding to it a quadratic logarithmic term of income and prices. QUAIDS allows goods to move from luxuries to necessities at different levels of income, providing the demand system with more flexibility to describe consumer's behavior. Yet another example of demand systems of rank 3 is the almost ideal quadratic logarithmic demand systems (AIQL) of Fry and Pashardes [36]. Finally, an example of a demand system of *rank 4* is the nearly log polynomial (NLP) demand system proposed by Lewbel [67]. NLP is consistent with the utility maximization and nests the popular log linear and log quadratic specifications. These systems are particularly suitable for demand systems with a large number of goods.

Some further implications of rank, for example, in connection to separability of demand systems, production models and welfare analysis, are discussed in Lewbel [63] (see also references therein).

Remark 2.3.3 Interestingly, the notion of rank as defined in Definition 2.2.1, turns out to be also relevant for portfolio separation in asset demands. Cass and Stiglitz [15] define the generalized portfolio separation of order R as the property that, for any value of the agent's total initial wealth, the agent's demand for J securities can be satisfied by the purchase of $R < J$ mutual funds (R portfolios), where a mutual fund is defined as a fixed weight basket of securities. Lewbel and Perraudin [68] showed that the separation property can be viewed as a rank restriction on the space of demand for risky assets considered as functions of the available rates of return for agents with the same preferences but different wealth. In other words, the portfolio fund separation of order R is equivalent to the fact that the demand system

$$y = f(x, z)$$

for risky assets, where y is a vector of expenditures for J assets of an agent, x is the total wealth of an agent and z is a vector of gross returns on one dollar's worth of J assets, has rank R.

2.4 Statistical and empirical studies of ranks

In statistical applications, a demand system model is no longer deterministic as in (2.4) but it is now stochastic. The basic assumptions are that a sample (or independent observations) (Y_i, X_i, Z_i), $i = 1, \ldots, N$, of the budget shares Y_i, the total incomes X_i and, possibly, the prices and demographic variables Z_i of N consumers is available, and that these observations are related through some functions f as

$$Y_i = f(X_i, Z_i) + \epsilon_i, \quad i = 1, \ldots, N, \tag{2.30}$$

where ϵ_i are some noise variables. (Hereafter, in order not to confuse deterministic and stochastic variables, we denote the former in lowercase and the latter in capitals.)

The function f in (2.30) can have a parametric form, for example, ignoring the variable z, the form $f(x) = (f_1(x), \dots, f_J(x))'$ with

$$f_j(x) = a_j + b_j \ln x,$$

where a_j and b_j are unknown parameters. Alternatively, one could let $f_j(x)$ have an unrestricted form. In the former case, one says that a demand system model (2.30) is *parametric* and in the the latter case, one says that it is *non-parametric*. (The term *semi-parametric* model refers to a model with some features of both parametric and non-parametric models, for example, our (SPF) model introduced in Section 3.1 which involves a particular structural form of f and also some unknown functions.)

The sample (Y_i, X_i, Z_i) is taken from one of the expenditure surveys conducted by government or private agencies. The most popular and widely used surveys are the Interview Survey Public-Use Tapes of the Consumer Expenditure Surveys (CEX, in short) published by the Bureau of Labor Statistics in the United States and Family Expenditure Surveys (FES, in short) conducted by Department of Employment in the United Kingdom. For a description of FES data, see for example Hildenbrand [51] (see also Remark 2.1.2). Other surveys available to researchers cover, for example, countries like France, Canada and New Zealand. It is important to note that none of these surveys include the price data (they may include, however, the data concerning demographic variables). We will discuss the price data at the end of this section.

There is an extensive literature on what models of demand system (2.30) are best to use for data sets at hand. A general belief is that, for a fixed level of prices and for homogeneous households, the Working-Leser specification $y_j = a_j + b_j \ln x$ gives a good fit for certain goods, in particular for the budget share of food and fuel. Other goods, such as clothing or alcohol, are believed to follow Engel curves of the form $y_j = a_j + b_j \ln x + c_j \phi(x)$ for some function ϕ of total income (e.g. $\phi(x) = (\ln x)^2$ or $1/x$). (The reader interested specifically at the problem of model specification and fitting may refer to some papers given below.)

Our focus here is on estimation of local or global ranks of a demand system (2.30) given a sample of observations (Y_i, X_i, Z_i). The local rank of (2.30) is defined as the local rank of the corresponding deterministic demand system $y = f(x, z)$, that is, as $\mathrm{rk}\{f(\cdot, z)\}$ for fixed z with the notation of Section 2.2. The global rank of (2.30) is defined again as the maximum of all local ranks, that is, $\sup_z \mathrm{rk}\{f(\cdot, z)\}$. We will now briefly describe the earlier statistical work related to rank estimation in parametric and non-parametric models of demand systems. This will then bring us naturally to a motivation behind the problems considered in this thesis.

Parametric demand systems. The author is aware only of a few studies aimed at a direct estimation of rank in a demand system (2.30) of a parametric form. (This is partly due, we feel, to the fact that rank estimation techniques became available only recently.) Most of the studies in a parametric setting are only related to rank estimation. For example, the papers by Leser [58] and Pollak and Wales [79] can be considered as the first rank studies. Leser [58] found that the non-linear Engel curve specification

$$y_j = a_j + b_j \ln x + \frac{c_j}{x} \tag{2.31}$$

is superior to the classical Leser-Working specification $y_j = a_j + b_j \ln x$. This finding indicates a rank possibly greater than 2 in a demand system (2.31) (in other words, $c_j \neq 0$). In the same vein, Hausman, Newey and Powell [50] used a parametric model

$$Y_{ji} = a_j + b_j \ln X_i + c_j (\ln X_i)^2 + d_j (\ln X_i)^3 + \epsilon_{ji},$$

where $j = 1, \ldots, J$ correspond to the budget shares and $i = 1, \ldots, N$ correspond to the elements of the sample, to fit CEX data. By using either instrumental variables or ordinary least squares, the authors found that the ratios c_j/d_j are statistically the same for different j's which is an indication of rank less or equal than 3. (In particular, this result is consistent with that of Gorman [41] discussed in Section 2.2.) Some other papers concerning estimation of rank in parametric models are Grodal and Hildenbrand [45], Kneip [55] and Nicol [76]. Since the paper by Nicol [76] is particularly relevant to this thesis, we will come back and discuss it in greater detail below.

Non-parametric demand systems. In a non-parametric setting, the (local) rank of a demand system was first estimated by Lewbel [63]. For *a fixed price regime* (in which case we can suppress the dependence on z), if a demand system (2.30) has rank R, then by Lemma 2.2.1,

$$Y_i = ah(X_i) + \epsilon_i,$$

where a is a $G \times R$ matrix and $h(x)$ is a $R \times 1$ vector of functions of total income. Suppose that $Q(x)$ is a $T \times 1$ (with $T \geq R$) vector of some rich enough family of functions of total income, for example, $Q(x) = (1, x, \ln x, x^2, 1/x, x \ln x)$. Then, the $R \times T$ matrix

$$M = EY_iQ(X_i)' = a\big(Eh(X_i)Q(X_i)'\big), \tag{2.32}$$

which always has rank equal or less than R, is likely to be of rank R and hence, R can be deduced from the rank of the matrix M. One can estimate the rank of M by using its estimator

$$\widehat{M} = \frac{1}{N}\sum_{i=1}^{N} Y_iQ(X_i)', \tag{2.33}$$

along with some method of rank estimation available in the literature, for example, the LDU test of Gill and Lewbel [38] (with the correction in Cragg and Donald [19]) or the minimum-χ^2 test of Cragg and Donald [20]. We will discuss these and other rank tests in Chapter 5 in connection to one of the problems studied in this thesis. Lewbel [63] applied the procedure described above to CEX and FES data, and found a strong evidence of rank 3 in both data sets. Interestingly, this finding is consistent with Gorman's result on rank 3 for exactly aggregable demand systems (see Section 2.3).

A rank test for a non-parametric model, which is more accurate and more powerful than that of Lewbel [63] described above, is due to Donald [28]. Suppose *a fixed price regime* as in Lewbel [63]. The basic idea of Donald [28] is as follows. First, drop one budget share of goods from the analysis and consider the reduced demand system

$$Y_i = F(X_i) + U_i, \quad i = 1, \dots, N, \tag{2.34}$$

where $F(x)$ is a $G \times 1$ vector (with $G = J - 1$) and U_i are noise variables. (We use the same notation Y_i for reduced collection of budget shares for convenience.) Unlike in relation (2.30), where budget shares in Y_i add up to 1 and hence impose singularity restrictions on the variance-covariance matrix of ϵ_i, one may now suppose that the variance-covariance matrix Σ of U_i is non-singular. This is one of the assumptions made by Donald [28]. Second, one may show that the rank $\mathrm{rk}\{f(\cdot)\}$ of the original demand system $y = f(x)$ is $L + 1$, where L is the smallest integer such that

$$F(x) = c + AH(x) \tag{2.35}$$

for some $G \times 1$ vector c, $G \times L$ matrix A and $L \times 1$ vector $H(x)$. Moreover, the condition (2.35) can be seen to be equivalent to the fact that the $G \times G$ matrix

$$\Gamma_w = Ew(X_i)(F(X_i) - EF(X_i))(F(X_i) - EF(X_i))', \tag{2.36}$$

where $w(x) > 0$ is a suitably chosen weight function, has $G - L$ zero eigenvalues. This suggests that, the rank $\mathrm{rk}\{f(\cdot)\}$ can be deduced from the eigenvalues of some estimator $\widehat{\Gamma}_w$ of the matrix Γ_w. Donald [28] considered two types of estimators $\widehat{\Gamma}_w$, a series based estimator and a kernel based estimator. For example, the kernel based estimator is defined as

$$\widehat{\Gamma}_w = \frac{1}{N(N-1)} \sum_{i \neq j}^{N} (Y_i - \overline{Y})(Y_j - \overline{Y})' K_h(X_i - X_j), \tag{2.37}$$

where $\overline{Y} = N^{-1} \sum_{i=1}^{N} Y_i$ is the sample mean, $K_h(\cdot) = h^{-m} K(h^{-1} \cdot)$ is a scaled kernel function and $h > 0$ is a bandwidth. By using the so-called Fujikoshi expansion techniques for eigenvalues of random matrices, Donald [28] established asymptotic laws for eigenvalues of $\widehat{\Gamma}_w \widehat{\Sigma}^{-1}$, where $\widehat{\Sigma}$, an estimator for the variance-covariance matrix Σ, plays the role of a normalization, and then developed statistical tests to determine the rank $\mathrm{rk}\{f(\cdot)\}$. When applying these tests in practice, Donald found a strong evidence of rank 3 in the CEX data. Finally, in the case of non-parametric model, let us also mention the papers by Lyssiotou, Pashardes and Stengos [71, 72]. These authors extend the work of Donald [28] by semi-parametrically controlling for variations in demographic variables.

Remark 2.4.1 Since demographic variables have been found to be important determinants in both theoretical and empirical work on demand systems and their ranks, they deserve some further discussion. We focus here on demographic variables and ranks of demand systems. (Other issues arising in connection to demographic variables, for example, how to incorporate them into theoretically reasonable demand systems, are discussed in Pollak and Wales [80] and Lewbel [65].) Many empirical studies in demand systems and their ranks, for example, the above mentioned Lewbel [63], Banks et al. [10], Donald [28] and Hausman et al. [50], were based on demographically homogeneous data sets and hence the problem of controlling for demographic variations in preferences did not arise. The pre-selection, however, was done at the expense of a sample size and also did not allow to evaluate welfare analysis in households of minority groups. These problems raised a need to investigate the effects of household heterogeneity. Such studies were undertaken, for example, in Lyssiotou et al. [71, 72] and Nicol [76]. Lyssiotou et al. [71, 72] considered the model $Y_i = \alpha Z_i + F(X_i) + U_i$, where Z_i is a vector of demographic characteristics taken as dummy variables, possibly correlated with total income. By using a nearest neighbor method proposed by Estes and Honoré [30] and Yatchew [103], they obtained an estimator $\widehat{\alpha}$ for α and hence were able to remove heterogeneity from Y_i by subtracting from it $\widehat{\alpha} Z_i$. After semi-parametrically controlling for heterogeneity, Lyssiotou et al. [71, 72] applied the Donald's [28] rank test to FES data. They found that taking into account for preference heterogeneity gives support for rank 3 of a demand system and that the exclusion of heterogeneity would make the rank estimator biased upwards. In another work, by analyzing CEX data, Nicol [76] found that age of head, labour force participation, vehicle ownership and tobacco consumption are important determinants of demand systems. Finally, let us mention some studies, e.g. Grodal and Hildenbrand [45] and Kneip [55], that use large samples of data but fail to control for heterogeneity. Their findings, not surprising in view of the work by Lyssiotou et al. [71, 72], point to higher ranks (typically between 4 and 6).

In all of the work on ranks thus far (with an exception of Nicol [76] discussed below), it is assumed that consumers face identical prices. This assumption is not realistic. For example, CEX data used in applications covers households across all the United States where prices of goods are clearly heterogeneous. In the same vein, most of the work on rank either concerns households with homogeneous demographics or ignores them altogether (see Remark 2.4.1 above). In the former case, a pre-selection is done, for example, at the expense of the sample size while, in the latter case, the assumption of homogeneous demographic is simply not realistic. Our goal in this thesis is then to extend some problems on rank estimation discussed above to the situations where variations in prices (or also demographic variables) are taken into account. As mentioned in the introduction (see Chapter 1), we will provide statistical tests to determine the local ranks in two types of demand systems given by the semi-parametric factor relation $Y_i = \theta(Z_i)V(X_i) + \epsilon_i$, where $\theta(z)$ is an unknown matrix and $V(x)$ is a known vector, and demand systems given by the non-parametric relation $Y_i = f(X_i, Z_i) + \epsilon_i$, where $f(x, z)$ is an unknown vector. In order to do so we will consider two statistical models, namely a semi-parametric factor (SPF) model and a non-parametric (NP) model, and study problems related to rank estimation in the two types of demand systems above. The (SPF) and (NP) models, related problems and connections to ranks of demand systems are discussed in the next chapter.

Remark 2.4.2 The work closely related to ours, is a recent paper by Nicol [76]. Nicol [76] proposed and tested a parametric model of demand systems which takes into account variations in prices. To our knowledge, this paper is the first attempt to deal with price changes across consumers. It is relevant to this thesis for a number of reasons. First, by using the American Chamber of Commerce Research Association (ACCRA) data, Nicol [76] created a unique price data set to go with the CEX data. Such data sets were not available or used before, partly owing to the problem that inter-regional price data is not published by government agencies. The ACCRA data mentioned above comes from private sources and contains indices for a range of goods and services (namely, housing, utilities, grocery items, transportation, health care and miscellaneous goods) in about 300 cities across the United States. See Table A1 on p. 288 in Nicol [76] or Table 7.1 in Section 7.1 below to get a feeling for the price data provided by ACCRA. It is somewhat striking to see to what extent the prices of goods are different in various locations across the United States. (For use of ACCRA data in other applications, see for example Frankel and Gould [34].) Second, Nicol [76] found evidence that the presence of price variations affected the rank of demands. Test results ignoring these variations indicated demands of rank 3 whereas those taking price variations into account supported the rank 2 hypothesis.

Chapter 3

Two Models

In this chapter, we introduce two models and discuss related problems which are studied in this thesis. We also state some assumptions on these models which will be used in the following chapters. The two models are a semi-parametric factor model considered in Section 3.1 and a non-parametric model considered in Section 3.2. Their connection to the theory of demand systems, being nontrivial, is explained in Section 3.3.

3.1 Semi-parametric factor model

Let $(X_i, Z_i) \in \mathbb{R}^n \times \mathbb{R}^m$ be independent variables and $Y_i \in \mathbb{R}^G$ be response variables explained by (X_i, Z_i). In a *semi-parametric factor model*, we will suppose that the relation between the variables Y_i and (X_i, Z_i) is given by

$$Y_i = \Theta(Z_i)V(X_i) + U_i, \quad i = 1, \dots, N, \tag{SPF}$$

where N is the number of observations, $\Theta(z)$ is an *unknown* $G \times d$ matrix of functions of z, $V(x)$ is a *known* $d \times 1$ vector of functions of x and U_i is a $G \times 1$ noise vector. Further assumptions on the variables X_i, Z_i and U_i, and on the functions Θ and V are stated below.

Example 3.1.1 By taking $n = 1$, fixed $d \geq 1$ and a vector $V(x) = (1, \ln x, \dots, (\ln x)^{d-1})$, a (SPF) model becomes

$$Y_{ij} = \theta_{j1}(Z_i) + \theta_{j2}(Z_i) \ln X_i + \cdots + \theta_{jd}(Z_i)(\ln X_i)^{d-1} + U_{ij}, \quad j = 1, \dots, G,$$

where $Y_i = (Y_{i1}, \ldots, Y_{iG})'$, $U_i = (U_{i1}, \ldots, U_{iG})'$ and $\Theta(z) = (\theta_{jk}(z))$. Such models are of interest in the theory of demand systems.

We will consider two types of problems related to (SPF) model, which we will refer to as *local* and *global* rank tests. Let $\mathrm{rk}\{\Theta(z)\}$ denote the rank of the matrix $\Theta(z)$.

Local tests for SPF model. *For a fixed* $1 \le L \le \min\{G, d\}$ *and* z*, to test the hypothesis* $H_0 : \mathrm{rk}\{\Theta(z)\} \le L$*, against the alternative* $H_1 : \mathrm{rk}\{\Theta(z)\} > L$*, as well as to determine* $\mathrm{rk}\{\Theta(z)\}$.

Global tests for SPF model. *For a fixed* $1 \le L \le \min\{G, d\}$*, to test the hypothesis* $H_0 : \sup_z \mathrm{rk}\{\Theta(z)\} \le L$*, against the alternative* $H_1 : \sup_z \mathrm{rk}\{\Theta(z)\} > L$*, as well as to determine the global rank* $\sup_z \mathrm{rk}\{\Theta(z)\}$.

The motivation behind these tests, explained in Section 3.3 below, lies in the theory of demand systems and their ranks. Local tests for (SPF) model are studied in Chapter 5 and global tests for (SPF) model are discussed in Chapter 6.

We now state assumptions on the variables X_i, Z_i and U_i, and on the functions Θ and V used in local tests. We will assume some of the following.

Assumption (SPF) L1: *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* $\mathcal{H}_x = [a_1, b_1] \times \cdots \times [a_n, b_n]$ *and* $\mathcal{H}_z = [c_1, d_1] \times \cdots \times [c_m, d_m]$*, and* (X_i, Z_i) *are continuously distributed with a density* $p(x, z)$ *which has an extension to* $\mathbb{R}^n \times \mathbb{R}^m$ *with* $s \ge r$ *continuous bounded derivatives. (The parameter* r *is defined in Assumption* (SPF) L5 *of Section 4.1.)*

Assumption (SPF) L2: *Suppose that the noise vectors* U_i*,* $i = 1, \ldots, N$*, are i.i.d.*

random vectors, independent of the sequence (X_i, Z_i) and such that $E(U_i|X_i, Z_i) = 0$ and

$$E(U_i U_j' | X_i, X_j, Z_i, Z_j) = \begin{cases} 0 & \text{if } i \neq j, \\ \Sigma & \text{if } i = j, \end{cases} \tag{3.1}$$

where Σ is a $G \times G$ positive definite matrix. Suppose also that $E|U_i|^4 < \infty$.

ASSUMPTION (SPF) L3: *The function $\Theta : \mathcal{H}_z \to \mathbb{R}^{Gd}$ is such that each of its component functions has an extension to $\mathbb{R}^m$ with $s \geq r$ continuous bounded derivatives. Each of the functions in a vector $V(x)$ has an extension to $\mathbb{R}^n$ which is bounded.*

ASSUMPTION (SPF) L4: *The $d \times d$ matrix*

$$\begin{aligned} Q(z) &= p(z)E(V(X_1)V(X_1)'|Z_1 = z) \\ &= \int_{\mathbb{R}^n} V(x_1)V(x_1)'p(x_1, z)dx_1, \end{aligned} \tag{3.2}$$

where $p(z)$ and $p(x, z)$ are the densities of Z_1 and (X_1, Z_1), respectively, is positive definite (invertible).

Remark 3.1.1 Assumptions (SPF) L1–L3 are in the spirit of those used by Donald [28] in connection to non-parametric rank estimation.

Remark 3.1.2 We supposed in Assumption (SPF) L2 that the variance-covariance matrix of U_i is constant for different values of X_i and Z_i. A more general assumption is

$$E(U_i U_j' | X_i, X_j, Z_i, Z_j) = \begin{cases} 0 & \text{if } i \neq j, \\ \Sigma(X_i, Z_i) & \text{if } i = j, \end{cases} \tag{3.3}$$

where $\Sigma(\cdot, \cdot)$ is some function. The noise satisfying (3.3) is called *heteroskedastic* and that satisfying (3.1) is called *homoskedastic*. In fact, some of our results for a (SPF) model with a homoskedastic noise, when slightly modified, will hold for a (SPF) model with a heteroskedastic noise as well. We will indicate in the sequel where this is so.

3.2 Non-parametric model

As in the (SPF) model in Section 3.1, consider independent variables (X_i, Z_i) and response variables Y_i. In a *non-parametric model*, we will suppose that the relation between the variables Y_i and (X_i, Z_i) is given by

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

where N is the number of observations, $F(x, z) = (F_1(x, z), \dots, F_G(x, z))'$ is an *unknown* $G \times 1$ vector of functions of x and z, and U_i is a $G \times 1$ noise vector. Further assumptions on the variables X_i, Z_i and U_i, and the function F can be found below.

To state the problems related to (NP) model, we need the following definition.

Definition 3.2.1 *(Adjusted rank) The adjusted rank of a $G \times 1$ vector $F(x, z)$, denoted by*

$$\operatorname{adrk}\{F(\cdot, z)\}, \tag{3.4}$$

is the smallest integer $L \in \mathbb{N} \cup \{0\}$ such that, for a $G \times 1$ vector $c(z)$, a $G \times L$ matrix $A(z)$ and a $L \times 1$ vector $H(x, z)$,

$$F(x, z) = c(z) + A(z)H(x, z). \tag{3.5}$$

Observe from Lemma 2.2.1 that without the additive term $c(z)$ in (3.5), Definition 3.2.1 is that of the rank $\operatorname{rk}\{F(\cdot, z)\}$ defined in Section 2.2. (Thus, the name "adjusted rank".) The motivation behind Definition 3.2.1 and the problems stated below will become clear in Section 3.3.

We will consider the following two problems related to (NP) model.

Local tests for NP model. *For a fixed* $0 \leq L \leq G$ *and* z, *to test the hypothesis* H_0 : $\mathrm{adrk}\{F(\cdot, z)\} \leq L$, *against the alternative* H_1 : $\mathrm{adrk}\{F(\cdot, z)\} > L$, *as well as to determine* $\mathrm{adrk}\{F(\cdot, z)\}$.

Global tests for NP model. *For a fixed* $0 \leq L \leq G$, *to test the hypothesis* H_0 : $\sup_z \mathrm{adrk}\{F(\cdot, z)\} \leq L$, *against the alternative* H_1 : $\sup_z \mathrm{adrk}\{F(\cdot, z)\} > L$, *as well as to determine the global adjusted rank* $\sup_z \mathrm{adrk}\{F(\cdot, z)\}$.

Local and global tests for (NP) model are explored in Chapters 5 and 6, respectively. Finally, we state assumptions on the variables X_i, Z_i and U_i, and on the function F used in local tests.

ASSUMPTION (NP) L1: *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. (The parameter* r *is defined in Assumption* (NP) L4 *of Section 4.1.)*

ASSUMPTION (NP) L2: *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. Suppose also that* $E|U_i|^4 < \infty$.

ASSUMPTION (NP) L3: *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.*

Remark 3.2.1 Assumptions (NP) L1–L3 are similar to those used by Donald [28]. Observe also that, according to Assumption (NP) L3, the function F in (NP) model is assumed to be smooth.

3.3 Connections to demand systems

We will now explain how the models and problems introduced in Sections 3.1 and 3.2 are related to demand systems and their ranks. In Section 3.3.1, we consider the case of a semi-parametric model and, in Section 3.3.2, we deal with a non-parametric model.

3.3.1 The case of semi-parametric factor model

Consider first (SPF) model introduced in Section 3.1. In applications, we would like to think of it as a model for demand systems where the variables X_i, Z_i and Y_i denote the total income, prices (or demographic variables) and shares of goods corresponding to the ith consumer. Moreover, since our focus is on ranks of demand systems, we would like to determine the rank of a demand system given by a (SPF) model, namely, $\mathrm{rk}\{F(\cdot,z)\}$ where $F(x,z) = \Theta(z)V(x)$. In that case, the following elementary lemma may seem to explain why local and global tests stated in Section 3.1 are of interest.

Lemma 3.3.1 *Consider a demand system $y = F(x,z) = \Theta(z)V(x)$, where $\Theta(z)$ is a $G \times d$ matrix and $V(x)$ is a $d \times 1$ vector. Suppose that the vector $V(x)$ consists of linearly independent functions. Then, for fixed z,*

$$\mathrm{rk}\{F(\cdot,z)\} = \mathrm{rk}\{\Theta(z)\}, \tag{3.6}$$

where $\mathrm{rk}\{\Theta(z)\}$ stands for the rank of the matrix $\Theta(z)$.

PROOF: The proof is elementary but it is included for the sake of completeness. Let $R(z) = \mathrm{rk}\{F(\cdot,z)\}$ and $L(z) = \mathrm{rk}\{\Theta(z)\}$. By the definition of $\mathrm{rk}\{F(\cdot,z)\}$, there are $G - R(z)$ elements of a vector $F(x,z) = \Theta(z)V(x)$ that can be expressed as linear combinations of the rest $R(z)$ elements of $F(x,z)$. Supposing without loss of generality that these are the last $G - R(z)$ elements of $F(x, z)$, we obtain that, for $i = R(z) + 1, \ldots, G$,

$$\theta_{i1}(z)V_1(x) + \cdots + \theta_{id}(z)V_d(x) = \sum_{k=1}^{R(z)} c_{ik}(z)(\theta_{k1}(z)V_1(x) + \cdots + \theta_{kd}(z)V_d(x)), \qquad (3.7)$$

where $\Theta(z) = (\theta_{ij}(z))$, $V(x) = (V_i(x))$ and $c_{ij}(z)$ are some functions. Since the functions $V_1(x), \ldots, V_d(x)$ are linearly independent by the assumption, relation (3.7) implies that

$$\theta_{ij}(z) = c_{i1}(z)\theta_{1j}(z) + \cdots + c_{iR(z)}(z)\theta_{R(z)j}(z),$$

for $i = R(z)+1, \ldots, G$ and $j = 1, \ldots, d$. This shows that $L(z) \leq R(z)$ since $G - R(z)$ rows of the matrix $\Theta(z)$ can be expressed as linear combinations of the other $R(z)$ rows. To obtain the inverse inequality $R(z) \leq L(z)$, observe that $\Theta(z) = \Theta_1(z)\Theta_2(z)$ where $\Theta_1(z)$ is a $G \times L(z)$ matrix and $\Theta_2(z)$ is a $L(z) \times d$ matrix. Then, $y = F(x, z) = \Theta(z)V(x) = \Theta_1(z)(\Theta_2(z)V(x))$. Since $\Theta_2(z)V(x)$ is a $L(z) \times 1$ vector, we obtain from Lemma 2.2.1 that $R(z) \leq L(z)$. Hence, $L(z) = R(z)$ which concludes the proof of (3.6). □

Lemma 3.3.1 may suggest that local tests formulated in Section 3.1 allow to determine the local rank of a demand system given by (SPF) model. This is, however, not true under the assumptions made on (SPF) model in Section 3.1. Since the budget shares of goods in Y_i add up to 1, the variance-covariance matrix of the noise vector U_i is necessarily singular which violates Assumption (SPF) L2 stated in that section. This problem can be dealt with in a number of ways. One way would be to eliminate the non-singularity condition on Σ from our assumptions. This approach, however, would require extending most of the statistical work on estimation of rank to situations allowing for singular variance-covariance matrices. Another way, which we adopt here, is to keep the non-singularity assumption, but make appropriate modifications when applying our results to demand systems.

In applications, the idea is to drop a share of good from the analysis, in which case the summing up condition becomes no longer relevant, and then determine the rank of a demand system from the rank of a reduced demand system estimated by using tests formulated in Section 3.1. This procedure, summarized in greater detail below, is based on the following elementary result. To state this result, we need to introduce some notation. Let

$$y = f(x,z) = \theta(z)V(x) \tag{3.8}$$

be a demand system with a $J \times d$ matrix $\theta(z)$ and a $d \times 1$ vector $V(x)$. Let also

$$y^{(j)} = F^{(j)}(x,z) = \Theta^{(j)}(z)V(x), \tag{3.9}$$

$j = 1, \dots, J$, be a reduced demand system obtained by dropping the kth share of goods from the demand system (3.8). (In other words, $F^{(j)}(x,z)$ is the vector $f(x,z)$ without its jth element and $\Theta^{(j)}(z)$ is the matrix $\theta(z)$ where the jth row is eliminated.)

Lemma 3.3.2 *With the above notation, if $J > d$ and a vector $V(x)$ consists of linearly independent functions, we have*

$$\mathrm{rk}\{f(\cdot,z)\} = \mathrm{rk}\{\theta(z)\} = \max_{1\le j\le J} \mathrm{rk}\{\Theta^{(j)}(z)\} = \max_{1\le j\le J} \mathrm{rk}\{F^{(j)}(\cdot,z)\}. \tag{3.10}$$

PROOF: The first and the third equalities follow from Lemma 3.3.1. We thus only need to show the second equality in (3.10). Since $\Theta^{(j)}(z)$ is obtained by eliminating a row in $\theta(z)$, we have $\mathrm{rk}\{\theta(z)\} \ge \mathrm{rk}\{\Theta^{(j)}(z)\}$ for all $j = 1, \dots, J$, and hence

$$\mathrm{rk}\{\theta(z)\} \ge \max_{1\le j\le J} \mathrm{rk}\{\Theta^{(j)}(z)\}. \tag{3.11}$$

On the other hand, setting $L(z) = \mathrm{rk}\{\theta(z)\}$, there is a $L(z) \times L(z)$ sub-matrix $\widetilde{\theta}(z)$ of $\theta(z)$ such that the determinant of $\widetilde{\theta}(z)$ is not zero. Since $L(z) \le d < J$, there is a row $j_0 = j(z)$

such that $\widetilde{\theta}(z)$ is a submatrix of $\Theta^{(j_0)}(z)$. It follows that

$$\max_{1\leq j\leq J} \mathrm{rk}\{\Theta^{(j)}(z)\} \geq \mathrm{rk}\{\Theta^{(j_0)}(z)\} \geq L(z) = \mathrm{rk}\{\theta(z)\}, \tag{3.12}$$

since the matrix $\Theta^{(j_0)}(z)$ has a $L(z) \times L(z)$ submatrix with a nonzero determinant. The second equality in (3.10) now follows from (3.11) and (3.12). $\square$

Remark 3.3.1 Taking the maximum of ranks $\mathrm{rk}\{\Theta^{(j)}(z)\}$ in (3.10) is crucial. Without the maximum, the second relation in (3.10) does not hold in general, that is, it is not true that $\mathrm{rk}\{\theta(z)\} = \mathrm{rk}\{\Theta^{(j)}(z)\}$ for any j. Consider the following elementary example. Suppressing the dependence on z (which is fixed anyway), consider the 4×3 matrix

$$\theta = \begin{pmatrix} 1/2 & 1/2 & 1/2 \\ 1/2 & 1/2 & 1/2 \\ 0 & -1 & 1 \\ 0 & 0 & -2 \end{pmatrix}$$

and the corresponding reduced matrices

$$\Theta^{(1)} = \Theta^{(2)} = \begin{pmatrix} 1/2 & 1/2 & 1/2 \\ 0 & -1 & 1 \\ 0 & 0 & -2 \end{pmatrix}, \quad \Theta^{(3)} = \begin{pmatrix} 1/2 & 1/2 & 1/2 \\ 1/2 & 1/2 & 1/2 \\ 0 & 0 & -2 \end{pmatrix}$$

and $\Theta^{(4)}$ defined in a similar way. Then, $\mathrm{rk}\{\theta(z)\} = 3$, and $\mathrm{rk}\{\Theta^{(1)}(z)\} = \mathrm{rk}\{\Theta^{(2)}(z)\} = 3$ and $\mathrm{rk}\{\Theta^{(3)}(z)\} = \mathrm{rk}\{\Theta^{(4)}(z)\} = 2$. Observe that $\mathrm{rk}\{\theta(z)\} \neq \mathrm{rk}\{\Theta^{(j)}(z)\}$ when $j = 3, 4$, but $\mathrm{rk}\{\theta(z)\} = \max_{1\leq j\leq 4} \mathrm{rk}\{\Theta^{(j)}(z)\}$. Observe also that the entries in the first column of θ add up to 1 and those in the other two columns add up to 0. Hence, in particular, for the chosen θ, the shares of the demand system $y = \theta V(x)$ add up to 1 as long as the first coordinate function of a vector $V(x)$ is identical to 1.

Estimation of rank using (SPF) model. Based on Lemma 3.3.2, we propose to estimate the local rank $R(z)$ of a demand system

$$Y_i = \theta(Z_i)V(X_i) + \epsilon_i, \quad i = 1, \dots, N, \tag{3.13}$$

where $\theta(z)$ is a $J \times d$ unknown matrix, $V(x)$ is a $d \times 1$ known vector, ϵ_i are noise variables and $J > d$, as follows:

1. for each $j = 1, \dots, J$, eliminate the jth budget share from the analysis and consider a reduced demand system

$$Y_i^{(j)} = \Theta^{(j)}(Z_i)V(X_i) + U_i, \quad i = 1, \dots, N, \tag{3.14}$$

   where $Y_i^{(j)}$ and $\Theta^{(j)}(z)$ are the vector Y_i and the matrix $\theta(z)$ with its jth element and its jth row, respectively, eliminated,

2. for each $j = 1, \dots, J$, estimate the local rank $L^{(j)}(z)$ of a reduced system (3.14), that is, the rank of the matrix $\Theta^{(j)}(z)$, by using methods proposed for local tests of Section 3.1, and

3. to determine $R(z)$, take the maximum of the estimated ranks $\widehat{L}^{(j)}(z)$ over all j's.

The following remarks provide further comments and insight on the estimation algorithm above.

Remark 3.3.2 The above algorithm can be applied only when $J > d$. Recall that this restriction was key to obtain relation (3.10). We do not believe that, when $J \leq d$, there is a relation analogous to (3.10). In applications, J typically ranges between 5 and 8, so that one still has flexibility in choosing V.

Remark 3.3.3 In the statistical literature on rank estimation of demand systems, the approach described above can not be found to our best knowledge. One indeed drops a share of goods from the analysis as we did but one does not take the maximum of all estimated ranks at the end (one just takes the rank of that one, arbitrarily chosen reduced demand system). See, for example, Cragg and Donald [18], p. 1306, or Robin and Smith [89], p. 161. The author is not aware why the rank estimation results in such demand systems should be invariant to a share of goods eliminated from the analysis. (Although in other questions related to singular covariance equations, this may be true. See, for example, Berndt and Savin [12].) Another approach found in the literature is to work with the original (full) demand system but then modify the rank tests that are used. See, for example, Donald [28], p. 123, where the so-called minimum-χ^2 statistic for the rank of a matrix is compared to a χ^2-distribution having less degrees of freedom than without singularity restrictions. The author, however, has not seen a rigorous proof of the aforementioned result.

Remark 3.3.4 Recall from Section 2.3 that demand systems $y = \theta(z)V(x)$, known as exactly aggregable demand systems, are of special interest both in Economic Theory and in applications.

3.3.2 The case of non-parametric model

We now turn to (NP) model introduced in Section 3.2. Our goal is to motivate the problem of adjusted rank estimation formulated in that section. As in the case of applications of (SPF) model, because of the adding up condition of budget shares and the entailing singularity problem, we want to eliminate one share of goods from the analysis. In order to do so, we need to know how this elimination changes the rank of a demand system. The following lemma, which is implicit in Donald [28], provides the answer. Consider a demand system

$$y = f(x, z), \tag{3.15}$$

where $f(x, z)$ is a $J \times 1$ vector, and let

$$y^{(j)} = F^{(j)}(x, z), \tag{3.16}$$

where $F^{(j)}(x, z)$ is a $(J-1) \times 1$ vector, be a reduced demand system obtained by eliminating the jth share of goods in the demand system (3.15). Recall also the definition of the adjusted rank $\text{adrk}\{F(\cdot, z)\}$ introduced in Section 3.2.

Lemma 3.3.3 *With the above notation, we have that, for fixed z and any $j = 1, \ldots, J$,*

$$\text{rk}\{f(\cdot, z)\} = \text{adrk}\{F^{(j)}(\cdot, z)\} + 1. \tag{3.17}$$

PROOF: Suppose without loss of generality that $j = 1$ and set $R(z) = \text{rk}\{f(\cdot, z)\}$. Then, by Lemma 2.2.1,

$$f(x, z) = a(z)h(x, z), \tag{3.18}$$

where $a(z) = (a_{kl}(z))$ is a $J \times R(z)$ matrix and $h(x, z) = (h_l(x, z))$ is a $R(z) \times 1$ vector. Since the J shares add up to 1, we obtain from (3.18) that

$$1 = \left(\sum_{k=1}^{J} a_{k1}(z)\right) h_1(x, z) + \cdots + \left(\sum_{k=1}^{J} a_{kR(z)}(z)\right) h_{R(z)}(x, z).$$

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

$$\begin{aligned} h_1(x, z) = {} & \left(\sum_{k=1}^{J} a_{k1}(z)\right)^{-1} - \left(\sum_{k=1}^{J} a_{k1}(z)\right)^{-1} \left(\sum_{k=1}^{J} a_{k2}(z)\right) h_2(x, z) \\ & - \cdots - \left(\sum_{k=1}^{J} a_{k1}(z)\right)^{-1} \left(\sum_{k=1}^{J} a_{kR(z)}(z)\right) h_{R(z)}(x, z). \end{aligned} \tag{3.19}$$

Substituting (3.19) into (3.18), we can conclude that

$$F^{(1)}(x,z) = c(z) + A(z)H(x,z), \tag{3.20}$$

where $A(z)$ is a $(J-1) \times (R(z)-1)$ matrix, $H(x,z)$ is a $(R(z)-1) \times 1$ vector and $c(z)$ is a $(J-1) \times 1$ vector. In view of Definition 3.2.1, (3.20) implies that

$$\text{adrk}\{F^{(1)}(\cdot,z)\} \leq R(z) - 1. \tag{3.21}$$

To show the converse, observe that, by using (3.5), the elements $f_2(x,z), \ldots, f_J(x,z)$ of $F^{(1)}(x,z)$ can be expressed as linear combinations of $\text{adrk}\{F^{(1)}(\cdot,z)\}+1$ functions of x and z. Since $f_1(x,z) = 1 - f_2(x,z) - \cdots - f_J(x,z)$, the function $f_1(x,z)$ can be also expressed as a linear combination of these $\text{adrk}\{F^{(1)}(\cdot,z)\} + 1$ functions. In view of Definition 2.2.1, we obtain that

$$R(z) = \text{rk}\{f(\cdot,z)\} \leq \text{adrk}\{F^{(1)}(\cdot,z)\} + 1. \tag{3.22}$$

The conclusion follows from (3.21) and (3.22). $\square$

Example 3.3.1 Suppressing the dependence on z, consider for example the 4×1 vector

$$f(x) = (1 - x - x^2 \quad -x \quad 2x \quad x^2)'$$

with its components summing up to 1, and the corresponding reduced vectors

$$F^{(1)}(x) = \begin{pmatrix} -x \\ 2x \\ x^2 \end{pmatrix}, \quad F^{(2)}(x) = \begin{pmatrix} 1 - x - x^2 \\ 2x \\ x^2 \end{pmatrix},$$

and $F^{(3)}(x)$ and $F^{(4)}(x)$ defined in a similar way. Observe that $\text{rk}\{f\} = 3$ and that

$\mathrm{adrk}\{F^{(j)}\} = 2$ for $j = 1,2,3,4$. To see, for example, the last equality for $j = 2$, observe first that

$$F^{(2)}(x) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} -1 & 1 \\ 2 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ x^2 \end{pmatrix},$$

showing that $\mathrm{adrk}\{F^{(2)}\} \leq 2$. If, however, there is a function $g(x)$ such that

$$F^{(2)}(x) = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} g(x)$$

(that is, $\mathrm{adrk}\{F^{(2)}\} \leq 1$), then this yields in particular that $2x = a_2 + b_2 g(x)$ and hence that $g(x) = c_1 + c_2 x$ for some c_1, c_2. This $g(x)$, however, cannot satisfy another required relation $x^2 = a_3 + b_3 g(x)$.

Estimation of rank using (NP) model. Based on Lemma 3.3.3, we propose to estimate the local rank $R(z) = \mathrm{rk}\{f(\cdot, z)\}$ of a demand system

$$Y_i = f(X_i, Z_i) + \epsilon_i, \quad i = 1, \ldots, N, \tag{3.23}$$

where $f(x,z)$ is a $J \times 1$ unknown vector and ϵ_i are noise variables, as follows:

1. fix one $j \in \{1, \ldots, J\}$, eliminate the jth budget share from the analysis and consider a reduced demand system

$$Y_i^{(j)} = F^{(j)}(X_i, Z_i) + U_i, \quad i = 1, \ldots, N, \tag{3.24}$$

   where $Y_i^{(j)}$ and $F^{(j)}(x,z)$ are the vectors Y_i and $f(x,z)$ with their jth element eliminated,

2. for that fixed j, estimate the local adjusted rank $L^{(j)}(z) = \mathrm{adrk}\{F^{(j)}(\cdot, z)\}$ of a

reduced demand system (3.24) by using methods proposed for local tests of Section 3.2, and

3. to determine $R(z)$, add 1 to the estimated adjusted rank $\widehat{L}^{(j)}(z)$.

Remark 3.3.5 The idea to determine the rank of demand system by adding 1 to the estimated adjusted rank, can be found in Donald [28]. (The term "adjusted rank" is not used by Donald [28].) Observe also that, in contrast to rank estimation for the semi-parametric factor model, it does not matter here which share of goods is eliminated.

Chapter 4

Kernel Based Estimators

In this chapter, we introduce some estimators for (SPF) and (NP) models, and establish some of their properties. Since these estimators are kernel based, we first recall in Section 4.1 the definition and the localization property of a kernel function. The estimators are defined and their properties are stated in Section 4.2. Sections 4.3 and 4.4 contain the proofs of the results of Section 4.2.

4.1 Kernel functions

We first give the definition of a kernel function (a kernel). It uses the notation $x^b = x_1^{b_1} \cdots x_m^{b_m}$ for $b = (b_1, \ldots, b_m) \in (\mathbb{N} \cup \{0\})^m$, $x = (x_1, \ldots, x_m) \in \mathbb{R}^m$ and $|b| = b_1 + \cdots + b_m$.

Definition 4.1.1 *(Kernel function) 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 $b \in (\mathbb{N} \cup \{0\})^m$ satisfying $1 \leq |b| < r$.*

Example 4.1.1 Some of the well-known and commonly used kernel functions on $\mathbb{R}$ are the uniform kernel (see Figure 4·1 below) given by

$$K(x) = \frac{1}{2} 1_{\{|x| \leq 1\}}, \quad x \in \mathbb{R},$$

the triangle kernel given by

$$K(x) = (1 - |x|) 1_{\{|x| \leq 1\}}, \quad x \in \mathbb{R},$$

or the Epanechnikov kernel given by

$$K(x) = \frac{3}{4}(1 - x^2)1_{\{|x| \le 1\}}, \; x \in \mathbb{R}.$$

For more examples, see Devroye [26], Devroye and Györfi [25] and Härdle [48, 49].

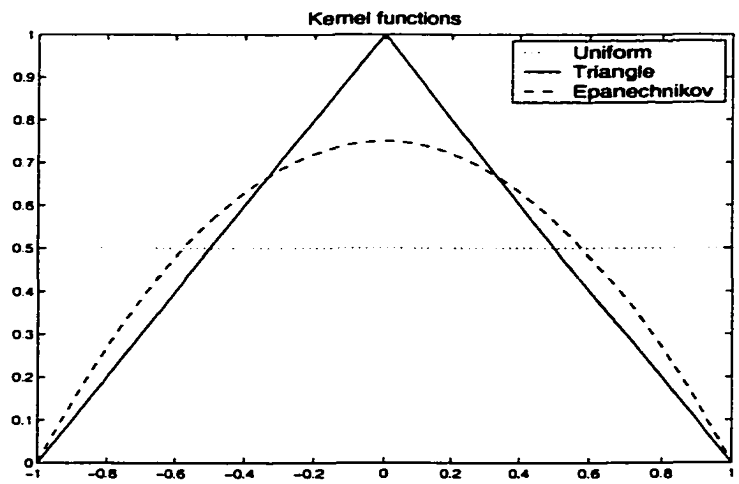

Figure 4·1: Uniform, triangle and Epanechnikov kernels

Remark 4.1.1 Observe that, if K is a kernel on $\mathbb{R}^m$ of order r and $\widetilde{K}$ is a kernel on $\mathbb{R}^n$ of order $\widetilde{r}$, then

$$K_0(x_1, x_2) = K(x_1)\widetilde{K}(x_2), \; (x_1, x_2) \in \mathbb{R}^m \times \mathbb{R}^n,$$

is a kernel on $\mathbb{R}^m \times \mathbb{R}^n$ of order $\min\{r, \widetilde{r}\}$. This provides a way to construct kernels on $\mathbb{R}^m$, $m \ge 2$, by using kernels on $\mathbb{R}$.

Remark 4.1.2 We assumed in Definition 4.1.1 that a kernel K has a compact support. One may consider kernels with unbounded support as well. However, it is a common belief that results with compactly supported kernels will continue to hold for unbounded support kernels as well when proper modifications in assumptions of these results are made. In our work, we decided not to include these technical details.

Kernel functions are used in statistics, as well as in other areas of applied or pure mathematics, because of their localization property. We recall the localization property in the following proposition. Since the proposition will be used many many times throughout the rest of the thesis, we provide its proof to the readers convenience. For notational simplicity, we set

$$K_h(x) = \frac{1}{h^m} K\left(\frac{x}{h}\right), \tag{4.1}$$

where $h > 0$ is the so-called bandwidth (or smoothing parameter).

Proposition 4.1.1 *(Localization property) Let K be a kernel on $\mathbb{R}^m$ of order $r \in \mathbb{N}$. Suppose that a function $g : \mathbb{R}^m \to \mathbb{R}^k$ is $r-$times continuously differentiable in a neighborhood of $z_0 \in \mathbb{R}^m$. Then, as $h \to 0$,*

$$\int_{R^m} g(z) K_h(z - z_0) dz = g(z_0) + O(h^r). \tag{4.2}$$

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

Moreover, if the function g ha[illegible], then the term $O(h^r)$ in (4.2) does not depend on z_0.

Proof : We will suppose for simplicity that $k = 1$, that is, the range of the function g is $\mathbb{R}$. The case of a general k can be proved by applying the proof below to each component of the vector g. By using the Taylor's formula for the function $g(z)$ around $z = z_0$, we can express the integral in (4.2) as

$$\int_{R^m} \left(\sum_{j=0}^{r-1} \frac{1}{j!} \sum_{|k|=j} \frac{\partial^k g}{\partial z^k}(z_0)(z - z_0)^k + \frac{1}{r!} \sum_{|k|=r} \frac{\partial^k g}{\partial z^k}(z_0 + \theta(z - z_0))(z - z_0)^k \right) K_h(z - z_0) dz, \tag{4.3}$$

where $k = (k_1, \ldots, k_m) \in (\mathbb{N} \cup \{0\})^m$, $\partial^k g / \partial^k z^k = \partial^{k_1} \cdots \partial^{k_m} g / \partial z_1^{k_1} \cdots \partial z_m^{k_m}$ and $\theta \in (0,1)$ may also depend on z. Since K is a kernel of order r, we have

$$\int_{R^m} (z - z_0)^k K_h(z - z_0) dz = 0$$

for any $k = (k_1, \ldots, k_m)$ such that $1 \leq |k| \leq r - 1$. Then, the integral (4.3) becomes

$$g(z_0) + \int_{R^m} \frac{1}{r!} \sum_{|k|=r} \frac{\partial^k g}{\partial z^k}(z_0 + \theta(z - z_0))(z - z_0)^k K_h(z - z_0) dz =: g(z_0) + I. \qquad (4.4)$$

Since $\partial^k g / \partial z^k$ is continuous in a neighborhood of z_0 by assumption, there is $\epsilon > 0$ such that

$$\max_{|k|=r, |z - z_0| < \epsilon} \left| \frac{\partial^k g}{\partial z^k}(z) \right| = M_\epsilon(z_0) < \infty.$$

Since K has a compact support $\text{supp}\{K\}$, there is $A > 0$ such that $\text{supp}\{K\} \subset (-A, A)^m$. Then, since $\text{supp}\{K_h\} \subset (-\epsilon, \epsilon)^m$ with $\epsilon > hA$ (or $h < \epsilon/A$), by using the change of variables $z = vh + z_0$ below, we obtain

$$\begin{aligned}
|I| &\leq \frac{1}{r!} \sum_{|k|=r} M_\epsilon(z_0) \int_{(-\epsilon,\epsilon)^m} |z - z_0|^k |K_h(z - z_0)| dz \\
&\leq \frac{1}{r!} \sum_{|k|=r} M_\epsilon(z_0) \int_{R^m} |z - z_0|^k \left| \frac{1}{h^m} K\left(\frac{z - z_0}{h}\right) \right| dz \\
&= \frac{1}{r!} \sum_{|k|=r} M_\epsilon(z_0) h^{|k|} \int_{R^m} |v|^k |K(v)| dv = Ch^r,
\end{aligned}$$

where the constant C is given by $C = (r!)^{-1} M_\epsilon(z_0) \sum_{|k|=r} \int_{R^m} |v|^k |K(v)| dv$. This proves the proposition. $\square$

Observe that, for any integer $j \geq 1$, a *symmetric* kernel K of order $r \geq 1$ induces a kernel $K^{(2j)}(x) = (K(x))^{2j}/\|K\|_{2j}^{2j}$ of order 2. We will often use the notation

$$K_{2j,h}(x) = \frac{1}{h^m} K^{(2j)}\left(\frac{x}{h}\right) = \frac{1}{\|K\|_{2j}^{2j} h^m} \left(K\left(\frac{x}{h}\right)\right)^{2j} \tag{4.5}$$

to denote the kernel $K^{(2j)}$ scaled by a bandwidth parameter $h > 0$.

Finally, we state the assumptions on the kernel functions which will be used in the sequel to state our results.

Assumption (SPF) L5: The function K is a symmetric kernel on $\mathbb{R}^m$ of order r.

Assumption (NP) L4: The functions $\widetilde{K}$ and K are symmetric kernels on $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, of order r.

4.2 Estimators for two models

In this section, we introduce some kernel based estimators for (SPF) and (NP) models. We also establish some of their properties that will be used in the following chapters. We first consider the (NP) model which involves well-known estimators.

In the case of (NP) model, we will need an estimator of an unknown function F. We will use for it the well-known Nadaraya-Watson estimator

$$\widehat{F}(x, z) = \frac{1}{N} \sum_{i=1}^{N} Y_i \widetilde{K}_h(x - X_i) K_h(z - Z_i)\ \widehat{p}(x, z)^{-1}, \tag{4.6}$$

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

$$\widehat{p}(x, z) = \frac{1}{N} \sum_{i=1}^{N} \widetilde{K}_h(x - X_i) K_h(z - Z_i). \tag{4.7}$$

Such estimators have been extensively studied in the statistical literature. See, for example, Johnston [54], Prakasa Rao [83], Devroye and Györfi [26], Devroye [25] or Härdle [48, 49].

The basic idea behind the estimator $\widehat{F}(x,z)$ can be expressed as

$$\begin{aligned} E\widehat{F}(x,z)\widehat{p}(x,z) &= EY_i\widetilde{K}_h(x-X_i)K_h(z-Z_i) \\ &= EF(X_i,Z_i)\widetilde{K}_h(x-X_i)K_h(z-Z_i) \\ &= \int_{\mathbb{R}^{n+m}} F(x_1,z_1)p(x_1,z_1)\widetilde{K}_h(x-x_1)K_h(z-z_1)dx_1dz_1 \\ &\approx F(x,z)p(x,z), \end{aligned} \tag{4.8}$$

where $p(x,z)$ denotes the density of (X_i,Z_i) and where in the last step we used the localization property of the kernel function $\widetilde{K}(\cdot)K(\cdot)$ (see Proposition 4.1.1). Under suitable conditions, one can show that $\widehat{F}(x,z)$ is a consistent and asymptotically normal estimator for $F(x,z)$. For more information on the estimators $\widehat{F}$ and $\widehat{p}$, see the references indicated above.

In addition to the estimator $\widehat{F}$, we will also consider

$$\widehat{\Sigma} = \frac{1}{N}\sum_{i=1}^{N}(Y_i-\widehat{F}(X_i,Z_i))(Y_i-\widehat{F}(X_i,Z_i))', \tag{4.9}$$

which is an estimator for the variance-covariance matrix Σ of the noise U_i. As we will see in Lemma 5.2.11 of Section 5.2.5 below, under suitable conditions, $\widehat{\Sigma} = \Sigma + o_p(1)$ and hence that $\widehat{\Sigma}$ is a consistent estimator for Σ.

In the case of (SPF) model, we will need an estimator of an unknown matrix $\Theta(z)$. We define it as follows.

Definition 4.2.1 *(Estimator for* (SPF) *model) For fixed z, let*

$$\widehat{\Theta}(z) = \frac{1}{N}\sum_{i=1}^{N}Y_iV(X_i)'K_h(z-Z_i)\widehat{Q}(z)^{-1}, \tag{4.10}$$

where

$$\widehat{Q}(z) = \frac{1}{N}\sum_{i=1}^{N} V(X_i)V(X_i)'K_h(z - Z_i). \tag{4.11}$$

Remark 4.2.1 The estimator $\widehat{\Theta}(z)$ is a generalization of an estimator considered by Li, Huang, Li and Fu [69]. These authors considered our (SPF) model with the function

$$V(x) = \begin{pmatrix} 1 \\ x \end{pmatrix}, \quad x \in \mathbb{R}^n,$$

and also defined the estimator of $\widehat{\Theta}(z)$ by (4.10) and (4.11) where $V(x)$ is replaced by $(1 \;\; x)'$. The definition of $\Theta(z)$ is also similar to local linear regression estimators (see, for example, Fan and Gijbels [31]) and to estimators in varying coefficient models (see, for example, Fan and Zhang [32]).

Remark 4.2.2 The estimator $\widehat{\Theta}(z)$ in (4.10) can be viewed as the solution to the minimization problem

$$\widehat{\Theta}(z) = \operatorname*{argmin}_{\Theta(z)} \sum_{i=1}^{N} |Y_i - \Theta(z)V(X_i)|^2 \, K_h(z - Z_i). \tag{4.12}$$

Remark 4.2.3 The estimator $\widehat{\Theta}(z)$ can be also expressed as

$$\widehat{\Theta}(z) = YDV'(VDV')^{-1}, \tag{4.13}$$

where Y is a $G \times N$ matrix with the entries Y_i for its N columns, V is a $d \times N$ matrix with $V(X_i)$ for its N columns and D is the $N \times N$ diagonal matrix

$$D = \begin{pmatrix} K_h(z - Z_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & K_h(z - Z_N) \end{pmatrix}.$$

The expression (4.13) is similar to that of the Generalized Least Squares estimator with the weight matrix D. (This fact can also be seen from (4.12).)

Observe that the expression (4.10) involves the inverse of the matrix $\widehat{Q}(z)$. By, the next proposition, the matrix $\widehat{Q}(z)$ converges in probability to the matrix $Q(z)$ defined in (3.2). Since, by Assumption (SPF) L4, the matrix $Q(z)$ is positive definite (thus invertible), we have that $\widehat{Q}(z)$ is invertible with probability approaching 1 as $N \to \infty$.

Proposition 4.2.1 *(Consistency of $\widehat{Q}(z)$) Under Assumptions* (SPF) L1–L3 *of Section 3.1 and* (SPF) L5 *of Section 4.1, as $h \to 0$ and $Nh^m \to \infty$, we have that, for fixed z,*

$$\widehat{Q}(z) \xrightarrow{p} Q(z), \tag{4.14}$$

where $\widehat{Q}(z)$ and $Q(z)$ are given by (4.11) and (3.2), respectively.

Proposition 4.2.1 is proved in the next section. The following two results show that the estimator $\widehat{\Theta}(z)$ is consistent and asymptotically normal. Asymptotic normality is key to local tests for (SPF) model. Note also that we obtain the rate of convergence in the result on consistency.

Theorem 4.2.1 *(Consistency of $\widehat{\Theta}(z)$) Under Assumptions* (SPF) L1–L5 *of Sections 3.1 and 4.1, we have for fixed z,*

$$\widehat{\Theta}(z) - \Theta(z) = O_p\left(h^r + \frac{1}{\sqrt{Nh^m}}\right). \tag{4.15}$$

Moreover, the optimal order in (4.15) is $N^{-r/(2r+m)}$, which is obtained by taking

$$h = O\left(N^{-1/(2r+m)}\right). \tag{4.16}$$

Theorem 4.2.2 *(Asymptotic normality of $\widehat{\Theta}(z)$) Under Assumptions* (SPF) L1–L5 *of*

Sections 3.1 and 4.1, for fixed z, we have

$$\sqrt{Nh^m}\text{vec}(\widehat{\Theta}(z) - \Theta(z)) \xrightarrow{d} \mathcal{N}(0, W(z)) \tag{4.17}$$

as

$$N \to \infty, \quad h \to 0, \quad Nh^m \to \infty \quad \textit{and} \quad Nh^{m+2r} \to 0, \tag{4.18}$$

with

$$W(z) = (Q(z)^{-1} \otimes \Sigma)\|K\|_2^2, \tag{4.19}$$

where $\otimes$ denotes the Kronecker product.

The proof of Theorem 4.2.1 can be found in Section 4.3 and that of Theorem 4.2.2 in Section 4.4.

Remark 4.2.4 Recall that the vec operation of a $m \times n$ matrix $A = (a_{ij})$ is defined as the $mn \times 1$ vector $\text{vec}(A) = (a_{11} \dots a_{n1} \dots a_{m1} \dots a_{mn})'$, that is, the columns of A are stacked one underneath the other. Recall also that the Kronecker product $\otimes$ of a $m \times n$ matrix $A = (a_{ij})$ and a $p \times q$ matrix $B = (b_{ij})$ is defined as the $mp \times nq$ matrix

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix}. \tag{4.20}$$

The basic properties of the Kronecker product, some of which be also used in this thesis below, are $A \otimes (B + C) = A \otimes B + A \otimes C$, $(A \otimes B)' = A' \otimes B'$, $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ (when A and B are non-singular), $\text{vec}(AB) = (B' \otimes I_m)\text{vec}(A) = (I_q \otimes A)\text{vec}(B)$ (when $n = p$ and where I_k is a $k \times k$ identity matrix) and others. See, for example, Rao and Rao [85] or Magnus and Neudecker [73]).

Remark 4.2.5 One can prove the asymptotic normality result under the heteroskedasticity assumption (3.3) in Remark 3.1.2. In this case, the limit variance-covariance matrix $W(z)$ can be shown to be expressed as

$$\begin{aligned} W(z) &= p(z)E\left((Q(z)^{-1}V(X_1)V(X_1)'Q(z)^{-1}) \otimes \Sigma(X_1, Z_1)|Z_1 = z\right) \|K\|_2^2 \\ &= \|K\|_2^2 \int_{\mathbb{R}^n} ((Q(z)^{-1}V(x_1)V(x_1)'Q(z)^{-1}) \otimes \Sigma(x_1, z))p(x_1, z)dx_1, \end{aligned} \quad (4.21)$$

where $p(z)$ is the density function of Z_i.

In practice, the limiting variance-covariance matrix $W(z)$ can be estimated by

$$\widehat{W}(z) = (\widehat{Q}(z)^{-1} \otimes \widehat{\Sigma})\|K\|_2^2,$$

where $\widehat{\Sigma}$ is now defined as

$$\widehat{\Sigma} = \frac{1}{N}\sum_{n=1}^{N}(Y_i - \widehat{\Theta}(Z_i)V(X_i))(Y_i - \widehat{\Theta}(Z_i)V(X_i))'. \quad (4.22)$$

One can show that $\widehat{\Sigma}$ is a consistent estimator of Σ. Since $\widehat{Q}(z)$ is a consistent estimator of $Q(z)$ by using Proposition 4.2.1, we have that $\widehat{W}(z) \to_p W(z)$.

Finally, we state a multi-dimensional analogue of Theorem 4.2.2. This result, which is proved in Section 4.4, will be used in Chapter 6 where we focus on global tests.

Theorem 4.2.3 *(Normality and independence of $(\widehat{\Theta}(z_1), \ldots, \widehat{\Theta}(z_q))$) Let $z_1, \ldots, z_q$ be fixed different values of z. Suppose that the conditions of Theorem 4.2.2 are satisfied for all z_i, $i = 1, \ldots, q$. Then, we have*

$$\sqrt{Nh^m}\left(\text{vec}(\widehat{\Theta}(z_1) - \Theta(z_1)), \ldots, \text{vec}(\widehat{\Theta}(z_q) - \Theta(z_q))\right)$$

$$\xrightarrow{d} (\mathcal{N}_1(0, W(z_1)), \ldots, \mathcal{N}_q(0, W(z_q))), \tag{4.23}$$

where $\mathcal{N}_i$, $i = 1, \ldots, q$, are independent normal random vectors with covariance matrices $W(z_i)$ defined by (4.19).

4.3 The proof of Theorem 4.2.1 and Proposition 4.2.1

To establish the properties of the estimator $\widehat{\Theta}(z)$ in (4.10), it is convenient to introduce the matrices

$$\Delta_1(z) = \frac{1}{N}\sum_{i=1}^{N}(\Theta(Z_i) - \Theta(z))V(X_i)V(X_i)'K_h(z - Z_i), \tag{4.24}$$

$$\Delta_2(z) = \frac{1}{N}\sum_{i=1}^{N}U_i V(X_i)'K_h(z - Z_i). \tag{4.25}$$

Observe that by using (4.10), the (SPF) relation $Y_i = \Theta(Z_i)V(X_i) + U_i = \Theta(z)V(X_i) + (\Theta(Z_i) - \Theta(z))V(X_i) + U_i$ and the expression (4.11) for $\widehat{Q}(z)$, we have

$$\widehat{\Theta}(z) = \Theta(z) + (\Delta_1(z) + \Delta_2(z))\ \widehat{Q}(z)^{-1}. \tag{4.26}$$

In the next result, we show that, under suitable conditions, the matrices $\Delta_1(z)$ and $\Delta_2(z)$ are asymptotically negligible and we also obtain their rates of convergence. Consistency of the estimator $\widehat{\Theta}(z)$ and the corresponding rate of convergence will then follow directly from this result and the decomposition (4.26).

Lemma 4.3.1 *Under Assumptions* (SPF) L1–L3, L5, *for fixed z, we have*

$$\Delta_1(z) = O_p\big(h^r + h/\sqrt{Nh^m}\big), \tag{4.27}$$

$$\Delta_2(z) = O_p\big(1/\sqrt{Nh^m}\big). \tag{4.28}$$

PROOF: We use the notation $M^2 = MM'$ for a matrix M. Since, by the Chebyshev-Markov inequality, with $\epsilon > 0$ and $a \in \mathbb{R}^G$ such that $|a| = 1$,

$$P\left(|a'\Delta_1(z)|^2 > \epsilon\right) = P(a'\Delta_1(z)^2 a > \epsilon) \leq \epsilon^{-1} a' E\Delta_1(z)^2 a, \tag{4.29}$$

it is enough to show that $E\Delta_1(z)^2 = O(h^{2r} + h^2/Nh^m)$. Indeed, by choosing a vector $a = (0, \ldots, 0, 1, 0, \ldots, 0)'$ with 1 in the ith place, one would get that each row of $\Delta_1(z)$ converges to 0 in probability with the specified rate. By using (4.24) and independence of (X_i, Z_i)'s, we have $E\Delta_1(z)^2 = N^{-1}S_1 + N^{-1}(N-1)S_2$, where

$$\begin{aligned} S_1 &= E\left((\Theta(Z_1) - \Theta(z))V(X_1)V(X_1)'K_h(z - Z_1)\right)^2, \\ S_2 &= \left(E(\Theta(Z_1) - \Theta(z))V(X_1)V(X_1)'K_h(z - Z_1)\right)^2. \end{aligned}$$

Consider first the term S_1, which can be expressed as

$$S_1 = \frac{\|K\|_2^2}{h^m} \int_{\mathbb{R}^n} \left\{ \int_{\mathbb{R}^m} ((\Theta(z_1) - \Theta(z))V(x_1)V(x_1)')^2 p(x_1, z_1) K_{2,h}(z - z_1) dz_1 \right\} dx_1, \tag{4.30}$$

where the kernel $K_{2,h}$ is defined by (4.5). By using Assumptions (SPF) L1, L3 and L5, since $K_{2,h}$ is a kernel of order 2, we can apply Proposition 4.1.1 to the integral in the braces in (4.30) to obtain that

$$S_1 = O\left(\frac{h^2}{h^m}\right).$$

Similarly, by using Proposition 4.1.1,

$$S_2 = \left(\int_{\mathbb{R}^n} \left\{ \int_{\mathbb{R}^m} (\Theta(z_1) - \Theta(z))V(x_1)V(x_1)'p(x_1, z_1) K_h(z - z_1) dz_1 \right\} dx_1 \right)^2 = O(h^{2r}),$$

since K_h is a kernel of order r. This shows that

$$E\Delta_1(z)^2 = \frac{S_1}{N} + \frac{N-1}{N}S_2 = O\left(\frac{h^2}{Nh^m} + h^{2r}\right).$$

Consider now the matrix $\Delta_2(z)$ in (4.25). As in the case of the matrix $\Delta_1(z)$, we need to consider

$$\begin{aligned} E\Delta_2(z)^2 &= \frac{\|K\|_2^2}{Nh^m}E((U_1V(X_1)')^2K_{2,h}(z-Z_1)) \\ &+ \frac{N-1}{N}E(U_1V(X_1)'V(X_2)U_2'K_h(z-Z_1)K_h(z-Z_2)) \\ &= \frac{\|K\|_2^2}{Nh^m}E\left(V(X_1)'V(X_1)\ \Sigma\ K_{2,h}(z-Z_1)\right), \end{aligned}$$

since $E(U_1U_2'|X_i,Z_j) = 0$ and $E(U_1U_1'|X_1,Z_1) = \Sigma$. Proposition 4.1.1 yields $E\Delta_2(z)^2 = O((Nh^m)^{-1})$. □

Proof of Theorem 4.2.1: The result (4.15) follows from (4.26) and Lemma 4.3.1. To obtain the optimal rate in (4.15), set $h^r = 1/\sqrt{Nh^m}$. This yields $h = O(N^{-1/(2r+m)})$. □

We will now prove Proposition 4.2.1 which concerns the convergence of the matrix $\widehat{Q}(z)$.

Proof of Proposition 4.2.1: We will show that $\widehat{Q}(z) - Q(z) \to_p 0$. Proceeding as in the proof of Lemma 4.3.1, we need to consider

$$E(\widehat{Q}(z) - Q(z))^2 = E\widehat{Q}(z)^2 - E\widehat{Q}(z)Q(z)' - Q(z)E\widehat{Q}(z)' + Q(z)^2. \tag{4.31}$$

Since

$$\begin{aligned} E\widehat{Q}(z) &= EV(X_i)V(X_i)'K_h(z-Z_i) \\ &= \int_{\mathbb{R}^n}\left\{\int_{\mathbb{R}^m} V(x_1)V(x_1)'p(x_1,z_1)K_h(z-z_1)dz_1\right\}dx_1, \end{aligned}$$

by applying Proposition 4.1.1 to the above integral in the braces, we obtain that

$$\begin{aligned} E\widehat{Q}(z) &= \int_{\mathbb{R}^n} V(x_1)V(x_1)'p(x_1,z)dx_1 + O(h^r) \\ &= p(z)\int_{\mathbb{R}^n} V(x_1)V(x_1)'p(x_1|Z_1=z)dx_1 + O(h^r) \\ &= Q(z) + O(h^r). \end{aligned}$$

As for $E\widehat{Q}(z)^2$, by using independence of (X_i, Z_i) and (X_j, Z_j) for $i \neq j$, we have

$$E\widehat{Q}(z)^2 = \frac{\|K\|_2^2}{Nh^m} E\left((V(X_1)V(X_1)')^2 K_{2,h}(z - Z_1)\right)$$

$$+\frac{N-1}{N}\left(E(V(X_1)V(X_1)'K_h(z-Z_1))\right)^2. \tag{4.32}$$

By using Proposition 4.1.1, the first term in (4.32) is of the order $O((Nh^m)^{-1})$. The order of the second term is that of $(E\widehat{Q}(z))^2 = Q(z)^2 + O(h^r)$. Now, by substituting the obtained orders back into (4.31), we obtain that $\widehat{Q}(z) = Q(z) + O_p(h^r + 1/\sqrt{Nh^m})$. This shows that $\widehat{Q}(z) \to_p Q(z)$. □

4.4 The proof of Theorem 4.2.2 and Theorem 4.2.3

In this section, we prove Theorems 4.2.2 and 4.2.3 which show that $\widehat{\Theta}(z)$ is asymptotically normal and also that $\widehat{\Theta}(z)$'s are asymptotically independent for different values of z.

Proof of Theorem 4.2.2: By using (4.26) and the property $\mathrm{vec}(AB) = (B' \otimes I_m)\mathrm{vec}(A)$, where A is an $m \times n$ matrix and B is an $n \times q$ matrix (see, for example, Theorem 2 on p.

30 in Magnus and Neudecker [73]), we have

$$\begin{aligned}\sqrt{Nh^m}\text{vec}(\widehat{\Theta}(z) - \Theta(z)) &= \sqrt{Nh^m}\text{vec}((\Delta_1(z) + \Delta_2(z))\widehat{Q}(z)^{-1}) \\ &= \sqrt{Nh^m}(\widehat{Q}(z)^{-1} \otimes I_G)\text{vec}(\Delta_1(z) + \Delta_2(z)) \\ &= \sqrt{Nh^m}(\widehat{Q}(z)^{-1} \otimes I_G)(\text{vec}(\Delta_1(z)) + \text{vec}(\Delta_2(z))).\end{aligned}$$

Since by Proposition 4.2.1 and Slutsky's theorem, $\widehat{Q}(z)^{-1} \to_p Q(z)^{-1}$, it is enough to show that

$$\sqrt{Nh^m}(Q(z)^{-1} \otimes I_G)\big(\text{vec}(\Delta_1(z)) + \text{vec}(\Delta_2(z))\big) \xrightarrow{d} \mathcal{N}(0, W(z)).$$

In view of the definition (4.19) of $W(z)$ and also the property $(A \otimes B)(C \otimes D)(E \otimes F) = ACE \otimes BDF$ of the Kronecker product, this is equivalent to showing that

$$\sqrt{Nh^m}\big(\text{vec}(\Delta_1(z)) + \text{vec}(\Delta_2(z))\big) \xrightarrow{d} \mathcal{N}(0, W_0(z)), \tag{4.33}$$

where $W_0(z) = (Q(z) \otimes \Sigma)\|K\|_2^2$, or by using (4.24), (4.25) and the notation

$$\xi_{N,i} = \sqrt{h^m}\big(\text{vec}((\Theta(Z_i) - \Theta(z))V(X_i)V(X_i)') + \text{vec}(U_iV(X_i)')\big)K_h(z - Z_i)$$

$$= \sqrt{h^m}\big(((V(X_i)V(X_i)') \otimes I_G)\text{vec}(\Theta(Z_i) - \Theta(z)) + (V(X_i) \otimes I_G)U_i\big)K_h(z - Z_i), \tag{4.34}$$

to the convergence

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \xi_{N,i} \xrightarrow{d} \mathcal{N}(0, W_0(z)).$$

The Cramér-Wald theorem (see, for example, p. 18 in Serfling [95]) implies that the last convergence is equivalent to

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \eta_{N,i} \xrightarrow{d} \mathcal{N}(0, W_1(z)), \tag{4.35}$$

where $\eta_{N,i} = \lambda'\xi_{N,i}$, $\lambda \in \mathbb{R}^{Gd} \setminus \{0\}$ is an arbitrary vector and $W_1(z) = \lambda' W_0(z)\lambda$. By using the Lyapunov's version of Central Limit Theorem for triangular arrays (see, for example, p. 32 in Serfling [95]), to show the convergence (4.35), it is enough to prove that

$$\frac{E|\eta_{N,1} - E\eta_{N,1}|^4}{N(E|\eta_{N,1} - E\eta_{N,1}|^2)^4} \to 0, \tag{4.36}$$

as $N \to \infty$, and also

$$E\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\eta_{N,i}\right)^2 = E\eta_{N,1}^2 + (N-1)(E\eta_{N,1})^2 \to W_1(z), \tag{4.37}$$

$$E\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\eta_{N,i}\right) = \sqrt{N}E\eta_{N,1} \to 0. \tag{4.38}$$

We first establish (4.38). Since $\eta_{N,i} = \lambda'\xi_{N,i}$, the convergence (4.38) will follow from $\sqrt{N}E\xi_{N,1} \to 0$. By using (4.34) and $E(U_1|X_1, Z_1) = 0$, we get that

$$E\xi_{N,1} = \sqrt{h^m}E\big(((V(X_1)V(X_1)') \otimes I_G)\mathrm{vec}(\Theta(Z_1) - \Theta(z))K_h(z - Z_1)\big)$$

$$= \sqrt{h^m}\int_{\mathbb{R}^n}\int_{\mathbb{R}^m}(V(X_1)V(X_1)' \otimes I_G)\mathrm{vec}(\Theta(z_1) - \Theta(z))p(x_1, z_1)K_h(z - z_1))\, dz_1 dx_1.$$

By applying Proposition 4.1.1, we deduce that

$$E\xi_{N,1} = O\left(h^{(m+2r)/2}\right). \tag{4.39}$$

This shows in particular that $\sqrt{N}E\xi_{N,1} \to 0$. To prove the convergence (4.37), it is enough to show that $E\xi_{N,1}^2 \to W_0(z)$ where $\xi_{N,1}^2 = \xi_{N,1}\xi_{N,1}'$, and $N(E\xi_{N,1})^2 \to 0$ where $(E\xi_{N,1})^2 =$

$E\xi'_{N,1}E\xi_{N,1}$. By using $E(U_1|X_1, Z_1) = 0$, we obtain that

$$\begin{aligned} E\xi_{N,1}^2 &= \|K\|_2^2 E\left(\left(((V(X_1)V(X_1)') \otimes I_G)\mathrm{vec}(\Theta(Z_1) - \Theta(z))\right)^2 K_{2,h}(z - Z_1)\right) \\ &+ \|K\|_2^2 E\left(\left((V(X_1) \otimes I_G)U_1\right)^2 K_{2,h}(z - Z_1)\right) =: I_1 + I_2, \end{aligned}$$

where $K_{2,h}$ is the kernel function defined by (4.5). Arguing as in Lemma 4.3.1, and by using the formula $((A \otimes I_G)B)^2 = AA' \otimes BB'$ for a $d \times 1$ vector A and a $G \times 1$ vector B, we conclude that $I_1 = O(h^2)$ and $I_2 = W_0(z) + O(h^2)$. (Indeed, for the term I_2, by using that formula, we have

$$\begin{aligned} I_2 &= \|K\|_2^2 E\left((V(X_1)V(X_1)' \otimes U_1U_1')K_{2,h}(z - Z_1)\right) \\ &= \|K\|_2^2 E\left((V(X_1)V(X_1)' \otimes \Sigma)K_{2,h}(z - Z_1)\right) = W_0(z) + O(h^2).) \end{aligned}$$

This yields

$$E\xi_{N,1}^2 = W_0(z) + O(h^2) \tag{4.40}$$

and, in particular, $E\xi_{N,1}^2 \to W_0(z)$. Since $(E\xi_{N,1})^2 = O(h^{m+2r})$ by (4.39), we obtain that $N(E\xi_{N,1})^2 = O(Nh^{m+2r}) \to 0$ by the assumption (4.18).

We still need to show (4.36). Let us first find the order of the numerator in (4.36). By using $\eta_{N,i} = \lambda'\xi_{N,i}$, we have

$$E|\eta_{N,1} - E\eta_{N,1}|^4 \leq \mathrm{const}|\lambda|^4(E|\xi_{N,1}|^4 + |E\xi_{N,1}|^4). \tag{4.41}$$

Since we already have the order of $E\xi_{N,1}$ in (4.39), we need to consider only $E|\xi_{N,1}|^4$. By using the definition (4.34) of $\xi_{N,1}$, we get $E|\xi_{N,1}|^4 \leq \mathrm{const}(S_1 + S_2)$, where

$$\begin{aligned} S_1 &= h^{-m}E|V(X_1)V(X_1)'|^4|\mathrm{vec}(\Theta(Z_1) - \Theta(z))|^4 K_{4,h}(z - Z_1), \\ S_2 &= h^{-m}E|V(X_1)|^4|U_1|^4 K_{4,h}(z - Z_1) \end{aligned}$$

and $K_{4,h}(\cdot)$ is defined by (4.5). By using Proposition 4.1.1, we conclude that $S_1 = O(h^{r-m})$, $S_2 = O(h^{-m})$ and hence that $E|\xi_{N,1}|^4 = O(h^{-m})$. Together with $E\xi_{N,1} = O\left(h^{(m+2r)/2}\right)$ from (4.39) and also (4.41), this shows that

$$E|\eta_{N,1} - E\eta_{N,1}|^4 = O(h^{-m}).$$

Finally, by using (4.37) and (4.38), we conclude that the order of the term in (4.36) is $O((Nh^m)^{-1})$ and hence (4.36) is valid since $Nh^m \to \infty$. $\square$

PROOF OF THEOREM 4.2.3: Arguing as in the proof of Theorem 4.2.2, it is enough to show that

$$\sqrt{Nh^m}(\text{vec}(\Delta_1(z_1) + \Delta_2(z_1)), \ldots, \text{vec}(\Delta_1(z_q) + \Delta_2(z_q)))$$

$$\xrightarrow{d} (\mathcal{N}_1(0, W_0(z_1)), \ldots, \mathcal{N}_q(0, W_0(z_q))), \tag{4.42}$$

where $W_0(z) = (Q(z) \otimes \Sigma)\|K\|_2^2$ and $\Delta_1(z)$, $\Delta_2(z)$ are defined by (4.24) and (4.25). As in the proof of Theorem 4.2.2, for fixed $a_1, \ldots, a_q \in \mathbb{R}^{Gd}$, the random variable

$$a_1'\sqrt{Nh^m}\text{vec}(\Delta_1(z_1) + \Delta_2(z_1)) + \cdots + a_q'\sqrt{Nh^m}\text{vec}(\Delta_1(z_q) + \Delta_2(z_q))$$

is asymptotically normal with the variance given by

$$\lim Nh^m E(a_1'\text{vec}(\Delta_1(z_1) + \Delta_2(z_1)) + \cdots + a_q'\text{vec}(\Delta_1(z_q) + \Delta_2(z_q)))^2. \tag{4.43}$$

The result (4.42) will then follow from the corresponding one-dimensional result (4.33) valid for $z_1, \ldots, z_q$, as long as

$$\lim Nh^m E(\text{vec}(\Delta_1(z_i) + \Delta_2(z_i))\text{vec}(\Delta_1(z_j) + \Delta_2(z_j)))' = 0, \tag{4.44}$$

for $i \neq j$. To show (4.44), suppose for simplicity that $G = d = 1$. In view of (4.24) and (4.25), we have

$$E(\Delta_1(z_i) + \Delta_2(z_i))(\Delta_1(z_j) + \Delta_2(z_j)) = E\Delta_1(z_i)\Delta_1(z_j) + E\Delta_2(z_i)\Delta_2(z_j)$$

$$= \frac{1}{N}E(\Theta(Z_1) - \Theta(z_i))(\Theta(Z_1) - \Theta(z_j))V(X_1)^4 K_h(z_i - Z_1)K_h(z_j - Z_1)$$

$$+\frac{N-1}{N}(E(\Theta(Z_1) - \Theta(z_i))V(X_1)^2 K_h(z_i - Z_1))E(\Theta(Z_1) - \Theta(z_j))V(X_1)^2 K_h(z_j - Z_1))^2$$

$$+\frac{\Sigma}{N}E(V(X_1)^2 K_h(z_i - Z_1)K_h(z_j - Z_1) =: I_1 + I_2 + I_3.$$

Since the kernel K has compact support, we have $EF(X_1, Z_1)K_h(z_i - Z_1)K_h(z_j - Z_1) = 0$ for small enough h and any bounded function F. This implies that $I_1 = 0$ and $I_3 = 0$ for small enough h. As for the term I_2, one may show by using Proposition 4.1.1 that $I_2 = O(h^{2r})$. Then,

$$\lim Nh^m(I_1 + I_2 + I_3) = \lim Nh^m I_2 = 0,$$

since $Nh^{m+2r} \to 0$. □

Chapter 5

Local Tests

In this chapter, we consider local tests for (SPF) and (NP) models. Sections 5.1 and 5.2 are on hypothesis testing, namely, for fixed z and L, to test $H_0 : \mathrm{rk}\{\Theta(z)\} \le L$ against $H_1 : \mathrm{rk}\{\Theta(z)\} > L$ in (SPF) model and $H_0 : \mathrm{adrk}\{F(\cdot, z)\} \le L$ against $H_1 : \mathrm{adrk}\{F(\cdot, z)\} > L$ in (NP) model. Section 5.3 concerns estimation of the ranks $\mathrm{rk}\{\Theta(z)\}$ and $\mathrm{adrk}\{F(\cdot, z)\}$ themselves.

5.1 Local tests for semi-parametric model

In this section, we consider local tests for (SPF) model where the basic problem is to test, for some fixed z and $L \le \min\{G, d\}$, the hypothesis $H_0 : \mathrm{rk}\{\Theta(z)\} \le L$, against the alternative $H_1 : \mathrm{rk}\{\Theta(z)\} > L$. In Section 5.1.1, we recall the well-known LDU test and apply it to the matrix $\Theta(z)$. In Section 5.1.2, we turn to the so-called minimum-χ^2 test and, in Section 5.1.3, we explore its connections to the eigenvalues of some random matrices. Other tests available in the literature are discussed in Section 5.1.4.

Notation. We suppose throughout this section that z is fixed. To simplify the notation, we will denote the dependence on z by a subscript 0. For example, the matrix $\Theta(z)$ and its estimator $\widehat{\Theta}(z)$ will be denoted by Θ_0 and $\widehat{\Theta}_0$, respectively, the rank $L(z) = \mathrm{rk}\{\Theta(z)\}$ by L_0, the matrix $Q(z)$ by Q_0 and so on. In many other less important situations, we will suppress the dependence on z overall.

5.1.1 LDU based test

The LDU test for the rank of a matrix is based on the Lower-Diagonal-Upper triangular (LDU, in short) decomposition of a matrix with a complete pivoting. The LDU decomposition is achieved by a successive applications of the usual Gaussian elimination procedure. A complete pivoting means that at each step of the Gaussian elimination procedure the largest element in absolute magnitude is shifted to the top left corner by column and row interchanges. For a background or more information on the LDU decomposition and a complete pivoting, see for example Golub and Van Loan [39].

Supposing that the matrix of study is $\Theta_0 = \Theta(z)$ for fixed z, which is $G \times d$ and has an unknown rank $L_0 = L_0(z) \leq d = \min\{G, d\}$, the LDU decomposition with a complete pivoting allows to express Θ_0 in a factor form

$$A\Theta_0 B = \mathrm{LDU}. \tag{5.1}$$

Here, A and B, the so-called permutation matrices with dimensions $G \times G$ and $d \times d$, respectively, correspond to the column and row interchanges in a complete pivoting procedure. The other matrices L, D and U are of the forms

$$\mathrm{L} = \begin{pmatrix} \mathrm{L}_{11} & 0 & 0 \\ \mathrm{L}_{21} & \mathrm{L}_{22} & 0 \\ \mathrm{L}_{31} & \mathrm{L}_{32} & \mathrm{L}_{32} \end{pmatrix}, \quad \mathrm{U} = \begin{pmatrix} \mathrm{U}_{11} & \mathrm{U}_{22} \\ 0 & \mathrm{U}_{22} \\ 0 & 0 \end{pmatrix} \tag{5.2}$$

and

$$\mathrm{D} = \begin{pmatrix} \mathrm{D}_{11} & 0 & 0 \\ 0 & \mathrm{D}_{22} & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{5.3}$$

where the partitions are into L_0, $d - L_0$ and $G - d$ rows and columns except the matrix U, where the columns are partitioned into L_0 and $d - L_0$ columns. The submatrices L_{11} and U'_{11} are unit lower triangular matrices. The matrix D (and hence the submatrices D_{11} and D_{22}) is diagonal.

By using the above LDU decomposition, Gill and Lewbel [38] introduced a test for the rank of a matrix. Let $\widehat{\Theta}_0$ be the estimator of the matrix Θ_0 and let also $\widehat{\mathrm{L}}, \widehat{\mathrm{D}}$ (with the corresponding submatrices $\widehat{\mathrm{D}}_{11}$ and $\widehat{\mathrm{D}}_{22}$) and $\widehat{\mathrm{U}}$ be the matrices in the LDU decomposition of $\widehat{\Theta}_0$. The basic idea behind the test of Gill and Lewbel is that the matrix Θ_0 has rank L_0 if and only if the submatrix D_{22} in the LDU decomposition is identically zero and the elements of the diagonal matrix D_{11} are non-zero. Hence, in order to determine the rank of the matrix, one may test whether the matrix D_{22} is significantly different from zero. To do so, Gill and Lewbel [38] first stated that, when the matrix Θ_0 has rank L_0, $\widehat{\mathrm{D}}_{22}$ is asymptotically normal with a specified asymptotic variance-covariance matrix. Then, based on this asymptotic normality result, the authors constructed the usual Wald type χ^2-test to determine whether $\mathrm{D}_{22} = 0$. However, as observed by Cragg and Donald [19], the asymptotic normality result of Gill and Lewbel is incorrect, except for special cases. Cragg and Donald [19] studied an appropriate modifications of Gill and Lewbel test. We will now summarize their rank test and apply it to our matrix Θ_0.

Let $\Theta_0(L)$ be the matrix obtained from the matrix Θ_0 after L steps of Gaussian elimination procedure with complete pivoting. More precisely, at each step $i = 1, \ldots, L$, the rows and columns of the matrix $\Theta_0(i-1)$ (with $\Theta_0(0) = \Theta_0$) are permuted according to complete pivoting and then the Gaussian elimination procedure is applied to the ith column making the elements in rows $i+1, \ldots, G$ zero. The matrix $\Theta_0(L)$ can then be expressed as

$$\Theta_0(L) = \begin{pmatrix} \Omega_{11}(L) & \Omega_{12}(L) \\ 0 & \Omega_{22}(L) \end{pmatrix}, \tag{5.4}$$

where $\Omega_{11}(L)$ is a $L \times L$ upper triangular matrix, and $\Omega_{12}(L)$ and $\Omega_{22}(L)$ are $L \times (d-L)$

and $(G-L)\times(d-L)$ matrices, respectively. As shown in Cragg and Donald [19] (to get a feeling for this result, consider the case $G=d=2$ and $\mathrm{rk}\{\Theta_0\}=1$), one has:

Lemma 5.1.1 *The matrix Θ_0 has rank L_0 if and only if $\Omega_{22}(L_0)=0$.*

Let now $\widehat{\Theta}_0(L)$ be the matrix $\widehat{\Theta}_0$ after L steps of Gaussian elimination procedure with complete pivoting as in the case of Θ_0, and $\widehat{\Omega}_{ij}(L)$ be the submatrices in the corresponding representation (5.4). As proved in Cragg and Donald [19], when $L=\mathrm{rk}\{\Theta_0\}$, the matrix $\widehat{\Omega}_{22}(L_0)$ is asymptotically normal.

Theorem 5.1.1 (Cragg and Donald) *Under the assumptions of Theorem 4.2.2, when $L=\mathrm{rk}\{\Theta_0\}$,*

$$\sqrt{Nh^m}\mathrm{vec}(\widehat{\Omega}_{22}(L)) \xrightarrow{d} \mathcal{N}(0,\Pi(L)W_0\Pi(L)'), \tag{5.5}$$

where $W_0=(Q_0^{-1}\otimes\Sigma)\|K\|_2^2$ is the limit variance-covariance matrix in Theorem 4.2.2 and, with matrices Θ_{ij}, $i,j=1,2$, defined below,

$$\Pi(L)=(-\Theta_{21}(L)\Theta_{11}(L)^{-1}\quad I_{G-L})\otimes(-\Theta_{12}(L)\Theta_{11}(L)^{-1}\quad I_{d-L}). \tag{5.6}$$

In the case when a complete pivoting is not used to obtain the matrix $\Theta_0(L)$, the matrices Θ_{11},Θ_{12}, Θ_{21} and Θ_{22} in (5.6) are $L\times L$, $L\times(d-L)$, $(G-L)\times L$ and $(G-L)\times(d-L)$, respectively, and appear in the following partition of the matrix Θ_0,

$$\Theta_0=\begin{pmatrix}\Theta_{11} & \Theta_{12}\\ \Theta_{21} & \Theta_{22}\end{pmatrix}. \tag{5.7}$$

In the case when a complete pivoting is necessary to obtain the matrix $\Theta_0(L)$, the rows and columns of Θ_0 can be permuted in advance so that a complete pivoting becomes unnecessary. (One needs however to take these permutations into account for the limit variance-covariance matrix.)

Let now $\widehat{\Theta}_{ij}$ be the corresponding estimators of submatrices Θ_{ij} and $\widehat{W_0}$ be the estimator of W_0. Consider the Wald type test statistic for the matrix $\Omega_{22}(L)$, namely,

$$\widehat{\xi}(\widehat{\Theta}_0, L) = Nh^m \mathrm{vec}(\widehat{\Omega}_{22}(L))' \left(\widehat{\Pi}(L)\widehat{W_0}\widehat{\Pi}(L)'\right)^{-1} \mathrm{vec}(\widehat{\Omega}_{22}(L)), \tag{5.8}$$

where $\widehat{\Pi}(L)$ is defined as in (5.6) by using the matrices $\widehat{\Theta}_{ij}$. The following result, which follows immediately from Theorem 5.1.1 and the discussion above, can be used to test the hypothesis $H_0 : \mathrm{rk}\{\Theta_0\} \leq L$ against the alternative $H_1 : \mathrm{rk}\{\Theta_0\} > L$. The notation $\chi^2(k)$ stands for the chi-square distribution with k degrees of freedom.

Theorem 5.1.2 *Under the assumptions of Theorem 4.2.2, (i) when $L < \mathrm{rk}\{\Theta_0\}$, we have $\widehat{\xi}(\widehat{\Theta}_0, L) \to_p \infty$ and (ii) when $L = \mathrm{rk}\{\Theta_0\}$, we have $\widehat{\xi}(\widehat{\Theta}_0, L) \to_d \chi^2((G-L)(d-L))$.*

5.1.2 Minimum-χ^2 test

The minimum-χ^2 test for the rank of a matrix was introduced and developed by Cragg and Donald [18, 19, 20]. Supposing that the matrix of interest is $\Theta_0 = \Theta(z)$ for fixed z, the minimum-χ^2 test is based on the following statistic.

Definition 5.1.1 *(Minimum-χ^2 statistic) Let*

$$\begin{aligned} \widehat{C}(\widehat{\Theta}_0, L) &= Nh^m \min_{\mathrm{rk}\{\Theta\} \leq L} \mathrm{vec}(\widehat{\Theta}_0 - \Theta)' \widehat{W}_0^{-1} \mathrm{vec}(\widehat{\Theta}_0 - \Theta) \\ &= Nh^m \|K\|_2^{-2} \min_{\mathrm{rk}\{\Theta\} \leq L} \mathrm{vec}(\widehat{\Theta}_0 - \Theta)'(\widehat{Q}_0^{-1} \otimes \widehat{\Sigma})^{-1} \mathrm{vec}(\widehat{\Theta}_0 - \Theta). \end{aligned} \tag{5.9}$$

The matrix $\widehat{W_0} = (\widehat{Q}_0^{-1} \otimes \widehat{\Sigma})\|K\|_2^2$ in (5.9) above is an estimator for the variance-covariance matrix $W_0 = (Q_0^{-1} \otimes \Sigma)\|K\|_2^2$ appearing in the asymptotic normality result for $\widehat{\Theta}_0$ (see Theorem 4.2.2). The term "minimum chi-square" used by Cramér [21] refers to a minimization of an expression of the type $g'Wg$. (See also Ferguson [33] and Rothenberg [90].)

Remark 5.1.1 The minimum-χ^2 statistic for the rank of a matrix can be obtained by using a standard generalized method of moments (GMM). The GMM approach introduced by Hansen [47], is based on the fact that a statistical model implies a set of orthogonality conditions. The GMM estimator is deduced from these orthogonality conditions.

The minimum-χ^2 test for the rank of the matrix Θ_0 is based on the following result which provides asymptotics of the test statistic $\widehat{C}(\widehat{\Theta}_0, L)$ in the case when $L < \text{rk}\{\Theta_0\}$, $L = \text{rk}\{\Theta_0\}$ and $L > \text{rk}\{\Theta_0\}$. We will prove the asymptotics for the first two cases by adapting and providing more details in some proof of Cragg and Donald [18, 20]. The asymptotics for the third case can be proved as in Theorem 1 of Cragg and Donald [20]. We omit its proof because, in the next section, we will prove a stronger result. (Recall also that a stochastic dominance $\xi \leq_d \eta$ means $P(\xi > x) \leq P(\eta > x)$ for all $x \in \mathbb{R}$.)

Theorem 5.1.3 (Cragg and Donald) *Let $L_0 = \text{rk}\{\Theta_0\}$. Then, under the assumptions of Theorem 4.2.2, we have (i) when $L < L_0$, $\widehat{C}(\widehat{\Theta}_0, L) \to_p \infty$, (ii) when $L = L_0$, $\widehat{C}(\widehat{\Theta}_0, L) \to_d \chi^2((G - L_0)(d - L_0))$, (iii) when $L > L_0$, $\widehat{C}(\widehat{\Theta}_0, L) \to_d \xi_0$, where $\xi_0 \leq_d \chi^2((G - L)(d - L))$.*

Proof: Part (i) follows since

$$\widehat{C}(\widehat{\Theta}_0, L)(Nh^m)^{-1} \xrightarrow{p} \|K\|_2^{-2} \min_{\text{rk}\{\Theta\} \leq L} \text{vec}(\Theta_0 - \Theta)'(Q_0^{-1} \otimes \Sigma)^{-1}\text{vec}(\Theta_0 - \Theta) > 0$$

for $L < L_0$ and $Nh^m \to \infty$. Consider now part (ii). Restriction $\text{rk}\{\Theta\} \leq L$ in (5.9) can be expressed as

$$\Theta = (\Theta_1 \ \ \Theta_1 \Xi_1), \tag{5.10}$$

where Θ_1 and Ξ_1 are any $G \times L$ and $L \times (d - L)$ matrices, respectively. This shows that there are $GL + L(d - L) =: s$ free parameters $\mu = \text{vec}(\Theta_1, \Xi_1)$ and hence

$$\widehat{C}(\widehat{\Theta}_0, L) = Nh^m \min_{\mu} \text{vec}(\widehat{\Theta}_0 - \Theta(\mu))'\widehat{W}_0^{-1}\text{vec}(\widehat{\Theta}_0 - \Theta(\mu)), \tag{5.11}$$

where $\Theta(\mu) = (\Theta_1 \quad \Theta_1\Xi_1)$. Let now $B(\mu)$ be a $Gd \times s$ matrix defined by $B(\mu) = \partial\mathrm{vec}(\Theta(\mu))/\partial\mu$. One can obtain from (5.10) that

$$B(\mu) = \begin{pmatrix} I_{GL} & 0_{GL\times L(d-L)} \\ \Xi_1' \otimes I_G & I_{d-L} \otimes \Theta_1 \end{pmatrix}. \tag{5.12}$$

Since $\mathrm{rk}\{\Theta_0\} = L$, we have $\Theta_0 = \Theta(\mu_0)$ for some μ_0 and, moreover, the corresponding submatrix Θ_1 has full column rank. In view of (5.12), we obtain that the matrix $B(\mu_0)$ is of full column rank, that is, $\mathrm{rk}\{B(\mu_0)\} = s$.

Let $\widehat{\mu}$ be μ minimizing the expression on the right-hand side of (5.11), that is,

$$\widehat{C}(\widehat{\Theta}_0, L) = Nh^m \mathrm{vec}(\widehat{\Theta}_0 - \Theta(\widehat{\mu}))'\widehat{W}_0^{-1}\mathrm{vec}(\widehat{\Theta}_0 - \Theta(\widehat{\mu})). \tag{5.13}$$

We have $\widehat{\mu} \to \mu_0$ in probability. Observe now that, by using the Taylor expansion and $\widehat{\mu} \to_p \mu_0$, we have

$$\mathrm{vec}(\widehat{\Theta}_0 - \Theta(\widehat{\mu})) = \mathrm{vec}(\widehat{\Theta}_0 - \Theta(\mu_0)) - B(\mu_0)(\widehat{\mu} - \mu_0) + o_p(1). \tag{5.14}$$

The first order conditions for minimizing (5.11), together with (5.14), imply that

$$\begin{aligned} 0 &= B(\widehat{\mu})'\widehat{W}_0^{-1}\mathrm{vec}(\widehat{\Theta}_0 - \Theta(\widehat{\mu})) \\ &= B(\mu_0)'\widehat{W}_0^{-1}\mathrm{vec}(\widehat{\Theta}_0 - \Theta(\mu_0)) - B(\mu_0)'\widehat{W}_0^{-1}B(\mu_0)(\widehat{\mu} - \mu_0) + o_p(1) \end{aligned}$$

and hence that

$$\widehat{\mu} - \mu_0 = (B(\mu_0)'\widehat{W}_0^{-1}B(\mu_0))^{-1}B(\mu_0)'\widehat{W}_0^{-1}\mathrm{vec}(\widehat{\Theta} - \Theta(\mu_0)) + o_p(1). \tag{5.15}$$

By substituting (5.15) into (5.14) and then (5.14) into (5.13), we get that

$$\widehat{C}(\widehat{\Theta}_0, L) = Nh^m \mathrm{vec}(\widehat{\Theta}_0 - \Theta(\mu_0))'\widehat{W}_0^{-1/2}.$$

$$\cdot \left(I_{Gd} - \widehat{W}_0^{-1/2} B(\mu_0)(B(\mu_0)'\widehat{W}_0^{-1}\, B(\mu_0))^{-1} B(\mu_0)'\widehat{W}_0^{-1/2} \right) \widehat{W}_0^{-1/2} \mathrm{vec}(\widehat{\Theta}_0 - \Theta(\mu_0)) + o_p(1)$$

$$= Nh^m \mathrm{vec}(\widehat{\Theta}_0 - \Theta(\mu_0))' W_0^{-1/2} (I_{Gd} - A_0(A_0'A_0)^{-1}A_0') W_0^{-1/2} \mathrm{vec}(\widehat{\Theta}_0 - \Theta(\mu_0)) + o_p(1),$$

where $A_0 = W_0^{-1/2} B(\mu_0)$. By Theorem 4.2.2 and since the matrix $B(\mu_0)$ or the matrix A_0 has full column rank, we obtain that

$$\widehat{C}(\widehat{\Theta}_0, L) \xrightarrow{d} \chi^2(Gd - \mathrm{rk}\{B(\mu_0)\}) = \chi^2((G-L)(d-L)). \quad \square$$

5.1.3 Connection to eigenvalues

In this section, we relate the minimum-χ^2 statistic $\widehat{C}(\widehat{\Theta}_0, L)$ to the eigenvalues of some random matrices. This connection will allow us to state an asymptotic result for the test statistic $\widehat{C}(\widehat{\Theta}_0, L)$ which is more accurate than that in part (*iii*) of Theorem 5.1.3.

Theorem 5.1.4 *(Connection of minimum-χ^2 statistic to eigenvalues) We have*

$$\widehat{C}(\widehat{\Theta}_0, L) = Nh^m \|K\|_2^{-2} \sum_{i=1}^{G-L} \widehat{\lambda}_i, \tag{5.16}$$

where $0 \le \widehat{\lambda}_1 \le \widehat{\lambda}_2 \le \cdots \le \widehat{\lambda}_G$ *are the eigenvalues of the matrix*

$$\widehat{\Gamma}_0 = \widehat{\Theta}_0 \widehat{Q}_0 \widehat{\Theta}_0' \widehat{\Sigma}^{-1}. \tag{5.17}$$

PROOF: The proof uses some ideas of the proof of Theorem 3 in Cragg and Donald [18]. For notational simplicity, we will omit the variable z in the proof. The restriction $\mathrm{rk}\{\Theta\} \le L$ can be written as

$$\Xi\Theta = 0, \tag{5.18}$$

where Ξ is a $(G-L)\times G$ matrix with its $G-L$ rows linearly independent. Moreover, after a proper normalization, we can assume that Ξ satisfies

$$\Xi\widehat{\Sigma}\Xi' = I_{G-L}. \tag{5.19}$$

Now, the restriction (5.18) can be written as $\text{vec}(\Xi\Theta) = 0$ or, by using the formula $\text{vec}(AB) = (I_p \otimes A)\text{vec}(B)$ for a $m \times n$ matrix A and a $n \times p$ matrix B, as

$$(I_d \otimes \Xi)\text{vec}(\Theta) = 0. \tag{5.20}$$

When Ξ is fixed, after a simple manipulation with Lagrange multipliers, the minimum value of the function $\text{vec}(\widehat{\Theta}_0 - \Theta)'(\widehat{Q}_0^{-1} \otimes \widehat{\Sigma})^{-1}\text{vec}(\widehat{\Theta}_0 - \Theta)$ under the linear constraints (5.20) on Θ, can be expressed as

$$\mathcal{F} = \big((I_d \otimes \Xi)\text{vec}(\widehat{\Theta}_0)\big)'\big((I_d \otimes \Xi)(\widehat{Q}_0^{-1} \otimes \widehat{\Sigma})(I_d \otimes \Xi)'\big)^{-1}\big((I_d \otimes \Xi)\text{vec}(\widehat{\Theta}_0)\big).$$

By using $(I_d \otimes \Xi)\text{vec}(\widehat{\Theta}_0) = \text{vec}(\Xi\widehat{\Theta}_0)$, the formulas $(A \otimes B)(C \otimes D) = AC \otimes BD$ and $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$, and also the condition (5.19), we can simplify $\mathcal{F}$ as

$$\mathcal{F} = \text{vec}(\Xi\widehat{\Theta}_0)'(\widehat{Q}_0 \otimes I_{G-L})\text{vec}(\Xi\widehat{\Theta}_0).$$

By using the formula $\text{tr}\{ABCD\} = (\text{vec}(D'))'(C' \otimes A)\text{vec}(B)$ where $\text{tr}\{\cdot\}$ denotes the trace of a matrix (see, for example, Theorem 3 on p. 31 in Magnus and Neudecker [73]), we can further rewrite $\mathcal{F}$ as

$$\mathcal{F} = \text{tr}\{\Xi\widehat{\Theta}_0\widehat{Q}_0\widehat{\Theta}_0'\Xi'\}.$$

The minimum-χ^2 statistic $\widehat{C}(\widehat{\Theta}_0, L)$ can then be obtained by minimizing $Nh^m\|K\|_2^{-2}\mathcal{F}$ under the constraint (5.19) on Ξ, that is,

$$\begin{aligned}
\widehat{C}(\widehat{\Theta}_0, L) &= Nh^m\|K\|_2^{-2} \min_{\Xi\widehat{\Sigma}\Xi'=I_{G-L}} \operatorname{tr}\{\Xi\widehat{\Theta}_0\widehat{Q}_0\widehat{\Theta}_0'\Xi'\} \\
&= Nh^m\|K\|_2^{-2} \min_{X'X=I_{G-L}} \operatorname{tr}\{X'\widehat{\Sigma}^{-1/2}\widehat{\Theta}_0\widehat{Q}_0\widehat{\Theta}_0'\widehat{\Sigma}^{-1/2}X\},
\end{aligned}$$

where, in the last step, we made the change of variables $X' = \Xi\widehat{\Sigma}^{1/2}$. Finally, by using the formula

$$\min_{X'X=I_k} \operatorname{tr}\{X'AX\} = \sum_{i=1}^{k} \lambda_i,$$

where $\lambda_1 \leq \cdots \leq \lambda_n$ are the eigenvalues of a $n \times n$ matrix A (see, for example, Theorem 13 on p. 211 in Magnus and Neudecker [73]), we conclude that

$$\widehat{C}(\widehat{\Theta}_0, L) = Nh^m\|K\|_2^{-2} \sum_{i=1}^{G-L} \widehat{\lambda}_i,$$

where $0 \leq \widehat{\lambda}_1 \leq \cdots \leq \widehat{\lambda}_G$ are the eigenvalues of the matrix $\widehat{\Sigma}^{-1/2}\widehat{\Theta}_0\widehat{Q}_0\widehat{\Theta}_0'\widehat{\Sigma}^{-1/2}$. (The eigenvalues $\widehat{\lambda}_i$ are all positive, since the matrix $\widehat{\Sigma}^{-1/2}\widehat{\Theta}_0\widehat{Q}_0\widehat{\Theta}_0'\widehat{\Sigma}^{-1/2}$ is semi-positive definite.) It is easy to see that $\widehat{\lambda}_i$, $i = 1, \ldots, G$, are also the eigenvalues of the matrix $\widehat{\Theta}_0\widehat{Q}_0\widehat{\Theta}_0\widehat{\Sigma}^{-1}$, which yields the result. □

The matrix $\widehat{\Gamma}_0$ in (5.17) is a consistent estimator for the matrix

$$\Gamma_0 = \Theta_0 Q_0 \Theta_0' \Sigma^{-1} =: \Upsilon_0 \Sigma^{-1}. \tag{5.21}$$

Its use for rank tests can be clarified by the following elementary lemma.

Lemma 5.1.2 *The matrix Θ_0 has rank L_0 if and only if the matrix $\Upsilon_0 = \Theta_0 Q_0 \Theta_0'$ has $G - L_0$ zero eigenvalues, or if and only if the matrix $\Gamma_0 = \Upsilon_0\Sigma^{-1}$ has $G - L_0$ zero eigenvalues.*

PROOF: Lemma follows from the following equivalence relations: $\mathrm{rk}\{\Theta_0\} = L_0$ if and only if $\mathrm{rk}\{\Theta_0 Q_0^{1/2}\} = L_0$ if and only if there are $G - L_0$ linearly independent vectors c_j, $j = 1, \ldots, G - L_0$, such that $c_j' \Theta_0 Q_0^{1/2} = 0$ if and only if $|c_j' \Theta_0 Q_0^{1/2}|^2 = c_j' \Theta_0 Q_0 \Theta_0' c_j = c_j' \Upsilon_0 c_j = 0$, $j = 1, \ldots, G - L_0$, if and only if the matrix Υ_0 has $G - L_0$ eigenvalues equal to 0 (or 0 is the eigenvalue of Υ_0 with the multiplicity $G - L_0$) if and only if the matrix $\Gamma_0 = \Upsilon_0 \Sigma^{-1}$ has $G - L_0$ eigenvalues equal to 0. □

The following lemma will be used to improve on Theorem 5.1.3. It follows directly from Theorem 3.1 in Robin and Smith [89] and Theorem 4.2.2 above.

Lemma 5.1.3 (Robin and Smith) *Under the assumptions of Theorem 4.2.2, the normalized eigenvalues $Nh^m \|K\|_2^{-2} \widehat{\lambda}_i$, $i = 1, \ldots, G - L_0$, of the matrix $\widehat{\Gamma}_0$ in (5.17) have the same limiting distribution as the ordered eigenvalues of the matrix*

$$Nh^m C'_{G-L_0} (\widehat{\Theta}_0 - \Theta_0) D_{d-L_0} D'_{d-L_0} (\widehat{\Theta}_0 - \Theta_0)' C_{G-L_0}, \tag{5.22}$$

where a $G \times (G - L_0)$ matrix C_{G-L_0} and a $d \times (d - L_0)$ matrix D_{d-L_0} are defined below.

To define the matrices C_{G-L_0} and D_{d-L_0} in Lemma 5.1.3, let c_i, $i = 1, \ldots, G - L_0$, be linearly independent eigenvectors corresponding to the $G - L_0$ zero eigenvalues of the matrix $\Theta_0 Q_0 \Theta_0' \Sigma^{-1}$ (see Lemma 5.1.2 above) and let d_i, $i = 1, \ldots, d - L_0$, be linearly independent eigenvectors corresponding to the $d - L_0$ zero eigenvalues of the matrix $\Theta_0' \Sigma^{-1} \Theta_0 Q_0$ (the number of zero eigenvalues can be obtained as in the proof of Lemma 5.1.2). One may suppose after a proper normalization that $c_i' \Sigma c_j = \delta_{ij}$, $i, j = 1, \ldots, G - L_0$, and $d_i' Q_0^{-1} d_j = \delta_{ij}$, $i, j = 1, \ldots, d - L_0$, where $\delta_{ij} = 0$ if $i \neq j$, and $\delta_{ij} = 1$ if $i = j$. Then, the matrices C_{G-L_0} and D_{d-L_0} are defined as

$$C_{G-L_0} = (c_1, \ldots, c_{G-L_0}), \quad D_{d-L_0} = (d_1, \ldots, d_{d-L_0}), \tag{5.23}$$

and they are normalized by

$$C'_{G-L_0}\Sigma C_{G-L_0} = I_{G-L_0}, \quad D'_{d-L_0}Q_0^{-1}D_{d-L_0} = I_{d-L_0}. \tag{5.24}$$

We can now state and prove an asymptotic result for the minimum-χ^2 statistic which improves on part (*iii*) of Theorem 5.1.3. Let $\mathcal{Y}_{n\times m}$, $n, m \geq 1$, be a $n \times m$ matrix with independent $\mathcal{N}(0,1)$ entries and let $\lambda_1(\mathcal{Y}^2_{n\times m}) \leq \cdots \leq \lambda_n(\mathcal{Y}^2_{n\times m})$ be the eigenvalues of the matrix $\mathcal{Y}^2_{n\times m} = \mathcal{Y}_{n\times m}\mathcal{Y}'_{n\times m}$.

Theorem 5.1.5 *Under the assumptions of Theorem 4.2.2, when $L \geq \mathrm{rk}\{\Theta_0\} = L_0$,*

$$\widehat{C}(\widehat{\Theta}_0, L) \xrightarrow{d} \sum_{i=1}^{G-L} \lambda_i(\mathcal{Y}^2_{(G-L_0)\times(d-L_0)}) \overset{d}{\leq} \chi^2((G-L)(d-L)), \tag{5.25}$$

where the stochastic dominance in (5.25) is, in fact, equality in distribution for $L = L_0$.

Remark 5.1.2 Observe that, by (5.25), $\lim P(\widehat{C}(\widehat{\Theta}_0, L) \geq x) \leq P(\chi^2((G-L)(d-L)) \geq x)$ for all x. This relation can be used in practice to choose the critical value for the test statistic $\widehat{C}(\widehat{\Theta}_0, L)$.

Proof: Let $\widehat{Y}_0 = \sqrt{Nh^m}C'_{G-L_0}(\widehat{\Theta}_0 - \Theta_0)D_{d-L_0}$ so that the matrix in (5.22) can be expressed as $\widehat{Y}_0\widehat{Y}_0'$. It follows from Theorem 4.2.2 that

$$\begin{aligned}
\mathrm{vec}(\widehat{Y}_0) &\xrightarrow{d} \mathcal{N}(0, (D'_{d-L_0} \otimes C'_{G-L_0})W_0(D_{d-L_0} \otimes C_{G-L_0})) \\
&= \mathcal{N}(0, (D'_{d-L_0}Q_0^{-1}D_{d-L_0}) \otimes (C'_{G-L_0}\Sigma C_{G-L_0})) \\
&= \mathcal{N}(0, I_{d-L_0} \otimes I_{G-L_0}) = \mathcal{N}(0, I_{(d-L_0)(G-L_0)}),
\end{aligned} \tag{5.26}$$

where in the last two equalities we used the expression (4.19) for W_0 and also the two relations in (5.24). The convergence (5.26) shows that $\widehat{Y}_0 \xrightarrow{d} \mathcal{Y}_{(G-L_0)\times(d-L_0)}$. Hence, by Lemma 5.1.3 above and the continuous mapping theorem we obtain that the convergence in (5.25) holds.

We still need to establish a dominance result in (5.25). To do so, we will use some ideas from the proof of Theorems 1 and 2 in Donald [28]. By using the Poincaré separation theorem (see, for example, Magnus and Neudecker [73], p. 209, or Rao [84], p. 65), we obtain that $\lambda_i(\mathcal{Y}_{(G-L_0)\times(d-L_0)}\mathcal{Y}'_{(G-L_0)\times(d-L_0)}) \leq \lambda_i(B'\mathcal{Y}_{(G-L_0)\times(d-L_0)}\mathcal{Y}'_{(G-L_0)\times(d-L_0)}B)$ for $i = 1,\ldots,G-L$, and any $(G-L_0)\times(G-L)$ matrix B such that $B'B = I_{G-L}$. Now take $B = (0_{(G-L)\times(L-L_0)}\ \ I_{G-L})'$ and note that $B'B = I_{G-L}$. Observe also that $B'\mathcal{Y}_{(G-L_0)\times(d-L_0)} =_d \mathcal{Y}_{(G-L)\times(d-L_0)}$ and hence

$$\sum_{i=1}^{G-L} \lambda_i(\mathcal{Y}^2_{(G-L_0)\times(d-L_0)}) \overset{d}{\leq} \sum_{i=1}^{G-L} \lambda_i(\mathcal{Y}^2_{(G-L)\times(d-L_0)}). \tag{5.27}$$

Since, for $i = 1,\ldots,G-L$, we have

$$\begin{aligned} \lambda_i(\mathcal{Y}^2_{(G-L)\times(d-L_0)}) &= \lambda_i((\mathcal{Y}'_{(G-L)\times(d-L_0)})^2) \overset{d}{=} \lambda_i(\mathcal{Y}^2_{(d-L_0)\times(G-L)}) \\ &\overset{d}{\leq} \lambda_i(\mathcal{Y}^2_{(d-L)\times(G-L)}) \overset{d}{=} \lambda_i(\mathcal{Y}^2_{(G-L)\times(d-L)}), \end{aligned}$$

where the last stochastic dominance is obtained by using the same arguments as to get (5.27), it follows that

$$\sum_{i=1}^{G-L} \lambda_i(\mathcal{Y}^2_{(G-L_0)\times(d-L_0)}) \overset{d}{\leq} \sum_{i=1}^{G-L} \lambda_i(\mathcal{Y}^2_{(G-L)\times(d-L)}). \tag{5.28}$$

Finally, by using the formula $\sum_{i=1}^{n} \lambda_i = \text{tr}\{A\}$, where A is a $n \times n$ matrix and λ_i, $i = 1,\ldots,n$, are its eigenvalues, we obtain that

$$\begin{aligned} \sum_{i=1}^{n} \lambda_i(\mathcal{Y}^2_{(G-L)\times(d-L)}) &= \text{tr}(\mathcal{Y}^2_{(G-L)\times(d-L)}) \\ &= \text{vec}(\mathcal{Y}_{(G-L)\times(d-L)})'\text{vec}(\mathcal{Y}_{(G-L)\times(d-L)}) \\ &\overset{d}{=} \chi^2((G-L)(d-L)), \end{aligned}$$

which together with (5.28), yields the stochastic dominance result in (5.25). □

Remark 5.1.3 Anderson [8, 7, 9] (one should also mention Hsu [53]) was the first author to study test statistics for the rank of a matrix based on eigenvalues of some random matrices. In the situation considered by Anderson, the matrix Θ_0 of an unknown rank is a regression coefficient matrix in a multivariate linear model, where the noise variables are assumed to be *normally* distributed. Supposing, as in (4.19), that the limiting variance-covariance for $\widehat{\Theta}_0$ has the Kronecker product form $W_0 = Q_0 \otimes \Sigma^{-1}$, Anderson has also found that, under the null hypothesis $\text{rk}\{\Theta_0\} = L$, the limit of properly normalized test statistic (5.16) is a $\chi^2((G-L)(d-L))$ random variable. (Hence, this result is a special case of Theorems 5.1.3 and 5.1.5.) The assumption of normality found in Anderson [8, 7, 9] allows for techniques inherent to normal distributions which cannot be used in more general settings (e.g. in the setting of this thesis). Despite this restriction on distributional properties of the underlying noise variables, the work by Anderson played a major guiding role in later developments related to rank tests of a matrix.

Remark 5.1.4 The minimum-χ^2 test statistic, when expressed in terms of estimated eigenvalues is seen to be a special case of rank tests developed by Robin and Smith [89]. As in (5.16) and (5.17), let $\widehat{\lambda}_i$ be estimated eigenvalues of the matrix $\widehat{\Theta}_0\widehat{Q}_0\widehat{\Theta}_0'\widehat{\Sigma}^{-1}$, where $\widehat{Q}_0$ and $\widehat{\Sigma}$ are consistent estimators of some matrices Q_0 and Σ. In contrast to (4.19), the matrices Q_0 and Σ in Robin and Smith [89] are not necessarily taken as Kronecker product factors for the asymptotic variance-covariance matrix W_0 of $\widehat{\Theta}_0$. They can be chosen to suit one's interests which depend on the situation at hand. Under some standard assumptions (like asymptotic normality of Θ_0), Robin and Smith [89] found asymptotic limits of some functionals $h(z)$ (e.g. $h(z) = z$ as in (5.16) or $h(z) = \log(1+z)$ related to likelihood ratios) of estimated eigenvalues $\widehat{\lambda}_i$. The limiting distribution turns out to be a weighted sum of $\chi^2(1)$ random variables, where the weights are eigenvalues of some matrix involving W_0 and also C_{G-L_0} and D_{d-L_0}, which are defined after Lemma 5.1.3. The major departure from earlier works is that the limit variance-covariance matrix W_0 need not to be assumed of a full rank. Robin and Smith [89] subsequently used their asymptotic results to construct and apply tests for the rank of a matrix. The matrices W_0, C_{G-L_0} and D_{d-L_0}, which appear in the characterizing limiting distribution, are replaced in practice by their empirical counterparts.

5.1.4 Other tests

In this section, we briefly describe two other tests for the rank of a matrix available in the literature, namely, the asymptotic least squares test and the test based on a singular value decomposition.

Asymptotic least squares test. Gourieroux, Monfort and Trognon [44] and Chamberlain [16, 17] showed that many estimation and hypothesis testing problems in statistics can be formulated in terms of a set of relations $f(\alpha, \beta) = 0$ between p parameters of interest α and q auxiliary parameters β for which $\sqrt{n}(\widehat{\beta}_n - \beta_0) \to_d \mathcal{N}(0, J_0)$. The related estimator of α, called the asymptotic least squares estimator (ALS estimator, in short), is then defined as

$$\widehat{\alpha} = \operatorname*{argmin}_{\alpha} f(\alpha, \widehat{\beta}_n)\, \widehat{S}\, f(\alpha, \widehat{\beta}_n),$$

where $\widehat{S}$ is a consistent estimator of

$$S = \left(\frac{\partial f}{\partial \beta'} J_0 \frac{\partial f'}{\partial \beta} \right)^{-1}.$$

Under suitable conditions, the ALS estimator is strongly consistent, asymptotically normal and is asymptotically equivalent to some other estimators used in statistics. The hypothesis testing problem related to the asymptotic model $f(\alpha, \beta) = 0$ is that of testing the hypothesis $H_0 : \exists\, \alpha : f(\alpha, \beta_0) = 0$. The corresponding test statistic, called the ALS statistic, is defined by

$$\widehat{C}_{\mathrm{ALS}} = f(\widehat{\alpha}, \widehat{\beta}_n)\, \widehat{S}\, f(\widehat{\alpha}, \widehat{\beta}_n).$$

Under the null hypothesis and some suitable conditions, the limit of the ALS statistic is a χ^2-distribution. For more information on ALS estimators and related hypothesis testing, see the papers mentioned above and also Gourieroux and Monfort [42] and Gourieroux, Monfort and Renault [43].

The ALS procedure was applied to testing and estimation of rank of a matrix by Robin and Smith [88]. Supposing that the matrix of interest is our $G \times d$ matrix Θ_0, let $\Theta_0(L)$ be the matrix obtained from Θ_0 as in Section 5.1.1 after L steps of Gaussian elimination procedure with complete pivoting. Then,

$$\Theta_0(L) = \begin{pmatrix} \Omega_{11}(L) & \Omega_{12}(L) \\ 0 & \Omega_{22}(L) \end{pmatrix} = (\Omega_1(L) \;\; \Omega_2(L)),$$

where $\Omega_1(L) = (\Omega_{11}(L) \;\; 0)'$ and $\Omega_2(L) = (\Omega_{12}(L) \;\; \Omega_{22}(L))'$. Observe now that the hypothesis $H_0 : \mathrm{rk}\{\Theta_0\} \le L$ can be rewritten as H_0 : there is Ξ_1 such that $\Omega_2(L) = \Omega_1(L)\Xi_1$ or, by setting

$$f(\Xi_1, \Omega_1(L), \Omega_2(L)) = \Omega_2(L) - \Omega_1(L)\Xi_1, \tag{5.29}$$

as $H_0 : \exists\, \Xi_1 : f(\Xi_1, \Omega_1(L), \Omega_2(L)) = 0$. The relation (5.29), when equated to 0, is then asymptotic model in the ALS procedure for rank testing, where Ξ_1 are the parameters of interest and $\Omega_1(L)$, $\Omega_2(L)$ are the auxiliary parameters. The corresponding ALS estimator $\widehat{\Xi}_1$ is defined as Ξ_1 minimizing

$$Nh^m \mathrm{vec}(\widehat{\Omega}_2(L) - \widehat{\Omega}_1(L)\Xi_1)' S\, \mathrm{vec}(\widehat{\Omega}_2(L) - \widehat{\Omega}_1(L)\Xi_1), \tag{5.30}$$

where

$$\begin{aligned} S &= \left(\frac{\partial \mathrm{vec}(\Omega_2(L) - \Omega_1(L)\Xi_1)'}{\partial \mathrm{vec}(\Omega_1(L), \Omega_2(L))} (Q_0^{-1} \otimes \Sigma) \|K\|_2^2 \frac{\partial \mathrm{vec}(\Omega_2(L) - \Omega_1(L)\Xi_1)}{\partial \mathrm{vec}(\Omega_1(L), \Omega_2(L))} \right)^{-1} \\ &= \left((-(\Xi_1 \otimes I_G) \;\; I_{d-L} \otimes I_G)(Q_0^{-1} \otimes \Sigma) \|K\|_2^2 (-(\Xi_1 \otimes I_G) \;\; I_{d-L} \otimes I_G)' \right)^{-1}. \end{aligned}$$

The corresponding ALS test statistic is defined by substituting $\widehat{\Xi}_1$ into (5.30), namely,

$$\widehat{C}_{\mathrm{ALS}}(\widehat{\Theta}_0, L) = Nh^m \|K\|_2^{-2} \mathrm{vec}(\widehat{\Omega}_2(L) - \widehat{\Omega}_1(L)\widehat{\Xi}_1)' \cdot$$

$$\cdot \Big((-(\widehat{\Xi}_1 \otimes I_G) \;\; I_{d-L} \otimes I_G)(\widehat{Q}_0^{-1} \otimes \widehat{\Sigma})(-(\widehat{\Xi}_1 \otimes I_G) \;\; I_{d-L} \otimes I_G)' \Big)^{-1} \cdot$$

$$\cdot \, \mathrm{vec}(\widehat{\Omega}_2(L) - \widehat{\Omega}_1(L)\widehat{\Xi}_1).$$

One may show that, under suitable conditions and under the null hypothesis, the ALS test statistic has also a $\chi^2((G-L)(d-L))$ limit distribution. This result can be used in hypothesis testing.

Remark 5.1.5 The ALS method is particularly appealing in practice because it is applied in the same way as the generalized least squares method. More precisely, to compute the ALS statistic $\widehat{C}_{\mathrm{ALS}}(\widehat{\Theta}_0, L)$ in practice, one first computes the ordinary least squares estimator $\widehat{\Xi}_{\mathrm{OLS}}$ by minimizing (5.30), where S is replaced by the appropriate identity matrix. One then substitutes the obtained $\widehat{\Xi}_{\mathrm{OLS}}$ into S and minimizes (5.30) again but this time by using this new S to obtain $\widehat{\Xi}_{\mathrm{GLS}}$. It is this estimator $\widehat{\Xi}_{\mathrm{GLS}}$ of $\widehat{\Xi}_1$ which is used in practice to compute the test statistic $\widehat{C}_{\mathrm{ALS}}(\widehat{\Theta}_0, L)$.

Singular value decomposition based test. Ratsimalahelo [86] constructed a rank test based on a singular value decomposition (SV decomposition or SVD, in short) of a matrix. Suppose that the matrix of study is a $G \times d$ matrix Θ_0 of rank L_0. In a SV decomposition, the matrix Θ_0 is expressed as a product of three matrices

$$\Theta = C\Psi D' = C \begin{pmatrix} \Psi_1 & 0 \\ 0 & \Psi_2 \end{pmatrix} D'. \tag{5.31}$$

The matrix $\Psi_1 = \mathrm{diag}\{\psi_1, \ldots, \psi_{L_0}\}$ in (5.31) is diagonal such that $\psi_1 \geq \cdots \geq \psi_{L_0} > 0$ and ψ_i^2, $i = 1, \ldots, L_0$, are the non-zero eigenvalues of the matrix $\Theta_0\Theta_0'$ (or, equivalently, the matrix $\Theta_0'\Theta_0$). Since the rank of the matrix Θ_0 is assumed to be L_0, the matrix Ψ_2 is identically zero. If the rank of Θ_0 were assumed to be greater than L_0, say $L > L_0$, then

$$\Psi_2 = \begin{pmatrix} \mathrm{diag}\{\psi_{L_0+1}\ldots,\psi_L\} & 0 \\ 0 & 0 \end{pmatrix},$$

where ψ_i^2, $i = L_0+1,\ldots,L$, are the rest of the non-zero eigenvalues of $\Theta_0\Theta_0'$. The matrices C and D in (5.31) are $G \times G$ and $d \times d$ orthogonal matrices, respectively. The matrix C is made of G linearly independent eigenvectors of the matrix $\Theta_0\Theta_0'$ and the matrix D consists of d linearly independent eigenvectors of the matrix $\Theta_0'\Theta_0$. For more information on singular value decomposition, see Golub and Van Loan [39] and Stewart and Sun [98].

Let now $\widehat{\Theta}_0$ be an estimator for the matrix Θ_0 and let also

$$\widehat{\Theta} = \widehat{C}\widehat{\Psi}\widehat{D}' = \widehat{C}\begin{pmatrix} \widehat{\Psi}_1 & 0 \\ 0 & \widehat{\Psi}_2 \end{pmatrix}\widehat{D}'$$

be its SV decomposition, where the diagonal matrix $\widehat{\Psi}_1$ is $L \times L$ for some $L \geq 1$. Ratsimalahelo [86] showed that, under suitable conditions, when $L = L_0$, the Wald type statistic

$$N\mathrm{vec}(\widehat{\Psi}_2)'\widehat{M}^{-1}\mathrm{vec}(\widehat{\Psi}_2), \tag{5.32}$$

where $\widehat{M}$ is some matrix and N is the sample size, is asymptotically $\chi^2((G-L_0)(d-L_0))$. By using this asymptotic result, the author then constructed a test to determine the rank of a matrix. For more details on this test and for the proof of the aforementioned asymptotic result, see Ratsimalahelo [86].

Remark 5.1.6 We believe that the rank test of Ratsimalahelo [86] is a special case of rank tests considered by Robin and Smith [89]. Suppose for instance that the matrix $\widehat{M}$ in (5.32) is identity. Then, $\mathrm{vec}(\widehat{\Psi}_2)'\mathrm{vec}(\widehat{\Psi}_2) = \sum_{i=L+1}^{G}\psi_i^2$ where ψ_i^2, $i = L+1,\ldots,G$, are the smallest $G-L$ eigenvalues of the matrix $\Theta_0\Theta_0'$. As mentioned in Remark 5.1.4, the asymptotic behavior of such sum statistics has been established by Robin and Smith [89]. Despite this fact, Ratsimalahelo [86] test would still be interesting as it would provide a connection between a special case of Robin and Smith [89] tests and a well-known SVD decomposition.

Concluding remarks We presented above four different test statistics for estimation of the rank in a matrix. Although all these statistics have the same asymptotic behavior, for example, they are $\chi^2((G-L)(d-L))$ in the limit when the true rank is L, it is very important to understand that they might and, in fact, do have different small sample properties. For example, it is well-known to practitioners that rank tests based on the minimum-χ^2 statistic will underestimate the rank whereas the rank estimated by the ALS statistic will be higher (see also Chapter 7 below). These facts are particularly useful in practice because they allow to get a better grip on the estimation object. It is thus advised in practice to draw conclusions not based on the results of one rank estimation method but on the results of several of them.

5.2 Local tests for non-parametric model

In this section, we introduce and study local tests for (NP) model. The problem, formulated in Section 3.2, is to test the hypothesis H_0 : for some fixed L and z, $\mathrm{adrk}\{F(\cdot,z)\} \leq L$ against the alternative H_1 : $\mathrm{adrk}\{F(\cdot,z)\} > L$. In Section 5.2.1, we explain the basic idea behind the statistics used for local tests. In Section 5.2.2, we prove their asymptotics and, in Section 5.2.3, we draw their connections to rank estimation in symmetric matrices. Sections 5.2.4 and 5.2.5 contain the proof of a result used in Section 5.2.2 and also some intermediate results.

5.2.1 Preliminaries

The idea behind local tests for (NP) model lies in the following lemma.

Lemma 5.2.1 *For some fixed z and L, we have* $\operatorname{adrk}\{F(\cdot,z)\} \leq L$ *if and only if the matrix*

$$\Gamma_{w,z} = E\gamma(X_i,z)\widetilde{F}(X_i,z)\widetilde{F}(X_i,z)', \tag{5.33}$$

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

$$\widetilde{F}(x,z) = F(x,z)E\beta(X_i,z) - EF(X_i,z)\beta(X_i,z) \tag{5.34}$$

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.

PROOF: The proof is similar to that of Lemma 5.1.2. Let us first show that $\operatorname{adrk}\{F(\cdot,z)\} \leq L$ implies that the matrix $\Gamma_{w,z}$ has $G-L$ zero eigenvalues. By Definition 3.2.1, we have $\operatorname{adrk}\{F(\cdot,z)\} \leq L$ if and only if

$$F(x,z) = c(z) + A(z)H(x,z), \tag{5.35}$$

for some $G\times 1$ vector $c(z)$, $G\times L$ matrix $A(z)$ and $L\times 1$ vector $H(x,z)$. Then, (5.35) implies that

$$\beta(x,z)F(x,z) = \beta(x,z)c(z) + A(z)\beta(x,z)H(x,z)$$

and, in particular, by substituting X_i for x and taking the expectation, that

$$E\beta(X_i,z)F(X_i,z) = E\beta(X_i,z)c(z) + A(z)E\beta(X_i,z)H(X_i,z). \tag{5.36}$$

Multiplying (5.35) by $E\beta(X_i,z)$ and subtracting from it the relation (5.36), we obtain that

$$\widetilde{F}(x,z) = A(z)(H(x,z)E\beta(X_i,z) - E\beta(X_i,z)H(X_i,z)). \tag{5.37}$$

It follows from relation (5.37) that there are $G-L$ linearly independent vectors $c_j(z)$, $j = 1,\dots,G-L$, 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

$$E\left(c_j(z)'\gamma(X_i,z)^{1/2}\widetilde{F}(X_i,z)\right)^2 = c_j(z)'\Gamma_{w,z}c_j(z) = 0, \tag{5.38}$$

for $j = 1,\dots,G-L$. Relation (5.38) holds if and only if the matrix $\Gamma_{w,z}$ has $G-L$ zero eigenvalues. One can, in fact, go back in the argument above which establishes the first "if and only if" part of the lemma. The second "if and only if" part is obvious. □

Local 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 5.2.1 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

$$\gamma(x,z) = \frac{p(x,z)^2}{\widetilde{p}(x)}, \quad \beta(x,z) = \frac{p(x,z)}{\widetilde{p}(x)}, \tag{5.39}$$

where $p(x,z)$ and $\widetilde{p}(x)$ are the densities of the vector (X,Z) and the variable X, respectively. The idea behind the definition of the estimator of $\Gamma_{w,z}$ that we will consider, is as follows. Observe first that

$$E\beta(X_i,z) = E\frac{p(X_i,z)}{\widetilde{p}(X_i)} = \int_{\mathbb{R}^n} p(x,z)dx = p(z) \approx \frac{1}{N}\sum_{i=1}^N K_h(z-Z_i) =: \widehat{p}(z), \tag{5.40}$$

where $p(z)$ is the density of the variable Z. Observe also that

$$\begin{aligned} Y(z) &:= E\beta(X_i,z)F(X_i,z) = E\frac{p(X_i,z)}{\widetilde{p}(X_i)}F(X_i,z) \approx EF(X_i,Z_i)K_h(z-Z_i) \\ &= EY_iK_h(z-Z_i) \approx \frac{1}{N}\sum_{i=1}^N Y_iK_h(z-Z_i) =: \overline{Y}(z), \end{aligned} \tag{5.41}$$

where the first approximation in (5.41) can be explained by Proposition 4.1.1 as

$$\begin{aligned} EF(X_i, Z_i)K_h(z - Z_i) &= \int_{\mathbb{R}^n}\int_{\mathbb{R}^m} p(x_i, z_i)F(x_i, z_i)K_h(z - z_i)dx_i dz_i \\ &\approx \int_{\mathbb{R}^n} p(x_i, z)F(x_i, z)dx_i = E\frac{p(X_i, z)}{\widetilde{p}(X_i)}F(X_i, z). \end{aligned}$$

Similarly, by Proposition 4.1.1,

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

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

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

Taking $j \neq i$ and using Proposition 4.1.1, we may get a further approximation of $\Gamma_{w,z}$ as

$$\Gamma_{w,z} \approx E(F(X_j, Z_j)p(z) - Y(z))(Y_i p(z) - Y(z))'\widetilde{K}_h(X_i - X_j)K_h(z - Z_i)K_h(z - Z_j).$$

Then, by using $E(U_j|X_i, Z_i, X_j, Z_j) = 0$ once again and the approximations (5.40) and (5.41), we obtain that

$$\begin{aligned} \Gamma_{w,z} &\approx E(Y_j p(z) - Y(z))(Y_i p(z) - Y(z))'\widetilde{K}_h(X_i - X_j)K_h(z - Z_i)K_h(z - Z_j) \\ &\approx E(Y_j \widehat{p}(z) - \overline{Y}(z))(Y_i \widehat{p}(z) - \overline{Y}(z))'\widetilde{K}_h(X_i - X_j)K_h(z - Z_i)K_h(z - Z_j). \quad (5.42) \end{aligned}$$

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

Definition 5.2.1 *(Estimator for local tests in* (NP) *model) Let*

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

where $\widehat{p}(z)$ and $\overline{Y}(z)$ are given by (5.40) and (5.41), respectively.

For the estimator of Σ in $\Gamma_{w,z}\Sigma^{-1}$, we will take $\widehat{\Sigma}$ defined in (4.9).

Remark 5.2.1 In this section, we introduced an estimator which is potentially useful in testing for the adjusted rank adrk$\{F(\cdot,z)\}$. Recall from Section 3.3 that the adjusted rank can be used to determine the rank of a non-parametric demand system. In other applications, one may be interested in testing for the rank rk$\{F(\cdot,z)\}$ itself. The test statistic can then be introduced in a similar (in fact, simpler) way. For example, as in Lemma 5.2.1, one may show that rk$\{F(\cdot,z)\} \leq L$ if and only if the matrix

$$\Upsilon_{w,z} = E(\gamma(X_i,z)F(X_i,z)F(X_i,z)') \tag{5.43}$$

has $G - L$ zero eigenvalues. Then, by taking $\gamma(x,z)$ defined in (5.39) and arguing as in (5.42), one may arrive at the following estimator of $\Upsilon_{w,z}$,

$$\widehat{\Upsilon}_{w,z} = \frac{1}{N(N-1)}\sum_{i\neq j}^{N} Y_iY_j'\widetilde{K}_h(X_i - X_j)K_h(z - Z_i)K_h(z - Z_j). \tag{5.44}$$

Observe that the difference between $\widehat{\Upsilon}_{w,z}$ and $\widehat{\Gamma}_{w,z}$ is that in the definition of $\Gamma_{w,z}$, by subtracting $\overline{Y}(z)$ from Y_i, we account for the additive term $c(z)$ which appears in Definition 3.2.1 of the adjusted rank. (Without $c(z)$, Definition 3.2.1 is that of the rank rk$\{F(\cdot,z)\}$.) In Remark 5.2.4 of the next section, we will state the asymptotic results for the eigenvalues of $\widehat{\Upsilon}_{w,z}\widehat{\Sigma}^{-1}$ which can be used to test H_0 : rk$\{F(\cdot,z)\} \leq L$ against H_1 : rk$\{F(\cdot,z)\} > L$.

5.2.2 Asymptotics

The following result is key to local tests for (NP) model. Let $\widehat{\lambda}_1(z) \leq \cdots \leq \widehat{\lambda}_G(z)$ be the eigenvalues of the matrix $\widehat{\Gamma}_{w,z}\widehat{\Sigma}^{-1}$. Set $V(z) = \left(2\|\widetilde{K}\|_2^2\|K\|_2^4 p(z)^4 \int p(x_i,z)^2 dx_i\right)^{-1/2}$ and let

$$\widehat{V}(z) = \left(2\|\widetilde{K}\|_2^2\|K\|_2^4\widehat{p}(z)^4 N^{-1}\sum_{i=1}^{N}\widehat{p}(X_i, Z_i)K_h(z - Z_i)\right)^{-1/2}, \tag{5.45}$$

where $\widehat{p}(z)$ is defined by (5.40) and $\widehat{p}(x, z)$ is given in (4.7). By Lemma 5.2.12 below, under suitable conditions, $\widehat{V}(z)$ is a consistent estimator of $V(z)$. (See also a remark following Lemma 5.2.12, where we provide a feeling for the estimator $\widehat{V}(z)$.) Let also $\mathcal{Z}_k$ be a symmetric $k \times k$ matrix having independent normal entries with variance 1 in the diagonal and variance 1/2 off the diagonal, and $\lambda_1(\mathcal{Z}_k) \leq \cdots \leq \lambda_k(\mathcal{Z}_k)$ be the eigenvalues of $\mathcal{Z}_k$ in increasing order.

Theorem 5.2.1 *Suppose that Assumptions* (NP) L1–L4 *of Sections 3.2 and 4.1 hold, and that*

$$Nh^{m+3n/2} \to \infty \quad \textit{and} \quad Nh^{m+n/2+2r} \to 0. \tag{5.46}$$

Set $L(z) = \mathrm{adrk}\{F(\cdot, z)\}$. *Then, for* $j = 1, \ldots, G - L(z)$,

$$\widehat{V}(z)Nh^{m+n/2}\widehat{\lambda}_j(z) \xrightarrow{d} \lambda_j(\mathcal{Z}_{G-L(z)}), \tag{5.47}$$

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

$$\widehat{V}(z)Nh^{m+n/2}\widehat{\lambda}_j(z) \xrightarrow{p} +\infty. \tag{5.48}$$

The proof of Theorem 5.2.1 is given in Section 5.2.4 below. We now state and prove two immediate corollaries of Theorem 5.2.1 which can be used in local tests for (NP) model, namely, to test $H_0 : \mathrm{adrk}\{F(\cdot, z)\} \leq L$ against $H_1 : \mathrm{adrk}\{F(\cdot, z)\} > L$. To state the first corollary, let

$$\widehat{T}_1(L,z) = \frac{\widehat{V}(z)Nh^{m+n/2}}{\sqrt{G-L}} \sum_{j=1}^{G-L} \widehat{\lambda}_j(z). \tag{5.49}$$

Theorem 5.2.2 *Under the assumptions of Theorem 5.2.1, we have that, under the hypothesis* $H_0 : \mathrm{adrk}\{F(\cdot,z)\} \le L$,

$$\widehat{T}_1(L,z) \xrightarrow{d} \frac{1}{\sqrt{G-L}} \sum_{j=1}^{G-L} \lambda_j(\mathcal{Z}_{G-L(z)}) \overset{d}{\le} \mathcal{N}(0,1), \tag{5.50}$$

where $\le_d$ *in (5.50) is, in fact,* $=_d$ *for* $L = \mathrm{adrk}\{F(\cdot,z)\}$, *and, under the hypothesis* H_1 : $\mathrm{adrk}\{F(\cdot,z)\} > L$, $\widehat{T}_1(L,z) \to_p +\infty$.

Remark 5.2.2 Observe from (5.50) that the eigenvalues $\widehat{\lambda}_j(z)$, $j = 1,\ldots,G-L(z)$, of $\widehat{\Gamma}_{w,z}\widehat{\Sigma}^{-1}$ can take negative values, in contrast to the eigenvalues of the limit matrix $\Gamma_{w,z}\Sigma^{-1}$ which are all positive. (Were the eigenvalues $\widehat{\lambda}_j(z)$ necessarily positive, then the limit of $\widehat{T}_1(L,z)$ would have support on the positive axis.) This observation can also be seen from Definition 5.2.1 which shows that the matrix $\widehat{\Gamma}_{w,z}$ is not positive definite.

Theorem 5.2.2 is proved below. Observe that the stochastic dominance result in (5.50) and the divergence of the test statistic $\widehat{T}_1(L,z)$ under the alternative hypothesis can be used to test for the adjusted rank $\mathrm{adrk}\{F(\cdot,z)\}$. At a significance level α, the hypothesis $H_0 : \mathrm{adrk}\{F(\cdot,z)\} \le L$ is accepted if $\widehat{T}_1(L,z) \le \mathcal{N}_\alpha(0,1)$ where $\mathcal{N}_\alpha(0,1)$ is the smallest ξ such that $P(\mathcal{N}(0,1) \ge \xi) = \alpha$.

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

$$\widehat{T}_2(L,z) = \widehat{V}(z)^2 N^2 h^{2m+n} \sum_{j=1}^{G-L} (\widehat{\lambda}_j(z))^2. \tag{5.51}$$

Theorem 5.2.3 *Under the assumptions of Theorem 5.2.1, we have that, under the hypothesis* $H_0 : \mathrm{adrk}\{F(\cdot, z)\} \leq L$*,*

$$\widehat{T}_2(L,z) \xrightarrow{d} \sum_{j=1}^{G-L} (\lambda_j(\mathcal{Z}_{G-L(z)}))^2 \overset{d}{\leq} \chi^2((G-L)(G-L+1)/2), \tag{5.52}$$

where $\leq_d$ *in (5.51) is, in fact,* $=_d$ *for* $L = \mathrm{adrk}\{F(\cdot, z)\}$*, and, under the hypothesis* H_1 : $\mathrm{adrk}\{F(\cdot, z)\} > L$*,* $\widehat{T}_2(L,z) \to_p +\infty$.

Observe that by Theorem 5.2.3, $\lim P(\widehat{T}_2(L,z) > x) \leq P(\chi^2((G-L)(G-L+1)/2) > x)$ for all x. The last relation can be used to choose the critical value for the test statistic $\widehat{T}_2(L,z)$.

We now prove Theorems 5.2.2 and 5.2.3.

Proof of Theorem 5.2.2: The convergence in (5.50) follows from (5.47) in Theorem 5.2.1. In order to show the stochastic dominance in (5.50), we use the proof of Theorems 1 and 2 in Donald [28]. By the Poincaré separation theorem (see Magnus and Neudecker [73], p. 209, or Rao [84], p. 65), we have $\lambda_i(\mathcal{Z}_{G-L(z)}) \leq \lambda_i(B'\mathcal{Z}_{G-L(z)}B)$ for $i = 1, \ldots, G-L$, where $L(z) = \mathrm{adrk}\{F(\cdot, z)\}$ and B is any $(G-L(z)) \times (G-L)$ matrix such that $B'B = I_{G-L}$. Now take $B = (0_{(G-L)\times(L-L(z))}\ I_{G-L})'$ so that $B'B = I_{G-L}$. Observe that $B'\mathcal{Z}_{G-L(z)}B =_d \mathcal{Z}_{G-L}$ and hence

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

The convergence under the hypothesis H_1 follows from (5.48) in Theorem 5.2.1. □

PROOF OF THEOREM 5.2.3: The convergence in (5.52) follows from (5.47) in Theorem 5.2.1. To prove the stochastic dominance in (5.52), observe first that

$$\sum_{j=1}^{G-L}(\lambda_j(\mathcal{Z}_{G-L(z)}))^2 = \sum_{j=1}^{G-L}\lambda_j(\mathcal{Z}^2_{G-L(z)}), \tag{5.53}$$

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 increasing order. Letting

$$B = (0_{(G-L)\times(L-L(z))} \quad I_{G-L})'$$

(I_k is a $k \times k$ identity matrix) and arguing as in the proof of Theorem 5.1.5, we can conclude that

$$\sum_{j=1}^{G-L}\lambda_j(\mathcal{Z}^2_{G-L(z)}) \overset{d}{\leq} \sum_{j=1}^{G-L}\lambda_j((B'\mathcal{Z}_{G-L(z)}B)(B'\mathcal{Z}_{G-L(z)}B)), \tag{5.54}$$

Since $B'\mathcal{Z}_{G-L(z)}B =_d \mathcal{Z}_{G-L}$, it follows from (5.53) and (5.54) that

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

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 $2(\mathcal{N}(0,1/2))^2 =_d \mathcal{N}(0,1)^2$). □

Remark 5.2.3 Theorem 5.2.2 is in the spirit of Theorems 1 and 2 in Donald [28]. To our best knowledge, a result like Theorem 5.2.3 does not appear elsewhere in connection to rank testing in a non-parametric relation. (See, however, Section 5.2.3 below.)

Remark 5.2.4 Let $\widehat{\mu}_j(z)$, $j = 1, \dots, G$, be the eigenvalues of the matrix $\widehat{\Upsilon}_{w,z}\Sigma^{-1}$, where $\widehat{\Upsilon}_{w,z}$ is given by (5.44) in connection to hypothesis testing for $R(z) = \mathrm{rk}\{F(\cdot, z)\}$. One may then show as in the proof of Theorem 5.2.1 that, under suitable conditions, for $j = 1, \dots, G - R(z)$,

$$\widehat{v}(z) N h^{m+n/2} \widehat{\mu}_j(z) \xrightarrow{d} \lambda_j(\mathcal{Z}_{G-R(z)}), \tag{5.55}$$

and, for $j = G - R(z) + 1, \dots, G$,

$$\widehat{v}(z) N h^{m+n/2} \widehat{\mu}_j(z) \xrightarrow{p} +\infty, \tag{5.56}$$

where $\widehat{v}(z) = \widehat{V}(z)\widehat{p}(z)^2$. The convergence in (5.55) and (5.56) can be used, similarly to Theorems 5.2.2 and 5.2.3, to test the hypothesis $H_0 : \mathrm{rk}\{F(\cdot, z)\} \leq L$ against the alternative $H_1 : \mathrm{rk}\{F(\cdot, z)\} > L$.

Remark 5.2.5 Observe from Definition 5.2.1 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. 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

$$\begin{aligned} \widehat{\Gamma}_{w,z} &= \frac{1}{N(N-1)} \sum_{i \neq j}^{N} (Y_i \widehat{p}(z) - \overline{Y}(z))(Y_j \widehat{p}(z) - \overline{Y}(z))' \cdot \\ &\qquad \cdot \widetilde{K}_{h_1}(X_i - X_j) K_{h_2}(z - Z_i) K_{h_2}(z - Z_j), \end{aligned} \tag{5.57}$$

where $h_1, h_2 > 0$ (compare with Definition 5.2.1). 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 (5.57)) satisfy the limit results analogous to those in Theorems 5.2.1, 5.2.2 and 5.2.3. The only difference is that the factor $Nh^{m+n/2}$ in the normalizations of (5.47), (5.48), (5.49) and

(5.51) should now be replaced by

$$Nh_2^m h_1^{n/2}. \tag{5.58}$$

In our applications and simulation study (Chapter 7), we will consider the test statistic (5.57) and use the normalization (5.58).

5.2.3 Connection to rank estimation in symmetric matrices

Observe that another way to formulate Lemma 5.2.1 is to say that $\mathrm{adrk}\{F(\cdot,z)\} \le L$ holds if and only if $\mathrm{rk}\{\Gamma_{z,w}\} \le L$. The problem of testing for the adjusted rank $\mathrm{adrk}\{F(\cdot,z)\}$ thus becomes that of testing for the rank of the matrix $\Gamma_{w,z}$. One may ask then why not to use any of the rank estimation methods (LDU, minimum χ^2, ALS and so on) described in Section 5.1 directly.

In fact, the test statistic $\widehat{T}_2(L,z)$ defined in (5.51) can be viewed as a minimum-χ^2 statistic for the matrix $\Gamma_{w,z}$. Indeed, observe first that

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

where $\widehat{\mu}_j(z) = (\widehat{\lambda}_j(z))^2$ are the eigenvalues of the matrix

$$(\widehat{\Gamma}_{w,z}\widehat{\Sigma}^{-1}\widehat{\Gamma}_{w,z})\widehat{\Sigma}^{-1}. \tag{5.59}$$

Then, by Theorem 5.1.4, we have

$$\widehat{T}_2(L,z) = \widehat{V}(z)^2 N^2 h^{2m+n} \min_{\mathrm{rk}\{\Gamma\}\le L} \mathrm{vec}(\widehat{\Gamma}_{w,z} - \Gamma)'(\widehat{\Sigma}\otimes\widehat{\Sigma})^{-1}\mathrm{vec}(\widehat{\Gamma}_{w,z} - \Gamma) \tag{5.60}$$

and hence that $\widehat{T}_2(L,z)$ looks like a minimum-χ^2 statistic for the matrix $\Gamma_{w,z}$ (compare (5.60) and (5.9)). Observe also that the sample variance-covariance matrix $\widehat{W}(z)$ in (5.9) is now replaced by $\widehat{\Sigma} \otimes \widehat{\Sigma}$.

Although $\widehat{T}_2(L, z)$, when expressed through (5.60), looks like a minimum-χ^2 statistic, there is something fundamentally different from the situation considered in Sections 5.1.2 and 5.1.3. The difference is that the matrix $\widehat{\Gamma}_{w,z}$ is now necessarily *symmetric*. This has a few immediate implications. Supposing that one wants to start with the minimum-χ^2 statistic for the matrix $\Gamma_{w,z}$ directly, it may not be immediately clear what normalization to use and what would be the limit of the corresponding test statistic. The problem here is that, since $\widehat{\Gamma}_{w,z}$ is symmetric, its variance-covariance matrix $W(z)$ is singular and hence $W(z)^{-1}$ is not defined. Theorem 5.2.3 and relation (5.60) show that one can, in fact, use the normalization $(\widehat{\Sigma} \otimes \widehat{\Sigma})^{-1}$ and that the corresponding limit (under the assumption $\mathrm{rk}\{\Gamma_{w,z}\} = L$) is still a χ^2-distribution but now with a $(G-L)(G-L+1)/2$ degrees of freedom. Note that the number of the degrees of freedom is less than $(G-L)(G-L)$ for the matrix $\widehat{\Gamma}_{w,z}$ without symmetry restrictions.

Remark 5.2.6 To get a feeling for the normalization $(\widehat{\Sigma} \otimes \widehat{\Sigma})^{-1}$ used in (5.60), observe first that $(\widehat{\Sigma} \otimes \widehat{\Sigma})^{-1/2}\mathrm{vec}(\widehat{\Gamma}_{w,z} - \Gamma_{w,z}) = \mathrm{vec}(\widehat{\Sigma}^{-1/2}(\widehat{\Gamma}_{w,z} - \Gamma_{w,z})\widehat{\Sigma}^{-1/2})$. One may argue as in the proof of Theorem 5.2.1 below that under the conditions of that theorem, $\widehat{V}(z)Nh^{m+n/2}\widehat{\Sigma}^{-1/2}(\widehat{\Gamma}_{w,z} - \Gamma_{w,z})\widehat{\Sigma}^{-1/2}$ converges in distribution to the matrix $\mathcal{Z}_G$ which is defined in the beginning of Section 5.2.2. Hence,

$$\widehat{V}(z)Nh^{m+n/2}(\widehat{\Sigma} \otimes \widehat{\Sigma})^{-1/2}\mathrm{vec}(\widehat{\Gamma}_{w,z} - \Gamma_{w,z}) \xrightarrow{d} \mathrm{vec}(\mathcal{Z}_G). \tag{5.61}$$

The convergence (5.61) should be enough to establish the asymptotics of the test statistic (5.60) directly without referring to Theorem 5.2.3, but the author is not aware of the proof.

Remark 5.2.7 The problem of testing whether the first $(G-L)$ eigenvalues of the type (5.59) matrix are equal to zero (or, equivalently, testing for the rank of $\Gamma_{w,z}$) has been also addressed by Robin and Smith [88, 89]. See also Remark 5.1.4. These authors have also found that the symmetry restriction on a matrix leads to $(G-L)(G-L+1)/2$ degrees of freedom for the limiting χ^2-distribution. See in particular the paper by Robin and Smith [88].

One may ask next what happens, for example, with the LDU and the ALS methods when matrices of interest are symmetric. Is it then enough to replace the normalization matrix $\widehat{W}(z)$ by $\widehat{\Sigma} \otimes \widehat{\Sigma}$ as in the case of minimum-χ^2 statistic and compare the result to a χ^2-distribution with $(G-L)(G-L+1)/2$ degrees of freedom? We leave this and other related questions for the future work.

5.2.4 The proof of Theorem 5.2.1

In this section, we prove Theorem 5.2.1 which describes the asymptotic behavior of the eigenvalues used for local tests.

PROOF OF THEOREM 5.2.1: The proof of the convergence (5.47) uses ideas of the proof of Lemma 2 in Section 2.2 of Donald [28]. To simplify notation, we set

$$\widetilde{K}_{ij} = \widetilde{K}_h(X_i - X_j), \quad K_{z,i} = K_h(z - Z_i), \tag{5.62}$$

and

$$\overline{F}(z) = \frac{1}{N}\sum_{i=1}^{N} F(X_i, z) K_h(z - Z_i), \tag{5.63}$$

$$\overline{\Delta F}(z) = \frac{1}{N}\sum_{i=1}^{N} \Delta F(X_i, Z_i, z) K_h(z - Z_i), \tag{5.64}$$

$$\overline{U}(z) = \frac{1}{N}\sum_{i=1}^{N} U_i K_h(z - Z_i), \tag{5.65}$$

where

$$\Delta F(x_i, z_i, z) = F(x_i, z_i) - F(x_i, z). \tag{5.66}$$

By using Definition 5.2.1 of $\widehat{\Gamma}_{w,z}$ and by writing $Y_i = F(X_i, Z_i) + U_i = F(X_i, z) + (F(X_i, Z_i) - F(X_i, z)) + U_i$, we can express the matrix $\widehat{\Gamma}_{w,z}$ as

$$\begin{aligned}\widehat{\Gamma}_{w,z} &= \mathcal{A}_1 + \delta\mathcal{A}_2 + \delta^2\mathcal{A}_3 \\ &= A_1 + \delta(A_2 + A_2') + \delta^2(A_3 + A_3' + A_4),\end{aligned} \tag{5.67}$$

where

$$\delta = \frac{1}{\sqrt{Nh^{m+n/2}}}, \tag{5.68}$$

$\mathcal{A}_1 = A_1$, $\mathcal{A}_2 = A_2 + A_2'$, $\mathcal{A}_3 = A_3 + A_3' + A_4$, the first order term A_1 is

$$A_1 = \widehat{p}(z)^2 A_{1,1} - \widehat{p}(z)\overline{F}(z)A_{1,2} - \widehat{p}(z)A_{1,2}'\overline{F}(z)' - \overline{F}(z)\overline{F}(z)'A_{1,3}, \tag{5.69}$$

where

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

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

$$A_{1,3} = \frac{1}{N(N-1)}\sum_{i\neq j}^{N} \widetilde{K}_{ij}K_{z,i}K_{z,j},$$

the second order term A_2 is

$$\begin{aligned}A_2 \;=\;& \delta^{-1}\widehat{p}(z)^2 A_{2,1} - \delta^{-1}\widehat{p}(z)A_{2,2}\overline{F}(z)' + \delta^{-1}\widehat{p}(z)^2 A_{2,3} - \delta^{-1}\widehat{p}(z)A_{2,4}\overline{F}(z)' \\ &- \delta^{-1}\widehat{p}(z)(\overline{\Delta F}(z) + \overline{U}(z))A_{1,2} + \delta^{-1}(\overline{\Delta F}(z) + \overline{U}(z))A_{1,3}\overline{F}(z)',\end{aligned} \tag{5.70}$$

where

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

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

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

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

the third order terms A_3 and A_4 are

$$\begin{aligned} A_3 &= \delta^{-2}\widehat{p}(z)^2 A_{3,1} + \delta^{-2}\widehat{p}(z)^2 \delta^{-2} A_{3,2} - \delta^{-2}\widehat{p}(z)(\overline{\Delta F}(z) + \overline{U}(z)) A_{2,2}' \\ &\quad - \delta^{-2}\widehat{p}(z)(\overline{\Delta F}(z) + \overline{U}(z)) A_{2,4}' + \delta^{-2}(\overline{\Delta F}(z) + \overline{U}(z))(\overline{\Delta F}(z) + \overline{U}(z))' A_{1,3}, \quad (5.71) \end{aligned}$$

where

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

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

and

$$A_4 = \frac{\delta^{-2}\widehat{p}(z)^2}{N(N-1)} \sum_{i \neq j}^{N} U_i U_j' \widetilde{K}_{ij} K_{z,i} K_{z,j}. \quad (5.72)$$

Recall that 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 5.2.2 below, the matrix $A_1\Sigma^{-1}$ has $G - L(z)$ zero eigenvalues and the remaining ones are strictly positive with probability approaching 1. Then, since we are interested in the 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

$$J'A_1\Sigma^{-1}J = J'A_1J(J'\Sigma J)^{-1} \tag{5.73}$$

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 that $C = I$ (otherwise, the proof goes along similar lines) 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

$$J'\Sigma J = J_1'J_0'\Sigma J_0J_1 = J_1'J_1 = I. \tag{5.74}$$

Now, relations (5.73) and (5.74), 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 5.2.3 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 (5.74), 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 5.2.11, $A_i = O_p(1)$, $i = 1, \ldots, 4$, in Lemmas 5.2.5–5.2.8 below and the discussion above, $\delta^{-2}\widehat{\lambda}_j(z)$ is equal to δ^{-2} times the jth smallest eigenvalue of the matrix

$$\begin{aligned}
J'\widehat{\Gamma}_{w,z}J(J'\widehat{\Sigma}J)^{-1} &= J'\widehat{\Gamma}_{w,z}J(J'\Sigma J + \delta J'BJ)^{-1} \\
&= J'\widehat{\Gamma}_{w,z}J(I + \delta J'BJ)^{-1} \\
&= J'(A_1 + \delta A_2 + \delta^2 A_3)J(I - \delta J'BJ + \delta J'B^2J - \ldots) \\
&= D_1 + \delta D_2 + \delta^2 D_3 + O_p(\delta^3),
\end{aligned}$$

where $D_1 = J'\mathcal{A}_1 J = J'A_1 J$ is diagonal, $D_2 = J'(\mathcal{A}_2 - \mathcal{A}_1 B)J = O_p(1)$ and $D_3 = J'(\mathcal{A}_3 - \mathcal{A}_2 B + \mathcal{A}_1 B)J = O_p(1)$. By applying Lemma 1 in Fujikoshi [37], we can conclude that $\widehat{\lambda}_j(z)$, $j = 1, \dots, G - L(z)$, are also the eigenvalues of the matrix

$$0I + \delta\widetilde{D}_2 + \delta^2\widetilde{D}_3 + O_p(\delta^3), \tag{5.75}$$

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

The matrix $\widetilde{D}_2$ in (5.75) 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 $D_2 = J'\mathcal{A}_2 J - J'\mathcal{A}_1 BJ$. Recall from (5.67) and the discussion above that $J'\mathcal{A}_2 J$ is a sum of two matrices $J'A_2 J$ and $J'A_2' J$, the matrix $J'A_2 J$ 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 $J'\mathcal{A}_2 J$ is identically zero. Turning to the second term $J'\mathcal{A}_1 BJ = J'\mathcal{A}_1 J(J'BJ)$ in the matrix D_2, since $J'\mathcal{A}_1 J$ is diagonal with its last $G - L(z)$ rows zero, we obtain that the $(G - L(z)) \times (G - L(z))$ matrix corresponding to $J'\mathcal{A}_1 BJ$ is identically zero as well. Then, $\widetilde{D}_2 = 0$ and hence $\widehat{\lambda}_j(z)$, $j = 1, \dots, G - L(z)$, are also the eigenvalues of the matrix

$$\delta^2\widetilde{D}_3 + O_p(\delta^3)$$

or $\delta^{-2}\widehat{\lambda}_j(z) = Nh^{m+n/2}\widehat{\lambda}_j(z)$, $j = 1, \dots, G - L(z)$, are the eigenvalues of the matrix

$$\widetilde{D}_3 + o_p(1). \tag{5.76}$$

According to Lemma 1 in Fujikoshi [37], the matrix $\widetilde{D}_3$ in (5.76) (or (5.75)) 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

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

Hence, it follows that

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

where a $(G - L(z)) \times 1$ vector $\widetilde{U}_j = J'U_i$ satisfies $E\widetilde{U}_j\widetilde{U}_j' = I$. By Lemma 5.2.10 below, we have

$$\widehat{V}(z)\widetilde{D}_3 \xrightarrow{d} \mathcal{Z}_{G-L_0(z)}. \tag{5.77}$$

The convergence (5.47) now follows from (5.76) and (5.77) by the continuous mapping theorem.

Finally, the convergence (5.48) holds, since by the continuous mapping theorem, $\widehat{\lambda}_j(z) \to \lambda_j(z)$ in probability, where $0 \leq \lambda_1(z) \leq \cdots \leq \lambda_G(z)$ are the eigenvalues of the matrix $\Gamma_{w,z}\Sigma^{-1}$ and, by Lemma 5.2.1, $\lambda_j(z) > 0$ for $j = G - L(z) + 1, \ldots, G$. □

We conclude this section with two elementary lemmas used above.

Lemma 5.2.2 *The matrix A_1 (or the matrix $A_1\Sigma^{-1}$) in (5.67) has $G - L(z)$ zero eigenvalues and the remaining ones are positive with probability approaching* 1.

PROOF: Observe first from (5.69) that

$$A_1 = \frac{1}{N(N-1)} \sum_{i \neq j}^{N} (F(X_i, z)\widehat{p}(z) - \overline{F}(z))(F(X_i, z)\widehat{p}(z) - \overline{F}(z))'\widetilde{K}_{ij}K_{z,i}K_{z,j}. \tag{5.78}$$

By the definition of $L(z) = \text{adrk}\{F(\cdot, z)\}$, we have

$$F(x,z) = c(z) + A(z)H(x,z), \tag{5.79}$$

where $A(z)$ is a $G \times L(z)$ matrix. This implies that

$$\overline{F}(z) = \widehat{p}(z)c(z) + A(z)\overline{H}(z), \tag{5.80}$$

where

$$\overline{H}(z) = \frac{1}{N}\sum_{i=1}^{N} H(X_i, z)K_h(z - Z_i).$$

By substituting X_i for x in (5.79), multiplying (5.79) by $\widehat{p}(z)$ and then subtracting from (5.79) the relation (5.80), we get that

$$F(X_i, z)\widehat{p}(z) - \overline{F}(z) = A(z)(H(X_i, z)\widehat{p}(z) - \overline{H}(z)). \tag{5.81}$$

By substituting (5.81) into (5.78), we further obtain that $A_1 = A(z)H_1A(z)'$, where

$$H_1 = \frac{1}{N(N-1)}\sum_{i \neq j}^{N}(H(X_i, z)\widehat{p}(z) - \overline{H}(z))(H(X_i, z)\widehat{p}(z) - \overline{H}(z))'\widetilde{K}_{ij}K_{z,i}K_{z,j}.$$

Since $A(z)$ is a $G \times L(z)$ matrix, there are $G - L(z)$ linearly independent vectors $c_j(z)$, $j = 1, \ldots, G - L(z)$, such that

$$c_j(z)A(z) = 0. \tag{5.82}$$

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, since, by Lemma 5.2.5 below, $A_1 \to_p \Gamma_{w,z}$ and by Lemma 5.2.1, the matrix $\Gamma_{w,z}$ has $G - L(z)$ zero eigenvalues, and the remaining ones are strictly positive. $\square$

Lemma 5.2.3 *The eigenvectors corresponding to $G - L(z)$ zero eigenvalues of the matrix A_1 in Lemma 5.2.2 are also eigenvectors for the matrix A_2 in (5.67) 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 5.2.2 and by using (5.82), we have $cA(z) = 0$. Observe now that A_2 in (5.70) can be expressed as

$$A_2 = \frac{\delta^{-1}}{N(N-1)} \sum_{i \neq j}^{N} \left(\Delta F(X_i, Z_i, z)\widehat{p}(z) + U_i\widehat{p}(z) - \overline{\Delta F}(z) - \overline{U}(z)\right) \cdot \\ \cdot \left(F(X_j, z)\widehat{p}(z) - \overline{F}(z)\right)' \widetilde{K}_{ij} K_{z,i} K_{z,j}.$$

Then, by using (5.81) in the proof of Lemma 5.2.2,

$$A_2 = \frac{\delta^{-1}}{N(N-1)} \sum_{i \neq j}^{N} \left(\Delta F(X_i, Z_i, z)\widehat{p}(z) + U_i\widehat{p}(z) - \overline{\Delta F}(z) - \overline{U}(z)\right) \cdot \\ \cdot \left(H(X_j, z)\widehat{p}(z) - \overline{H}(z)\right)' \widetilde{K}_{ij} K_{z,i} K_{z,j}\, A(z)'.$$

Since $cA(z) = 0$, it follows that $A_2 c' = 0$. This concludes the proof. □

5.2.5 Intermediate results

In this section, we establish the results used in the proof of Theorem 5.2.1 above. Their proofs often use the notion of a second order U-statistic whose definition we recall next for later reference.

Definition 5.2.2 *(Second order U-statistic) Let W_i, $i = 1, \ldots, N$, be i.i.d. random variables and $a_N : \mathbb{R} \times \mathbb{R} \mapsto \mathbb{R}$ be a symmetric kernel (that is, $a_N(x, y) = a_N(y, x)$). Then,*

$$U_N = \binom{N}{2}^{-1} \sum_{1 \leq i < j \leq N} a_N(W_i, W_j) \tag{5.83}$$

is called a second order U-statistic for the sequence (W_i).

We now give a result on the limit behavior of a second order U-statistic which will also be used many many times below. This result follows easily from the proof of Lemma 3.1 in Powell, Stock and Stoker [82]. It is, however, often easier to use and yields stronger results than a direct applications of Lemma 3.1 in Powell et al. [82] itself.

Lemma 5.2.4 *(Limit behavior of a second order U-statistic) Let U_N be a second order U-statistic defined by (5.83). Then,*

$$U_N = Ea_N(W_i, W_j) + O_p\left(\sqrt{\frac{E(E(a_N(W_i, W_j)|W_i)^2)}{N}} + \sqrt{\frac{Ea_N(W_i, W_j)^2}{N^2}}\right). \tag{5.84}$$

PROOF: Let

$$\widehat{U}_N = Ea_N(W_i, W_j) + \frac{2}{N}\sum_{i=1}^{N}\Big(E(a_N(W_i, W_j)|W_i) - Ea_N(W_i, W_j)\Big) \tag{5.85}$$

be the so-called projection of the U-statistic U_N (see Serfling [95] or Powell et al. [82]). Then, as in the proof of Lemma 3.1 in Powell et al. [82],

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

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

$$E\Big(U_N - \widehat{U}_N\Big)^2 = O\left(\frac{Ea_N(W_i, W_j)^2}{N^2}\right)$$

or

$$U_N - \widehat{U}_N = O_p\left(\sqrt{\frac{Ea_N(W_i, W_j)^2}{N^2}}\right). \tag{5.86}$$

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

$$E\left(\frac{2}{N}\sum_{i=1}^{N}\Big(E(a_N(W_i, W_j)|W_i) - Ea_N(W_i, W_j)\Big)\right)^2$$

$$= \frac{4}{N}E\Big(E(a_N(W_i, W_j)|W_i) - Ea_N(W_i, W_j)\Big)^2 \leq \frac{4E(E(a_N(W_i, W_j)|W_i)^2)}{N},$$

the result (5.84) follows from (5.85) and (5.86). □

The next four lemmas concern the orders of the terms A_1, A_2, A_3 and A_4 in the decomposition (5.67).

Lemma 5.2.5 *(Order of A_1) Under the assumptions of Theorem 5.2.1, we have*

$$A_1 = \Gamma_{w,z} + o_p(1), \tag{5.87}$$

where A_1 is given in (5.69) and $\Gamma_{w,z}$ is defined by (5.33), (5.34) and (5.39).

PROOF: By using (5.98) in Lemma 5.2.9 below, it is enough to show that

$$A_{1,1} = E\frac{p(X_i, z)^2}{\widetilde{p}(X_i)}F(X_i, z)F(X_i, z)' + o_p(1), \tag{5.88}$$

$$A_{1,2} = E\frac{p(X_i, z)^2}{\widetilde{p}(X_i)}F(X_i, z)' + o_p(1), \tag{5.89}$$

$$A_{1,3} = E\frac{p(X_i, z)^2}{\widetilde{p}(X_i)} + o_p(1), \tag{5.90}$$

where $A_{1,1}$, $A_{1,2}$ and $A_{1,3}$ are defined by (5.69). We will prove only relation (5.88) since the proofs of (5.89) and (5.90) are similar. We will also consider only the case $G = 1$, that is, when the dimension of a vector F is 1. The result in the case $G \geq 2$ can be proved by considering each component of the matrix $A_{1,1}$ separately.

Observe that $A_{1,1}$ can be expressed as a second order U-statistic (5.83) with $W_i = (X_i, Z_i)$ and

$$a_N(W_i, W_j) = F(X_i, z)F(X_j, z)\widetilde{K}_{ij}K_{z,i}K_{z,j}.$$

Then, by Lemma 5.2.4 above, to prove (5.88), it is enough to show that

$$Ea_N(W_i, W_j) \to E\frac{p(X_i, z)^2}{\widetilde{p}(X_i)}F(X_i, z)F(X_i, z) \tag{5.91}$$

and

$$E\left(E(a_N(W_i, W_j)|W_i)^2\right) = o(N), \quad Ea_N(W_i, W_j)^2 = o(N^2). \tag{5.92}$$

By using the assumptions of Theorem 5.2.1 and by applying Proposition 4.1.1, we obtain that

$$\begin{aligned} Ea_N(W_i, W_j) &= \int_{\mathbb{R}^n} dx_i\, F(x_i, z)\Big\{\int_{\mathbb{R}^m}\int_{\mathbb{R}^n}\int_{\mathbb{R}^m} dz_i dx_j dz_j \cdot \\ &\quad \cdot F(x_j, z)p(x_i, z_i)p(x_j, z_j)\widetilde{K}_h(x_i - x_j)K_h(z - z_i)K_h(z - z_j)\Big\} \\ &= \int_{\mathbb{R}^n} F(x_i, z)F(x_i, z)p(x_i, z)^2 dx_i + o(1) \\ &= E\left(\frac{p(X_i, z)^2}{\widetilde{p}(X_i)}F(X_i, z)F(X_i, z)\right) + o(1), \end{aligned}$$

which shows (5.91). To show the first relation in (5.92), observe that, by using Proposition 4.1.1,

$$\begin{aligned} E\left(E(a_N(W_i, W_j)|W_i)^2\right) &= E\left(F(X_i, z)^2K_{z,i}^2E(F(X_j, z)\widetilde{K}_{ij}K_{z,j}|X_i)^2\right) \\ &= \|K\|_2^2h^{-m}\int_{\mathbb{R}^n}\int_{\mathbb{R}^m} dx_i dz_i\; F(x_i, z)^2p(x_i, z_i)K_{2,h}(z - z_i)\,\cdot \end{aligned}$$

$$\cdot \left\{ \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} F(x_j, z) p(x_j, z_j) \widetilde{K}_h(x_i - x_j) K_h(z - z_j) dx_j dz_j \right\}^2$$

$$= \|K\|_2^2 h^{-m} \int_{\mathbb{R}^n} F(x_i, z)^2 p(x_i, z) \left(F(x_i, z) p(x_i, z)\right)^2 dx_i + o(h^{-m}) = O(h^{-m}).$$

Since $Nh^m \to \infty$, we obtain that $E(E(a_N(W_i, W_j)|W_i)^2) = o(N)$. As for the second relation in (5.92), by setting $\widetilde{K}_{2,ij} = \widetilde{K}_{2,h}(X_i - X_j)$, $K_{2,z,i} = K_{2,h}(z - Z_i)$ and by using Proposition 4.1.1, we obtain that

$$Ea_N(W_i, W_j)^2 = EF(X_i, z)^2 F(X_j, z)^2 \widetilde{K}_{ij}^2 K_{z,i}^2 K_{z,j}^2$$

$$= \|\widetilde{K}\|_2^2 \|K\|_2^4 h^{-2m-n} EF(X_i, z)^2 F(X_j, z)^2 \widetilde{K}_{2,ij} K_{2,z,i} K_{2,z,j} = O(h^{-2m-n}) = o(N^2),$$

since $Nh^{m+n/2} \to \infty$. □

Lemma 5.2.6 *(Order of A_2) Under the assumptions of Theorem 5.2.1, we have $A_2 = o_p(1)$, where A_2 is defined by (5.70).*

PROOF: We will show that

$$A_{2,i} = O_p\left(h^r + \sqrt{\frac{h^r}{Nh^m}} + \sqrt{\frac{h^r}{N^2 h^{2m+n}}}\right), \quad i = 1, 2, \tag{5.93}$$

$$A_{2,i} = O_p\left(\frac{1}{\sqrt{Nh^m}} + \frac{1}{\sqrt{N^2 h^{2m+n}}}\right), \quad i = 3, 4, \tag{5.94}$$

where $A_{2,i}$, $i = 1, 2, 3, 4$, are defined by (5.70). Then, by using Lemma 5.2.9 below, relations (5.89) and (5.90), and also (5.68) and (5.70), the order of A_2 can be shown to be

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

This proves that $A_2 = o_p(1)$ since $Nh^{m+n/2+2r} \to 0$ and $Nh^{m+n/2} \to \infty$.

Since the proof of (5.93) is similar when $i = 1$ and $i = 2$, we show (5.93) only in the case $i = 2$. We also consider only the case $G = 1$. Observe that $A_{2,2}$ is a second order

U-statistic (5.83) with $W_i = (X_i, Z_i)$ and

$$a_N(W_i, W_j) = \frac{1}{2}\big(\Delta F(X_i, Z_i, z) + \Delta F(X_j, Z_j, z)\big)\widetilde{K}_{ij}K_{z,i}K_{z,j}.$$

To find the order of $A_{2,2}$, we will use Lemma 5.2.4 above, in which case we need to obtain the orders of $Ea_N(W_i, W_j)$, $E(E(a_N(W_i, W_j)|W_i)^2)$ and $Ea_N(W_i, W_j)^2$. By using the assumptions of Theorem 5.2.1 and Proposition 4.1.1, we have

$$Ea_N(W_i, W_j) = E\Delta F(X_i, Z_i, z)\widetilde{K}_{ij}K_{z,i}K_{z,j} = \int_{\mathbb{R}^n} dx_i \bigg\{ \int_{\mathbb{R}^m} \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} dz_i dx_j dz_j \cdot$$

$$\cdot \, \Delta F(x_i, z_i, z) p(x_i, z_i) p(x_j, z_j) \widetilde{K}_h(x_i - x_j) K_h(z - z_i) K_h(z - z_j) \bigg\} = O(h^r).$$

Similarly, setting $K_{2,z,i} = K_{2,h}(z - Z_i)$, we obtain that

$$E\left(E(a_N(W_i, W_j)|W_i)^2\right) \le \frac{\|K\|_2^2}{2h^m} E\Big(\Delta F(X_i, Z_i, z)^2 K_{2,z,i} E(\widetilde{K}_{ij}K_{z,j}|X_i)^2\Big)$$

$$+\frac{\|K\|_2^2}{2h^m} E\Big(K_{2,z,i} E(\Delta F(X_j, Z_j, z)\widetilde{K}_{ij}K_{z,j}|X_i)^2\Big) = O(h^{r-m} + h^{2r-m}) = O(h^{r-m})$$

and, by setting $\widetilde{K}_{2,ij} = \widetilde{K}_{2,h}(X_i - X_j)$, that

$$Ea_N(W_i, W_j)^2 \le 4E\Delta F(X_i, Z_i, z)^2 \widetilde{K}_{ij}^2 K_{z,i}^2 K_{z,j}^2$$

$$= \frac{4\|\widetilde{K}\|_2^2 \|K\|_2^4}{h^{2m+n}} E\Delta F(X_i, Z_i, z)^2 \widetilde{K}_{2,ij} K_{2,z,i} K_{2,z,j} = O\left(\frac{h^r}{h^{2m+n}}\right).$$

Relation (5.93) with $i = 2$ now follows from Lemma 5.2.4 above. We will now show (5.94) with $i = 4$ when $G = 1$. (The proof of (5.94) with $i = 3$ is similar.) Observe that $A_{2,4}$ is a second order U-statistic (5.83) with $W_i = (Y_i, X_i, Z_i)$ and

$$a_N(W_i, W_j) = \frac{1}{2}(U_i + U_j)\widetilde{K}_{ij}K_{z,i}K_{z,j}.$$

Since $Ea_N(W_i, W_j) = 0$, we only need to find the orders of the terms $Ea_N(W_i, W_j)^2$ and $E(E(a_N(W_i, W_j)|W_i)^2)$. By using Proposition 4.1.1, we have

$$E\left(E(a_N(W_i, W_j)|W_i)^2\right) = \frac{1}{4}E\Big(U_i^2 K_{z,i}^2 E(\widetilde{K}_{ij}K_{z,i}|X_i)^2\Big)$$

$$= \frac{\|K\|_2^2}{4h^m}E\Big(K_{2,z,i}^2 E(\widetilde{K}_{ij}K_{z,i}|X_i)^2\Big) = O(h^{-m})$$

and

$$Ea_N(W_i, W_j)^2 = \frac{1}{2}EU_i^2\widetilde{K}_{ij}^2 K_{z,i}^2 K_{z,j}^2$$

$$= \frac{\|\widetilde{K}\|_2^2\|K\|_2^4}{2h^{2m+n}}E\widetilde{K}_{2,ij}K_{2,z,i}K_{2,z,j} = O(h^{-2m-n}).$$

Relation (5.94) with $i = 4$ now follows from Lemma 5.2.4 above. □

Lemma 5.2.7 *(Order of A_3) Under the assumptions of Theorem 5.2.1, we have $A_3 = o_p(1)$, where A_3 is defined by (5.71).*

PROOF: We will show that

$$A_{3,1} = O_p\left(h^{2r} + \sqrt{\frac{h^{3r}}{Nh^m}} + \sqrt{\frac{h^{2r}}{N^2h^{2m+n}}}\right), \tag{5.95}$$

$$A_{3,2} = O_p\left(\sqrt{\frac{h^{2r}}{Nh^m}} + \sqrt{\frac{h^{r}}{N^2h^{2m+n}}}\right), \tag{5.96}$$

where $A_{3,1}$ and $A_{3,2}$ are defined by (5.71). Then, by using Lemma 5.2.9 below, relations (5.90), (5.93) and (5.94), and also (5.71), one can deduce that

$$A_3 = O_p\left(Nh^{m+n/2+2r} + \frac{1}{\sqrt{Nh^m}} + h^{r/2} + h^{n/2}\right).$$

This proves that $A_3 = o_p(1)$ since $Nh^{m+n/2+2r} \to 0$ and $Nh^m \to \infty$.

We show (5.95) and (5.96) for $G = 1$ only. Observe that $A_{3,1}$ is a second order U-statistic (5.83) 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 setting $\widetilde{K}_{2,ij} = \widetilde{K}_{2,h}(X_i - X_j)$, $K_{2,z,i} = K_{2,h}(z - Z_i)$ and using Proposition 4.1.1, we have

$$Ea_N(W_i, W_j) = \int_{\mathbb{R}^n}\int_{\mathbb{R}^m}\int_{\mathbb{R}^n}\int_{\mathbb{R}^m} dx_i dz_i dx_j dz_j \cdot$$

$$\cdot \Delta F(x_i, z_i, z)\Delta F(x_j, z_j, z)p(x_i, z_i)p(x_j, z_j)\widetilde{K}_h(x_i - x_j)K_h(z - z_i)K_h(z - z_j) = O(h^{2r}),$$

$$E\left(E(a_N(W_i, W_j)|W_i)^2\right) \le E\Big(\Delta F(X_i, Z_i, z)^2 K_{z,i}^2 E(\Delta F(X_j, Z_j, z)\widetilde{K}_{ij}K_{z,j}|X_i)^2\Big)$$

$$= \frac{\|K\|_2^2}{h^m}E\Big(\Delta F(X_i, Z_i, z)^2 K_{2,z,i}E(\Delta F(X_j, Z_j, z)\widetilde{K}_{ij}K_{z,j}|X_i)^2\Big) = O(h^{3r-m})$$

and

$$Ea_N(W_i, W_j)^2 = E\Delta F(X_i, Z_i, z)^2\Delta F(X_j, Z_j, z)^2\widetilde{K}_{ij}^2 K_{z,i}^2 K_{z,j}^2$$

$$= \frac{\|\widetilde{K}\|_2^2\|K\|_2^4}{h^{2m+n}}E\Delta F(X_i, Z_i, z)^2\Delta F(X_j, Z_j, z)^2\widetilde{K}_{2,ij}K_{2,z,i}K_{2,z,j} = O(h^{2r-2m-n}).$$

Relation (5.95) now follows from Lemma 5.2.4. As for $A_{3,2}$, it is a second order U-statistic (5.83) with $W_i = (Y_i, X_i, Z_i)$ and

$$a_N(W_i, W_j) = \frac{1}{2}\Big(\Delta F(X_i, Z_i, z)U_j + \Delta F(X_j, Z_j, z)U_i\Big)\widetilde{K}_{ij}K_{z,i}K_{z,j}.$$

Since $Ea_N(W_i, W_j) = 0$ and, by Proposition 4.1.1,

$$E\left(E(a_N(W_i, W_j)|W_i)^2\right) = \frac{1}{4}E\Big(K_{z,i}^2 E(\Delta F(X_j, Z_j, z)\widetilde{K}_{ij}K_{z,j}|X_i)^2\Big)$$

$$= \frac{\|K\|_2^2}{4h^m}E\Big(K_{2,z,i}E(\Delta F(X_j, Z_j, z)\widetilde{K}_{ij}K_{z,j}|X_i)^2\Big) = O(h^{2r-m}),$$

$$Ea_N(W_i, W_j)^2 \le E\Delta F(X_i, Z_i, z)^2\widetilde{K}_{ij}^2 K_{z,i}^2 K_{z,j}^2$$

$$= \frac{\|\widetilde{K}\|_2^2\|K\|_2^4}{h^{2m+n}} E\Delta F(X_i, Z_i, z)^2 \widetilde{K}_{2,ij} K_{2,z,i} K_{2,z,j} = O(h^{r-2m-n}),$$

relation (5.96) follows from Lemma 5.2.4 above. □

Lemma 5.2.8 *(Order of A_4) Under the assumptions of Theorem 5.2.1, we have $A_4 = O_p(1)$, where A_4 is defined by (5.72).*

PROOF: Arguing as in the proof of Lemma 5.2.10 below, one may show that A_4 is asymptotically normal. Hence, $A_4 = O_p(1)$. □

The next result was used a number of times in the proofs of Lemmas 5.2.5–5.2.8 above.

Lemma 5.2.9 *(Auxiliary orders) Under the assumptions of Theorem 5.2.1, we have*

$$\overline{\Delta F}(z) = O_p\left(h^r + \frac{h}{\sqrt{Nh^m}}\right), \quad \overline{U}(z) = O_p\left(\frac{1}{\sqrt{Nh^m}}\right), \tag{5.97}$$

where $\overline{\Delta F}(z)$ and $\overline{U}(z)$ are defined by (5.64) and (5.65), respectively, and

$$\overline{F}(z) = EF(X_i, z)\frac{p(X_i, z)}{\widetilde{p}(X_i)} + o_p(1), \quad \widehat{p}(z) = p(z) + o_p(1), \tag{5.98}$$

where $\overline{F}(z)$ and $\widehat{p}(z)$ are defined by (5.63) and (5.40), respectively, and $p(x, z)$, $\widetilde{p}(x)$ and $p(z)$ are the densities of (X_i, Z_i), X_i and Z_i, respectively.

PROOF: We consider only the case $G = 1$. By using Proposition 4.1.1, we have

$$E\overline{\Delta F}(z)^2 = \frac{1}{N} E\Delta F(X_i, Z_i, z)^2 K_h^2(z - Z_i) + \frac{N-1}{N}\left(E\Delta F(X_i, Z_i, z)K_h(z - Z_i)\right)^2$$

$$= \frac{\|K\|_2^2}{Nh^m} E\Delta F(X_i, Z_i, z)^2 K_{2,h}(z - Z_i) + \frac{N-1}{N}\left(E\Delta F(X_i, Z_i, z)K_h(z - Z_i)\right)^2$$

$$= O\left(\frac{h^r}{Nh^m}\right) + O(h^{2r})$$

and

$$E\widetilde{U}(z)^2 = \frac{1}{N}EU_i^2K_h^2(z - Z_i) = \frac{\|K\|_2^2}{Nh^m}EK_{2,h}(z - Z_i) = O\left(\frac{1}{Nh^m}\right),$$

which shows the two relations in (5.97). To show the first relation in (5.98), observe first that

$$E\left(\overline{F}(z) - EF(X_i, z)\frac{p(X_i, z)}{\widetilde{p}(X_i)}\right)^2$$

$$= E\overline{F}(z)^2 - 2E\overline{F}(z)EF(X_i, z)\frac{p(X_i, z)}{\widetilde{p}(X_i)} + \left(EF(X_i, z)\frac{p(X_i, z)}{\widetilde{p}(X_i)}\right)^2. \tag{5.99}$$

Since, by using Proposition 4.1.1,

$$E\overline{F}(z)^2 = \frac{1}{N}EF(X_i, z)^2K_h^2(z - Z_i) + \frac{N-1}{N}(EF(X_i, z)K_h(z - Z_i))^2$$

$$= O\left(\frac{1}{Nh^m}\right) + \left(\int_{\mathbb{R}^n} F(x_i, z)p(x_i, z)dx_i\right)^2 + o(1) = \left(EF(X_i, z)\frac{p(X_i, z)}{\widetilde{p}(X_i)}\right)^2 + o(1)$$

and

$$E\overline{F}(z) = EF(X_i, z)K_h(z - Z_i) = EF(X_i, z)\frac{p(X_i, z)}{\widetilde{p}(X_i)} + o(1),$$

the first relation in (5.98) follows from (5.99). The second relation in (5.98) can be proved in a similar way. □

We now prove an asymptotic normality result (5.77) used in the proof of Theorem 5.2.1.

Lemma 5.2.10 *Under the assumptions and with the notation of Theorem 5.2.1 and its proof, we have*

$$\widehat{V}(z)\widehat{p}(z)^2\frac{h^{m+n/2}}{N}\sum_{i \neq j}^{N}\widetilde{U}_i\widetilde{U}_j'\widetilde{K}_{ij}K_{z,i}K_{z,j} \xrightarrow{d} \mathcal{Z}_{G-L(z)}. \tag{5.100}$$

PROOF: Setting $t = G - L(z)$ and denoting $\widetilde{U}_i = (\widetilde{U}_{i1}, \ldots, \widetilde{U}_{it})'$, we can write the sum in (5.100) without the multiplicative term $\widehat{V}(z)\widehat{p}(z)^2$ as

$$\frac{h^{m+n/2}}{N} \sum_{i \neq j}^{N} \widetilde{U}_i \widetilde{U}_j' \widetilde{K}_{ij} K_{z,i} K_{z,j} = \Big(A_{pq}(N) \Big)_{p,q=1,\ldots,t},$$

where

$$A_{pq}(N) = \frac{h^{m+n/2}}{N} \sum_{i \neq j}^{N} \widetilde{U}_{ip} \widetilde{U}_{jq} \widetilde{K}_{ij} K_{z,i} K_{z,j}.$$

We will first show that, for fixed p and q,

$$A_{pq}(N) \xrightarrow{d} \mathcal{N}(0, \sigma_{pq}^2 V(z)^{-2} p(z)^{-4}), \tag{5.101}$$

where $V(z)$ is defined in the beginning of Section 5.2.2 and

$$\sigma_{pq}^2 = \begin{cases} 1 & \text{if } p = q, \\ 1/2 & \text{if } p \neq q. \end{cases}$$

Then, by Lemma 5.2.12 below, (5.101) shows that the convergence (5.100) holds componentwise. In order to show (5.101), we will follow the proof of Theorem 4.5 in White and Hong [101] (see also Lemma B.2 in Donald [28]). Since $\widetilde{U}_i$ can be expressed in terms of $W_i = (Y_i, X_i, Z_i)$, we can write

$$A_{pq}(N) = \sum_{i \neq j}^{N} \widetilde{a}_N(W_i, W_j) = \sum_{1 \leq i < j \leq N} a_N(W_i, W_j),$$

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

$$a_N(W_i, W_j) = \tilde{a}_N(W_i, W_j) + \tilde{a}_N(W_j, W_i).$$

Observe that, for $i < j$, $E(a_N(W_i, W_j)|W_i) = 0$. Hence, by Proposition 3.2 in de Jong [22], convergence (5.101) holds if (1) $\text{Var}(A_{pq}(N)) \to \sigma_{pq}^2$, and (2) $G_{N,i} = o(\text{Var}(A_{pq}(N))^2) = o(1)$ for $i = 1, 2, 4$, where with the notation $a_{ij} = Ea_N(W_i, W_j)$,

$$G_{N,1} = \sum_{1 \le i < j \le N} Ea_{ij}^4,$$

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

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

To show part (1), observe first that, by using the notation $\widetilde{K}_{2,ij} = \widetilde{K}_{2,h}(X_i - X_j)$ and $K_{2,z,i} = K_{2,h}(z - Z_i)$,

$$\text{Var}(A_{pq}(N)) = 2\sigma_{pq}^2 \frac{h^{2m+n}}{N^2} \sum_{i \ne j}^{N} E\widetilde{U}_{ip}^2 \widetilde{U}_{jq}^2 \widetilde{K}_{ij}^2 K_{z,i}^2 K_{z,j}^2$$

$$= 2\sigma_{pq}^2 h^{2m+n} E\widetilde{K}_{ij}^2 K_{z,i}^2 K_{z,j}^2 + o(1) = \sigma_{pq}^2 2\|\widetilde{K}\|_2^2 \|K\|_2^4 E\widetilde{K}_{2,ij} K_{2,z,i} K_{2,z,j} + o(1).$$

Since, by using Proposition 4.1.1, $E\widetilde{K}_{2,ij} K_{2,z,i} K_{2,z,j} = \int_{\mathbb{R}^n} p(x_i, z)^2 dx_i + o(1)$, we obtain that

$$\text{Var}(A_{pq}(N)) = \sigma_{pq}^2 2\|\widetilde{K}\|_2^2 \|K\|_2^4 \int_{\mathbb{R}^n} p(x_i, z)^2 dx_i + o(1) = \sigma_{pq}^2 V(z)^{-2} p(z)^{-4}.$$

As for part (2), observe that, by using the notation $\widetilde{K}_{4,ij} = \widetilde{K}_{4,h}(X_i - X_j)$, $K_{4,z,i} = K_{4,h}(z - Z_i)$ and Proposition 4.1.1,

$$G_{N,1} \le \text{const} \frac{h^{4m+2n}}{N^4} \sum_{i \ne j} E\widetilde{K}_{ij}^4 K_{z,i}^4 K_{z,j}^4$$

$$= \frac{\text{const}}{N^2h^{2m+n}} E\widetilde{K}_{4,ij}K_{4,z,i}K_{4,z,j} + o(1) = \frac{\text{const}}{N^2h^{2m+n}} \int_{\mathbb{R}^n} p(x_i,z)^2 dx_i + o(1) = o(1).$$

Similarly,

$$G_{N,2} \leq \text{const}\frac{h^{4m+2n}}{N^4} \sum_{i\neq j\neq l} E\widetilde{K}_{ij}^2 K_{z,i}^2 K_{z,j}^2 \widetilde{K}_{il}^2 K_{z,i}^2 K_{z,l}^2$$

$$= \frac{\text{const}}{Nh^m} E\widetilde{K}_{2,ij}\widetilde{K}_{2,il}K_{4,z,i}K_{2,z,j}K_{2,z,l} + o(1) = \frac{\text{const}}{Nh^m}\int_{\mathbb{R}^n} p(x_i,z)^3 dx_i + o(1) = o(1)$$

and

$$G_{N,4} \leq \text{const}\frac{h^{4m+2n}}{N^4} \sum_{i\neq j\neq l\neq k} E\widetilde{K}_{ij}K_{z,i}K_{z,j}\widetilde{K}_{ik}K_{z,i}K_{z,k}\widetilde{K}_{lj}K_{z,l}K_{z,j}\widetilde{K}_{lk}K_{z,l}K_{z,k}$$

$$= \text{const}\frac{h^{4m+2n}}{N^4} \sum_{i\neq j\neq l\neq k} E\widetilde{K}_{ij}\widetilde{K}_{ik}\widetilde{K}_{lj}\widetilde{K}_{lk}K_{z,i}^2K_{z,j}^2K_{z,k}^2K_{z,l}^2$$

$$= \text{const}\, h^{2n} E\widetilde{K}_{ij}\widetilde{K}_{ik}\widetilde{K}_{lj}\widetilde{K}_{lk}K_{2,z,i}K_{2,z,j}K_{2,z,k}K_{2,z,l} + o(1)$$

$$= \text{const}\, h^{2n}\left(\int_{\mathbb{R}^n} p(x_i,z)^4 dx_i\right) + o(1) = o(1).$$

Arguing in a similar way as above, one may show that, for any $c_i \in \mathbb{R}$, $p_i, q_i \in \{1,\ldots,t\}$, a linear combination $\sum_{i=1}^d c_i A_{p_iq_i}(N)$ is asymptotically normal with the limiting variance $\sigma(\boldsymbol{p},\boldsymbol{q})^2$ characterized by

$$\text{Var}\left(\sum_{i=1}^d c_i A_{p_iq_i}(N)\right) \to \sigma(\boldsymbol{p},\boldsymbol{q})^2.$$

Since $EA_{pq}(N)A_{p'q'}(N) = 0$ for different pairs (p,q) and (p',q'), we conclude that $\sigma(\boldsymbol{p},\boldsymbol{q})^2 = \sigma_{p_1q_1}^2 + \cdots + \sigma_{p_dq_d}^2$, which, together with the convergence (5.101), shows that (5.100) holds. □

The following result was used in the proof of Theorem 5.2.1 to replace the variance-covariance matrix Σ by its estimator $\widehat{\Sigma}$.

Lemma 5.2.11 *Under the assumptions of Theorem 5.2.1, we have* $\widehat{\Sigma} = \Sigma + \delta B$ *with* $B = o_p(1)$.

PROOF: As shown in the proof of Lemma 2 in Donald [28], pp. 126–127,

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

Then, by using the assumptions of Theorem 5.2.1 and since $\delta^{-1} = \sqrt{Nh^{m+n/2}}$, we obtain that $\widehat{\Sigma} = \Sigma + \delta B$ with $B = o_p(1)$. □

Finally, we prove that $\widehat{V}(z)$, defined by (5.45), is a consistent estimator for $V(z)$.

Lemma 5.2.12 *Under the assumptions of Theorem 5.2.1, we have*

$$\frac{1}{N}\sum_{i=1}^{N} \widehat{p}(X_i, Z_i)K_h(z - Z_i) \xrightarrow{P} \int_{\mathbb{R}^n} p(x_i, z)^2 dx_i \tag{5.102}$$

and hence $\widehat{V}(z) \to_p V(z)$.

Remark 5.2.8 The idea behind (5.102) can be expressed as follows

$$\frac{1}{N}\sum_{i=1}^{N} \widehat{p}(X_i, Z_i)K_h(z - Z_i) \approx Ep(X_i, Z_i)K_h(z - Z_i)$$

$$= \int_{\mathbb{R}^n}\int_{\mathbb{R}^m} p(x_i, z_i)^2 K_h(z - Z_i)dx_i dz_i = \int_{\mathbb{R}^n} p(x_i, z)^2 dx_i.$$

PROOF: By using (4.7), we can write the sum in (5.102) as

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \widetilde{K}_h(X_i - X_j)K_h(z - Z_i)K_h(Z_i - Z_j) = \frac{\widetilde{K}(0)K(0)}{Nh^{m+n}}\left(\frac{1}{N}\sum_{i=1}^{N} K_h(z - Z_i)\right)$$

$$+\frac{1}{N^2}\sum_{i\neq j}^{N} \widetilde{K}_h(X_i - X_j)K_h(z - Z_i)K_h(Z_i - Z_j) =: I_1 + I_2.$$

We have

$$\widehat{p}(z) := \frac{1}{N}\sum_{i=1}^{N} K_h(z - Z_i) \to p(z)$$

in probability and hence $I_1 \to_p 0$ since $Nh^{m+n} \to \infty$. Arguing as in the proof of Lemma 5.2.5, one can show that

$$I_2 = E\widetilde{K}_h(X_i - X_j)K_h(z - Z_i)K_h(Z_i - Z_j) + o(1) = \int_{\mathbb{R}^n}\int_{\mathbb{R}^m}\int_{\mathbb{R}^n}\int_{\mathbb{R}^m} dx_i dz_i dx_j dz_j \cdot$$

$$\cdot\, p(x_i, z_i)p(x_j, z_j)\widetilde{K}_h(x_i - x_j)K_h(z - z_i)K_h(z_i - z_j) + o(1) = \int_{\mathbb{R}^n} p(x_i, z)^2 dx_i + o(1).$$

This proves (5.102). □

5.3 Estimation of rank

In Sections 5.1 and 5.2, we studied local tests for (SPF) and (NP) models, namely, the problems to test $H_0 : \mathrm{rk}\{\Theta(z)\} \le L$ against $H_1 : \mathrm{rk}\{\Theta(z)\} > L$ in the case of (SPF) model and to test $H_0 : \mathrm{adrk}\{F(\cdot, z)\} \le L$ against $H_1 : \mathrm{adrk}\{F(\cdot, z)\} > L$ in the case of (NP) model, where z and L are fixed. In this section, we use these local tests to estimate the true ranks $\mathrm{rk}\{\Theta(z)\}$ and $\mathrm{adrk}\{F(\cdot, z)\}$ in (SPF) and (NP) models, respectively. 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. The model selection criteria is only mentioned at the end of this section because it has been found to perform poorly in small samples (see Cragg and Donald [20]).

To describe the sequential procedure, consider for example the problem of determining the true rank $\mathrm{rk}\{\Theta(z)\}$ in (SPF) model by using the minimum-χ^2 test of Section 5.1.2. (One may use any other hypothesis test for the rank of a matrix, for example, the LDU based test of Section 5.1.1.) Let $\widehat{T}(L, z)$, $L = 1, \ldots, G$, denote the value of the minimum-χ^2 statistic defined by (5.9). The sequential testing is based on the following procedure:

first, for increasing integer values $L = 1, \ldots, G$, by using the statistic $\widehat{T}(L, z)$, test the hypothesis $H_0 : \mathrm{rk}\{\Theta(z)\} \leq L$ against the alternative $H_1 : \mathrm{rk}\{\Theta(z)\} > L$ at a given level of significance α, that is, determine whether

$$\widehat{T}(L, z) \leq \chi^2_\alpha((G - L)(d - L)), \tag{5.103}$$

where $\chi^2_\alpha((G - L)(d - L))$ is the minimum ξ such that $P(\chi^2((G - L)(d - L)) > \xi) = \alpha$; second, stop at the first value of L which does not reject the hypothesis H_0, that is, when (5.103) holds. Denote this value of L by $\widehat{L}(z)$. In view of Theorem 5.1.3, $\widehat{L}(z)$ will not be a consistent estimator of $\mathrm{rk}\{\Theta(z)\}$ because, as N increases, $\widehat{L}(z)$ will overestimate $\mathrm{rk}\{\Theta(z)\}$ with probability $\alpha > 0$ (which is a fixed significance level). The idea then, proposed by Pötscher [81] in the context of determining the order of an autoregressive moving average (ARMA) model and by Bauer, Pötscher and Hackl [11] in the context of model selection is to make α depend on N, that is, $\alpha = \alpha(N)$, and let $\alpha(N) \to 0$ as $N \to \infty$. In this way, one can obtain a consistent estimator $\widehat{L}(z)$ of $\mathrm{rk}\{\Theta(z)\}$.

Theorem 5.3.1 *(Consistency of rank estimator in* (SPF) *model) With the above notation and under the assumptions of Theorem 5.1.3, we have* $\widehat{L}(z) \to_p \mathrm{rk}\{\Theta(z)\}$ *as long as* $\alpha(N) \to 0$ *and* $-\ln\alpha(N)/Nh^m \to 0$.

Proof: The proof is similar to that of Theorem 3 in Donald [28] or Theorem 5.2 in Robin and Smith [89]. Let $\mathcal{A}_L$ denote the event that the null hypothesis $H_0 : \mathrm{rk}\{\Theta(z)\} \leq L$ is rejected by using the minimum-χ^2 statistic $\widehat{T}(L, z)$ at the significance level $\alpha = \alpha(N)$. Then, we have

$$P(\widehat{L} = L) = P\left(\mathcal{A}_1 \cap \cdots \cap \mathcal{A}_{L-1} \cap \mathcal{A}_L^c\right), \tag{5.104}$$

where $\mathcal{A}_L^c$ denotes the complement of $\mathcal{A}_L$. Let $\chi^2_{\alpha(N)}((G - L)(d - L))$ be the minimum ξ such that $P(\chi^2((G - L)(d - L)) \geq \xi) = \alpha(N)$. By Theorem 5.8 in Pötscher [81], we have $\chi^2_{\alpha(N)}((G-L)(d-L)) \to \infty$ if $\alpha(N) \to 0$, and $\chi^2_{\alpha(N)}((G-L)(d-L))/Nh^m \to 0$ if $-\ln\alpha(N)/Nh^m \to 0$. Then, for any $L < \mathrm{rk}\{\Theta(z)\}$, we obtain from (5.104) that

$$P(\widehat{L} = L) \leq P(\mathcal{A}_L^c) = 1 - P(\mathcal{A}_L) = 1 - P(\widehat{T}(L,z) > \chi^2_{\alpha(N)}((G-L)(d-L)))$$

$$= 1 - P(\widehat{C}(\widehat{\Theta}, L) > \chi^2_{\alpha(N)}((G-L)(d-L))/Nh^m) \to 0,$$

as $N \to \infty$, by using $\widehat{C}(\widehat{\Theta}, L) \to C(\Theta_0, L) > 0$ (see the proof of Theorem 5.1.3) and $\chi^2_{\alpha(N)}((G-L)(d-L))/Nh^m \to 0$. Observe also that, by setting $L_0 = \mathrm{rk}\{\Theta(z)\}$,

$$P(\widehat{L} > L_0) \leq P(\mathcal{A}_{L_0}) = P(\widehat{T}(L_0, z) > \chi^2_{\alpha(N)}((G-L_0)(d-L_0))) \to 0,$$

by Theorem 5.1.3, *(ii)*, $\chi^2_{\alpha(N)}((G-L_0)(d-L_0)) \to \infty$ and Theorem 2.1, *(iii)*, in Billingsley [14]. The result of the theorem now follows from the two convergence above. □

In the case of (NP) model, the sequential testing procedure is similar. Let $\widehat{T}_1(L,z)$ be the test statistic (5.49) used for local tests in (NP) model. Let also

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

where $\mathcal{N}_{\alpha(N)}(0,1)$ is the smallest ξ such that $P(\mathcal{N}(0,1) > \xi) = \alpha(N)$, be the minimum L which does not reject the null hypothesis H_0 : $\mathrm{adrk}\{F(\cdot,z)\} \leq L$ at a significance level $\alpha(N)$. According to the next result, $\widehat{L}(z)$ is a consistent estimator of $\mathrm{adrk}\{F(\cdot,z)\}$, provided the specified conditions on the significance levels $\alpha(N)$ hold.

Theorem 5.3.2 *(Consistency of rank estimator in* (NP) *model) With the above notation and under the assumptions of Theorem 5.2.1, we have* $\widehat{L}(z) \to_p \mathrm{adrk}\{F(\cdot,z)\}$ *as long as* $\alpha(N) \to 0$ *and* $(-\ln\alpha(N))^{1/2}/Nh^{m+n/2} \to 0$.

PROOF: The proof of the theorem is analogous to that of Theorem 5.3.1 above by using Theorem 5.2.2 and Lemma 5.3.1 below. □

Lemma 5.3.1 *Let $f(T) \to \infty$, as $T \to \infty$, and $\xi(T) \to \infty$ be such that $P(\mathcal{N}_{\alpha(T)}(0,1) > \xi(T)) = \alpha(T)$. Then, $\xi(T)/f(T) \to 0$ if and only if $(-\ln \alpha(T))^{1/2}/f(T) \to 0$.*

PROOF: The proof is elementary. There are $a, b > 0$ such that

$$\exp\{-a\xi(T)^2\} \leq P\left(\mathcal{N}_{\alpha(T)}(0,1) > \xi(T)\right) = \alpha(T) \leq \exp\{-b\xi(T)^2\}$$

for large enough T. Hence, $a^{1/2}\xi(T) \leq (-\ln \alpha(T))^{1/2} \leq b^{1/2}\xi(T)$ and

$$a^{1/2}\frac{\xi(T)}{f(T)} \leq \frac{(-\ln \alpha(T))^{1/2}}{f(T)} \leq b^{1/2}\frac{\xi(T)}{f(T)},$$

for large T, from which the result of the lemma follows. □

Finally, we briefly describe ideas behind estimation of the true rank by model selection criteria. Consider, for example, the case of (SPF) model and let $\widehat{T}(L,z)$ be the minimum-χ^2 statistic (5.9) used in local tests for (SPF) model. The model selection criteria is based on the quantity

$$S(L) = \frac{\widehat{T}(L,z)}{f(N)} - g(L), \tag{5.105}$$

where f and g are some functions. The function g, which is typically a strictly decreasing function in L, corresponds to a penalty factor. The behavior of the function f differs from one situation to another. According to the model selection criteria, the estimator $\widehat{L}(z)$ of $\mathrm{rk}\{\Theta(z)\}$ is then defined as the integer L for which the function $S(L)$ in (5.105) takes its minimal value. The goal is to find conditions on the functions f and g for which the estimator $\widehat{L}(z)$ is a (strongly or weakly) consistent estimator of $\mathrm{rk}\{\Theta(z)\}$. Such necessary conditions were found by Cragg and Donald [20] in a context similar to our. These authors indicated, however, that the rank estimators based on model selection criteria perform poorly on small sample data. We decided therefore not to include here results on these rank estimators. Despite this negative fact, let us conclude by saying that some model selection criteria, for example, Akaike information criteria (AIC) of Akaike [6] or Bayesian information criteria (BIC) of Schwarz [94], are widely and successfully used in other areas of Statistics.

Chapter 6

Discussion on Global Tests

In this chapter, we focus on estimation and testing of global ranks of demand systems given by non-parametric and semi-parametric factor models. We will *not* provide here global tests. We will instead introduce some global test statistics, explain the difficulties in establishing their limit laws and outline some possible approaches to solution. The chapter thus lays a path and points to directions for the future research.

6.1 Semi-parametric factor model

Consider first the problem of testing for the global rank in a semi-parametric factor (SPF) model discussed in Section 3.1, namely, the problem to test $H_0 : \max_z \mathrm{rk}\{\Theta(z)\} \leq L$, against the alternative $H_1 : \max_z \mathrm{rk}\{\Theta(z)\} > L$, for some fixed L. Since we are now interested in the maximum of local ranks over all possible values of z, it is natural to introduce a test statistic which is the maximum of some statistics used for local tests. Suppose, for example, that local test statistics are minimum-χ^2 statistics $\widehat{C}(\widehat{\Theta}(z), L)$ discussed in Section 5.1.2. Then, consider the global test statistic

$$\widehat{C}_{\max}(L) = \max_z \widehat{C}(\widehat{\Theta}(z), L)$$
$$= Nh^m \|K\|_2^{-2} \max_z \min_{\mathrm{rk}\{\Theta\} \leq L} \mathrm{vec}(\widehat{\Theta}(z) - \Theta)'(\widehat{Q}(z)^{-1} \otimes \widehat{\Sigma})^{-1} \mathrm{vec}(\widehat{\Theta}(z) - \Theta).$$

Observe that, under the hypothesis H_1, there is z_0 such that $\mathrm{rk}\{\Theta(z_0)\} > L$. Then, by Theorem 5.1.3,(i), $\widehat{C}(\widehat{\Theta}(z_0), L) \to +\infty$ in probability and hence $\widehat{C}_{\max}(L) \to +\infty$ in probability as well. On the other hand, under the hypothesis H_0, we have $\mathrm{rk}\{\Theta(z)\} \le L$ for all values of z. Then, in view of Theorem 5.1.3,(*ii*)–(*iii*), one would expect that the statistic $\widehat{C}_{\max}(L)$ has a non-degenerate limit distribution. The goal then would be to find this limit distribution and, by using it, to construct the corresponding global tests. This goal, however, turns out to be nontrivial to achieve. The difficulties lie in the following theorem. Its implications on global tests are further discussed after the proof.

Theorem 6.1.1 *Let $z_1, ..., z_n$ be fixed different values of z. Suppose that the assumptions of Theorem 4.2.2 hold for all $z_1, \ldots, z_n$. Then, if $L_i := \mathrm{rk}\{\Theta(z_i)\} \le L$ for $i = 1, \ldots, n$, we have*

$$\left(\widehat{C}(\widehat{\Theta}(z_1), L), \ldots, \widehat{C}(\widehat{\Theta}(z_n), L)\right)$$

$$\xrightarrow{d} \left(\sum_{j=1}^{G-L} \lambda_j(\mathcal{Y}^2_{1,(G-L_1)\times(d-L_1)}), \ldots, \sum_{j=1}^{G-L} \lambda_j(\mathcal{Y}^2_{n,(G-L_n)\times(d-L_n)})\right), \tag{6.1}$$

where $\mathcal{Y}^2_{i,(G-L_i)\times(d-L_i)} = \mathcal{Y}_{i,(G-L_i)\times(d-L_i)}\mathcal{Y}'_{i,(G-L_i)\times(d-L_i)}$, $\lambda_j(\mathcal{Y}^2_{i,(G-L_i)\times(d-L_i)})$, $j = 1, \ldots, (G-L_i)$, are the eigenvalues of $\mathcal{Y}^2_{i,(G-L_i)\times(d-L_i)}$ in increasing order and $\mathcal{Y}_{i,(G-L_i)\times(d-L_i)}$, $i = 1, \ldots, n$, are independent $(G-L_i)\times(d-L_i)$ matrices with independent $\mathcal{N}(0,1)$ entries. In particular, the statistics $\widehat{C}(\widehat{\Theta}(z_i), L)$, $i = 1, \ldots, n$, are asymptotically independent.

Proof: The proof is similar to that of Theorem 5.1.5. Letting $\widehat{Y}_i = \sqrt{Nh^m}C'_{i,G-L_i}(\widehat{\Theta}(z_i) - \Theta(z_i))D_{i,d-L_i}$, $i = 1, \ldots, n$, where $C_{i,G-L_i}$ and $D_{i,d-L_i}$ are defined as in (5.23) and (5.24) for the matrix $\Theta(z_i)$, we have by Lemma 5.1.3 that the test statistics $\widehat{C}(\widehat{\Theta}(z_i), L)$, $i = 1, \ldots, n$, are asymptotically equivalent to the sum of the first $G-L$ eigenvalues of the matrices $\widehat{Y}_i\widehat{Y}'_i$. As in the proof of Theorem 5.1.5, we have $\widehat{Y}_i\widehat{Y}'_i \to_d \mathcal{Y}^2_{(G-L_i)\times(d-L_i)}$. Moreover, by Theorem 4.2.3, since matrices $\sqrt{Nh^m}(\widehat{\Theta}(z_i) - \Theta(z_i))$, $i = \ldots, n$, are asymptotically independent, we obtain that the matrices $\widehat{Y}_i\widehat{Y}'_i$, $i = \ldots, n$, are asymptotically independent as well. This concludes the proof. $\square$

According to Theorem 6.1.1, the minimum-χ^2 test statistics are asymptotically independent for different values of z. This asymptotic independence implies, in particular, that the global test statistic $\widehat{C}_{\max}(L)$ does *not* have a non-degenerate limit distribution under the assumption $\sup_z \mathrm{rk}\{\Theta(z)\} \leq L$. In fact, according to the next result, $\widehat{C}_{\max}(L) \to +\infty$ in probability.

Corollary 6.1.1 *Suppose that the assumptions of Theorem 4.2.2 hold for all values of z. Then, under the assumption* $\sup_z \mathrm{rk}\{\Theta(z)\} \leq L$, *we have*

$$\widehat{C}_{\max}(L) \xrightarrow{P} +\infty. \tag{6.2}$$

PROOF: We need to show that $P(\widehat{C}_{\max}(L) \leq c) \to 0$ for all $c > 0$. Since $\max_z \mathrm{rk}\{\Theta(z)\} \leq L$, we may suppose that there are different z_n, $n \geq 1$, such that $\mathrm{rk}\{\Theta(z_n)\} = L_0 \leq L$ for some value L_0. Then, by using Theorem 6.1.1 above and Theorem 2.1, (*iii*), in Billingsley [14], we have that, for any $n \geq 1$,

$$\overline{\lim}\, P(\widehat{C}_{\max}(L) \leq c) \leq \overline{\lim}\, P\left(\bigcap_{k=1}^{n} \{\widehat{C}(\Theta(z_k), L) \leq c\}\right)$$

$$\leq \prod_{k=1}^{n} P\left(\sum_{j=1}^{G-L} \lambda_j(\mathcal{Y}^2_{k,(G-L_0)\times(d-L_0)}) \leq c\right) \leq \left(P(\lambda_1(\mathcal{Y}^2_{1,(G-L_0)\times(d-L_0)}) \leq c)\right)^n.$$

Since n is an arbitrary integer, it is enough to show that $P(\lambda_1(\mathcal{Y}^2_{m\times n}) > c) > 0$ for any $c > 0$ and $m, n \geq 1$. By using Theorem 13 in Magnus and Neudecker [73], we have

$$\lambda_1(\mathcal{Y}^2_{m\times n}) = \min_{x'x=1} x'\mathcal{Y}_{m\times n}\mathcal{Y}'_{m\times n}x.$$

Then, denoting the first element of the vector $x'\mathcal{Y}_{m\times n}$ by $x_1\xi_1 + \cdots + x_m\xi_m$ where $x = (x_1, \ldots, x_m)'$ and ξ_i, $i = 1, \ldots, n$ are i.i.d. $\mathcal{N}(0, 1)$ random variables, we obtain that

$$P(\lambda_1(\mathcal{Y}^2_{m\times n}) > c) \geq P(\min_{x'x=1} (x_1\xi_1 + \cdots + x_n\xi_n)^2 > c).$$

One can verify by using Lagrange multiplier and Kühn-Tücker conditions that

$$\min_{x'x=1}(x_1\xi_1+\cdots+x_n\xi_n)^2=\xi_1^2+\cdots+\xi_n^2\stackrel{d}{=}\chi^2(n),$$

with the minimizing $x_i=\xi_i(\xi_1^2+\cdots+\xi_n^2)^{-1/2}$, $i=1,\ldots,n$, or $x_i=-\xi_i(\xi_1^2+\cdots+\xi_n^2)^{-1/2}$, $i=1,\ldots,n$. Hence,

$$P(\lambda_1(\mathcal{Y}_{m\times n}^2)>c)\geq P(\chi^2(n)>c)>0. \qquad \square$$

According to Corollary 6.1.1 and the discussion preceding Theorem 6.1.1, the maximum test statistic $\widehat{C}_{\max}(L)\to_p+\infty$ under both null hypothesis $H_0:\max_z\mathrm{rk}\{\Theta(z)\}\leq L$ and its alternative $H_1:\max_z\mathrm{rk}\{\Theta(z)\}>L$. This may appear as a major problem because there is no longer simple distinction between the behavior of test statistic under the two hypotheses. We believe, however, that there is a way out. The idea is in the spirit of that used in connection to global measures of density function estimates which we briefly discuss next.

Suppose that $p(z)$ is a density function and that

$$\widehat{p}(z)=\frac{1}{N}\sum_{i=1}^{N}K_h(z-Z_i)$$

is its kernel based estimator. It is well-known (see, for example, Härdle [49] or Pagan and Ullah [77]) that, under suitable conditions, the estimator $\widehat{p}(z)$ is asymptotically normal, namely,

$$\frac{\sqrt{Nh}}{p(z)^{1/2}\|K\|_2}(\widehat{p}(z)-p(z))\xrightarrow{d}\mathcal{N}(0,1).$$

Moreover, as it is the case with the minimum-χ^2 statistic $\widehat{C}(\widehat{\Theta}(z),L)$, the estimators $\widehat{p}(z)$'s can be shown to be asymptotically independent for different values of z. Despite this fact, there are a number of results in the statistical literature on the asymptotic behavior of the maximum deviation measure of the estimator $\widehat{p}(z)$, namely, the quantity

$$\widehat{T}_{\max} = \max_{z} \frac{\sqrt{Nh}}{p(z)^{1/2}\|K\|_2} |\widehat{p}(z) - p(z)|. \tag{6.3}$$

(Such quantity is interesting in connection to goodness-of-fit tests for density estimates.) In a fundamental work related to the global measure (6.3), Bickel and Rosenblatt [13] showed that, under suitable conditions, for some constants c_0, c_1, c_2, c_3,

$$(c_1 \ln N)^{1/2} \left(\widehat{T}_{\max} - (c_1 \ln N)^{1/2} - \frac{c_2 + c_3 \ln \ln N}{(c_1 \ln N)^{1/2}} \right) + c_0 \xrightarrow{d} \Lambda, \tag{6.4}$$

where Λ is a random variable with the distribution function

$$P(\Lambda \leq z) = e^{-e^{-z}}, \quad z \in \mathbb{R}. \tag{6.5}$$

The result (6.4) shows in particular that, as $N \to \infty$, the measure $\widehat{T}_{\max}$ concentrates around the points $(c_1 \ln N)^{1/2} + (c_2 + c_3 \ln \ln N)(c_1 \ln N)^{-1/2}$ which tend to infinity. When centered around these points and properly normalized (interestingly the normalization $(c_1 \ln N)^{1/2} \to \infty$), the measure $\widehat{T}_{\max}$ converges in distribution to the limit law Λ shifted by c_0.

The basic idea behind the result (6.4) is as follows. Since the estimates $\widehat{p}(z)$ are asymptotically normal and asymptotically independent for different values of z, one may think of $\widehat{T}_{\max}$ as the maximum of $M = M(N)$ independent $|\mathcal{N}(0,1)|$ random variables. It is well-known that a maximum of M i.i.d. random variables, when properly centered and normalized, can have only one of the three types of limit distributions: Fréchet, Weibull or Gumbell (see, for example, Embrechts, Klüppelberg and Mikosch [29], Resnick [87] or Leadbetter, Lindgren and Rootzén [57]). The distribution which appears in (6.5) is Gumbell. Its maximum domain of attraction contains most of the light-tailed distributions like Normal, Gamma, Lognormal and others. For example, if $\widehat{T}_{\max}$ is the maximum of N i.i.d.

$\mathcal{N}(0,1)$ random variables, the relation (6.4) holds with $c_0 = 0$, $c_1 = 2$, $c_2 = -(\ln 4\pi)/2$ and $c_3 = -1/2$ (see, for example, p. 147 in Embrechts et al. [29]).

Although the above provides idea behind the result (6.4), it is by no means easy to establish. The approach used by Bickel and Rosenblatt [13] essentially involves two steps. First, the authors show that, in the limit, the difference $\widehat{p}(z) - p(z)$ can be approximated by a Gaussian process (through so-called Komlós-Major-Tusnády [56] type approximations). Second, the result (6.4) is then established by considering the maximum of (the absolute value) of this Gaussian process which is easier to deal with. There is by now a substantial amount of literature on asymptotics of maxima of Gaussian processes (see, for example, Leadbetter et al. [57]).

Turning back to the maximum test statistic $\widehat{C}_{\max}(L)$ for the semi-parametric factor model, we expect that, when properly normalized and when $\max_z \operatorname{rank}\{\Theta(z)\} \le L$, it will also converge (at least be dominated by) a Gumbell distribution in the limit. The heuristics behind this statement are as follows. Let $L_z = \operatorname{rank}\{\Theta(z)\}$ and suppose that one works under the assumption $H_0 : \max_z L_z \le L$. Observe then that, by Theorem 5.1.4,

$$\widehat{C}_{\max}(L) = \max_z \widehat{C}(\widehat{\Theta}(z), L) \le \max_z \widehat{C}(\widehat{\Theta}(z), L_z).$$

The advantage of replacing L by L_z in $\widehat{C}(\widehat{\Theta}(z), L)$ is that we can now approximate the statistic $\widehat{C}(\widehat{\Theta}(z), L_z)$ as in the proof of Theorem 5.1.3 to obtain

$$\begin{aligned} \widehat{C}_{\max}(L) &\le \max_z \widehat{C}(\widehat{\Theta}(z), L_z) \\ &\approx c \max_z Nh^m \operatorname{vec}(\widehat{\Theta}(z) - \Theta(z))' W(z)^{-1/2} (I - S(z)) W(z)^{-1/2} \operatorname{vec}(\widehat{\Theta}(z) - \Theta(z)), \end{aligned} \tag{6.6}$$

where $S(z) = A(z)(A(z)'A(z))^{-1}A(z)$ with $A(z) = W(z)^{-1/2}B(z)$. We believe that one should be able to proceed now as in Bickel and Rosenblatt [13]. First, approximate $\sqrt{Nh^m}\operatorname{vec}(\widehat{\Theta}(z) - \Theta(z))'W(z)^{-1/2}$ by a Gaussian process and then work with the maximum of this Gaussian process squared. Since (6.6) can be thought as the maximum of i.i.d. $(\mathcal{N}(0,1))^2$ (more generally, χ^2) random variables and since the square of a normal random variable belongs to the maximum domain of attraction of a Gumbell distribution, we expect that the normalized (6.6) would converge to a Gumbell distribution as well. The established asymptotics for the bound $\widehat{C}_{\max}(L)$ and the fact that, $\widehat{C}_{\max}(L) \to_p +\infty$ at the rate Nh^m under the alternative hypothesis $H_1 : \max_z L_z > L$ (see the proof of Theorem 5.1.3), we hope, will allow to distinguish between the two test hypotheses in the limit.

More precisely, suppose that $a(N)\max_z \widehat{C}(\widehat{\Theta}(z), L_z) - b(N)$ converges to a Gumbell distribution Λ where $a(N), b(N)$ are some positive normalization and shift constants, and, for a given significance level α, let Λ_α be such that $P(\Lambda \geq \Lambda_\alpha) = \alpha$. Then, under the null hypothesis H_0, one would have

$$\begin{aligned} &\overline{\lim} P\Big(\widehat{C}_{\max}(L) \geq a(N)^{-1}(\Lambda_\alpha + b(N))\Big) \\ &\leq \overline{\lim} P\Big(a(N)\max_z \widehat{C}(\widehat{\Theta}(z), L_z) - b(N) \geq \Lambda_\alpha\Big) = P(\Lambda \geq \Lambda_\alpha) = \alpha. \end{aligned} \tag{6.7}$$

On the other hand, under the alternative hypothesis H_1, by using the fact that $\widehat{C}_{\max}(L) \to_p \infty$ at the rate Nh^m, we expect that

$$P\Big(\widehat{C}_{\max}(L) \geq a(N)^{-1}(\Lambda_\alpha + b(N))\Big) \to 1. \tag{6.8}$$

Based on (6.7) and (6.8), one then would not reject H_0 at a significance level α as long as $\widehat{C}_{\max}(L) \leq a(N)^{-1}(\Lambda_\alpha + b(N))$.

Remark 6.1.1 An alternative statistic for global tests in the semi-parametric factor model can be defined by integrating local statistics, namely,

$$\widehat{C}_{1,\mathrm{int}}(L) = \int \widehat{C}(\widehat{\Theta}(z), L)w(z)dz \tag{6.9}$$

or by the empirical version of (6.9) as

$$\widehat{C}_{2,\text{int}}(L) = \frac{1}{N}\sum_{i=1}^{N} \widehat{C}(\widehat{\Theta}(Z_i), L)\widetilde{w}(Z_i), \tag{6.10}$$

where $w(z) > 0$ and $\widetilde{w}(z) > 0$ are some weight functions. Global measures similar to (6.9) and (6.10) have been extensively studied in the statistical literature (see Bickel and Rosenblatt [13], Lii [70], Hall [46] or others). They are typically thought to be easier to work with than the corresponding tests based on the maxima. The basic idea behind the test statistic (6.9) or (6.10) is that, under the alternative hypothesis $H_1 : \max_z \text{rk}\{\Theta(z)\} > L$, one expects $\widehat{C}_{i,\text{int}} \to_p +\infty$, $i = 1, 2$. Under the hypothesis $H_0 : \max_z \text{rk}\{\Theta(z)\} \leq L$, since one is essentially summing (integrating) independent finite-variance random variables, one expects to get a normal distribution in the limit. This can be achieved, we believe, by using the (6.6) type approximations.

6.2 Non-parametric model

Consider now the problem of testing for the global adjusted rank in a non-parametric (NP) model discussed in Section 3.2, namely, the problem to test $H_0 : \max_z \text{adrk}\{F(\cdot, z)\} \leq L$, against the alternative $H_1 : \max_z \text{adrk}\{F(\cdot, z)\} > L$. Focus on the local test statistic $\widehat{T}_2(L, z)$ defined in (5.51) and consider the corresponding global test statistic

$$\widehat{T}_{2,\max}(L) = \max_z \widehat{T}_2(L, z). \tag{6.11}$$

As in Section 6.1, we expect that $\widehat{T}_2(L, z)$'s are asymptotically independent for different values of z and, as a consequence, that $\widehat{T}_2(L, z) \to_p +\infty$ under both hypotheses H_0 and H_1. Despite this fact, we believe that one can describe the asymptotics of the global test statistic $\widehat{T}_{2,\max}(L)$ by following the approach outlined in Section 6.1.

Recall from Section 5.2.3 that the statistic $\widehat{T}_2(L, z)$ can be expressed in the form of the minimum-χ^2 statistic as in (5.60). Setting $L_z = \text{adrk}\{F(\cdot, z)\}(= \text{rk}\{\Gamma_{w,z}\})$, $\widehat{W}(z) = \widehat{\Sigma} \otimes \widehat{\Sigma}$

and arguing as in the proof of Theorem 5.1.3, one obtains that, under the assumption $L \le \max_z L_z$,

$$\widehat{T}_{2,\max}(L) \le \max_z \widehat{T}_2(L_z, z)$$
$$\approx \max_z \widehat{V}(z)^2 N^2 h^{2m+n} \mathrm{vec}(\widehat{\Gamma}_{w,z} - \Gamma_{w,z})'(\widehat{W}(z)^{-1} - \widehat{W}(z)^{-1} B(z) \cdot$$
$$\cdot (B(z)'\widehat{W}(z)^{-1}B(z))^{-1}B(z)'\widehat{W}(z)^{-1})\mathrm{vec}(\widehat{\Gamma}_{w,z} - \Gamma_{w,z}). \quad (6.12)$$

One should be able to approximate now $Nh^{m+n/2}\mathrm{vec}(\widehat{\Gamma}_{w,z} - \Gamma_{w,z})$ by a Gaussian process, substitute it in the relation (6.12) and then work with the maximum of the square like transformation of this Gaussian process. At the end, we expect to obtain a Gumbell distribution as in the case of (SPF) model discussed in Section 6.1.

Remark 6.2.1 The discussion above is based on the local test statistics $\widehat{T}_2(L, z)$. How to deal with, say, the maximum of the alternative local test statistics $\widehat{T}_1(L, z)$ in (5.49) is still an open question. Note also that, as in (6.9) and (6.10), one may define a global test statistic for (NP) model as an integral or a sum of the local test statistics $\widehat{T}_2(L, z)$.

Chapter 7

Applications and Simulation Results

In this chapter, we apply (SPF) and (NP) models to estimate the local rank of a demand system from economic data. In Section 7.1, we describe the data set that we use. Applications of the two models for rank estimation can be found in Sections 7.2 and 7.3. Finally, in Section 7.4, we perform some Monte Carlo simulations to support the observations made in applications.

7.1 Description of the data set

The data set that we will use contains information on expenditures, total income and prices faced by a number of households across the United States. Expenditures and total income are taken from Interview Survey Public-Use Tapes of the Consumer Expenditure Surveys data for the United States (the CEX data, in short). The CEX data set and the selection procedure are described in greater detail below. The CEX data, however, contains no information on prices faced by different households. Prices are drawn from the American Chamber of Commerce Research Association price data for various cities across the United States (the ACCRA data, in short). We are able to associate these prices to households by using some location variables reported in the CEX data set as matching variables. The ACCRA data and the matching procedure are also described in greater detail below. Our matching procedure is a more refined version of that found in Nicol [76] because we use confidential information to attain a better match of the CEX and ACCRA data sets.

The CEX data set is the major source of information on household expenditures and types reported by the Bureau of Labor Statistics in the United States every quarter since 1980. In our study, we use the CEX data set of the first quarter of 2000 available at Inter-university Consortium for Political and Social Research [4]. We work with the family (FMLY) data file which is one of the four data files in the CEX data set. The FMLY file of the first quarter of 2000 contains records on 7860 households (also called consumer units) across the United States. A record corresponding to one of these households consists of a large number of variables or characteristics of a household, for example, different kinds of income, expenditures on various kinds of goods, region of residence, size of the household, number of vehicles owned and many many others. Each variable has a name, for example, the variable FOODCQ denotes expenditures on food during the current quarter, and a preassigned position in the FMLY file. The full list and description of these variables can be found in the codebook accompanying the CEX data set for 1999-2000 [4].

In order to have a somewhat homogeneous sample, we select from the FMLY file only those households which contain married couples, whose tenure status is renter household, homeowner with mortgages or without mortgages, whose age of the head is between 18 and 65, and which have no self-employed members. A selection similar to ours was also used by Nicol [76], Donald [28], Lewbel [63] and others. The total number of households which met these criteria were 1771 (out of 7860).

With each of the selected households, we retain the variables of interest to our study, namely, total income of a household, expenditures on various goods and some location variables. By the total income we mean wages and salaries before deductions received by all household members in the past 12 months, denoted by the variable SALARYX in the FMLY data file. (There are also other total income variables that one can use, for example, income after taxes.) We group expenditure variables into six categories of goods: food (FOODCQ variable in the FMLY data file), health care (HEALTHCARE variable), transportation

(TRANSCQ variable), household (HOUSECQ), apparel (APPARCQ variable to which we also added the personal care variable PERSCACQ) and miscellaneous goods (essentially the sum of variables corresponding to expenditures on entertainment, books, alcohol and others). This grouping is motivated by the following two considerations. First, the majority of the above groups of goods are already defined as categories of goods for the expenditures in the FMLY data file. Second, since one would like to match expenditures for various goods to their prices, we have to take into account the groups of goods considered in the ACCRA data. We will see that the categories of goods defined above are essentially those that appear in the ACCRA data. (This latter point is not so much relevant to the applications presented in this thesis because we end up using the average prices of the ACCRA data; see below. It is nevertheless important whenever a more general study of ranks is undertaken.)

The location variables associated with a household are the population size variable of a principal statistical unit (PSU, in short; see below for an explanation) denoted by POPSIZE in the FMLY file, the dummy variable SMSASTAT indicating whether a selected household belongs to a metropolitan statistical area (MSA, in short; see below) and the STATE variable indicating the state where the household belongs to. These variables will be used as matching variables below to associate prices from the ACCRA data to different households. The notions of MSA and PSU used earlier are easy to explain. MSA's are just metropolitan areas of the Unites States, for example, Atlanta (GA), Nashville (TN) and so on. They are defined by the Office of Management and Budget at the White House [3] and each of them is assigned a number. Unlike MSA's which are just used in CEX data sets, PSU's are directly linked to them. PSU's are groups of one, two or more MSA's whose households were selected for a CEX survey. They are created and used by CEX for confidentiality reasons.

The ACCRA data is the premier source of information on living cost differentials among U.S. urban areas. It is published quarterly since 1968, covers a varying number of urban areas (for example, 302 urban areas in the first quarter of 2001 and 317 urban areas in the first quarter of 2000) and contains price indices for six categories of goods of the corresponding urban areas: food (grocery items), housing, utilities, transportation, health care and others (miscellaneous goods and services). It also contains a composite index which is computed as a weighted average of the six categories of goods. For a detailed description of these categories of goods, corresponding indices and a composite index, see the ACCRA Cost of Living Index Manual [2]. Since we want to match price data to expenditures, we use the ACCRA price data for the first quarter of 2000 which we acquired directly from ACCRA [5]. To give an idea of the prices (price indices), we provide in Table 7.1 below a sample of prices for eight urban areas across the United States. Observe that prices and even composite prices are quite different for various urban areas. (For example, the price index 112.6 in Food category for Boston (MA) means that the food prices in Boston are 12.6 percent higher than the average of food prices for all areas participating in the ACCRA survey, while 88.4 in Housing category for Houston (TX) means that the housing prices in Houston are on average 11.6 percent lower than elsewhere.)

Area	Composite	Food	Housing	Utilities	Transp.	Health	Others
Boston (MA)	136.3	112.6	188.3	130.3	117.5	126.6	112.2
Cincinnati (OH)	98.3	95.8	93.5	110.8	97.3	94.1	101.4
Detroit (MI)	110.5	108.0	130.3	106.4	104.6	113.2	97.2
Houston (TX)	95.7	93.5	88.4	98.4	107.4	110.1	96.5
Jacksonville (NC)	97.5	100.6	89.7	102.9	88.6	97.7	104.1
Knoxville (TN)	94.5	95.5	87.8	96.6	91.5	91.7	100.4
Minneapolis (MN)	107.5	102.7	108.2	101.5	114.2	132.0	105.1
Pittsburgh (PA)	109.8	100.7	109.4	139.6	104.8	103.0	107.8

Table 7.1: ACCRA data of First Quarter 2000 for eight selected urban areas

Having obtained the ACCRA data, we then need to match it to households selected from the CEX data set. The matching procedure is carried out by using location variables POPSIZE, SMSASTAT and STATE in the FMLY data file of CEX (see above) as follows. On our request, the Bureau of Labor Statistics provided a confidential list of all (numbers of) PSU's and corresponding MSA's that were used in the CEX survey of the first quarter of 2000. Then, by using the list of metropolitan areas [3], we first identify each MSA number in the confidential list with the MSA area name and the state. By using population sizes of metropolitan areas published by the U.S. Census Bureau [1], we next also assign to each identified MSA its population size. Then, to each such MSA we can assign prices of goods from the ACCRA data. The majority of MSA's have their prices reported in the ACCRA data. In cases when MSA price data is not available, we use the average prices of areas in the neighborhood counties as proxies. Next, with each MSA used, we also associate the value of its POPSIZE variable. To do so, we first list all PSU's and their MSA's, and then compute their population sizes by summing up the sizes of the corresponding MSA's. The POPSIZE variable assigned to an MSA is a dummy variable of the population size of the PSU that the MSA belongs to. It equals 1 if the PSU size is more than 4 million, it equals 2 if the size is between 1.2 and 4 million, and so on. After performing these steps, we obtain entries with the following information: PSU number, MSA number, MSA name, vector of prices, STATE variable indicating the state where MSA belongs to and POPSIZE variable indicating the PSU size.

Finally, we want to match the prices in these created entries to households selected from the CEX data. To do so, we use the triplet of STATE, POPSIZE and SMSASTAT variables in the CEX data set as matching variables. If the variable SMSASTAT indicates that a household does not belong to an MSA, we drop it from our analysis. (There were 236 such households out of 1771 selected.) If a household belongs to an MSA, we match its STATE and POPSIZE variables to those already associated with the price data. For the majority of STATE and POPSIZE variable values, there is only one corresponding entry with price information. These prices are then taken as a price data associated with a household (see also a further discussion below). In cases when there are a few corresponding entries with the same STATE and POPSIZE values (this can happen, for example, when one PSU has more than one MSA or when there are a number of cities of comparable size in the same state), we take the average of these prices as a price data associated with a household.

We thus match the ACCRA price data to households selected from the CEX data file. In addition, we drop from our analysis those households who refused to report their total income, whose income was lower than \$2,000 or higher than \$150,000, and whose total expenditures were reported zero. After these and the eliminations discussed earlier, we are left with 755 households. We will work hereafter with expenditures, income and price data for these 755 households.

We still need to make clear exactly what price vectors we associate with households. According to Economic Theory, we have to take a vector of prices of the categories of goods used in a demand system. We can, in principle, do this for our data set by taking the prices of food, health care and transportation for the corresponding categories of goods in the FMLY file, the average of the housing and the utilities prices for the household category and, for example, by taking the prices of other goods as proxies for apparel and miscellaneous goods. We think, however, that the constructed data set would not be feasible for our purposes. Recall that estimation of local rank involves localization at a fixed price variable z. If z is, say, 6-dimensional (as it would be in our case) and if there are 755 observations, then there would be very few observations around a fixed value of z at our disposal. To understand what we mean, suppose for instance that there are $N = 1000$ observations (of prices) uniformly distributed in a 6-dimensional box $[0,1]^6$ and that the bandwidth h which is used in localization, is equal to 0.2. Then, the average number of observations in a 6-dimensional box of the size $h^6 = (0.2)^6$ is $N(h^6) = 1000(0.2)^6 = 0.064$. There is no way one can make a sound inference from 0.064 observations. If the box were 7-dimensional or 8-dimensional, then the average numbers would be even smaller, 0.0128 and 0.00256, respectively. In contrast, if the box were 1-dimensional or 2-dimensional, the corresponding average number of observations would be 200 and 40, respectively. One can see that the problem is a large dimension of a vector z. Then, either one needs more observations or one has to use a z of a lower dimension as a proxy. (Similar observations, regarding what is known as the empty-space phenomenon, are also made in the context of kernel based estimation of multi-dimensional densities. See, for example, Silverman [97], pp. 92-93.)

We shall follow here the second approach (but see also remarks below). More precisely, we will take z to be the ACCRA composite price index which is one-dimensional. If one agrees to take a z of a lower dimension, then this assumption has a few advantages. First, it is the simplest one in the sense that one does not have to worry what dimension to take, which prices to eliminate or average and so on. Second, by its definition, a composite price index is supposed to capture best the overall price state of an urban area, and hence is quite natural to use. Third, it is also most convenient to work with since there are now more observations around a fixed value of a price z (as compared to two or higher dimensions). And, fourth, one may also expect that this simplest case becomes a guide to more general situations of multi-dimensional z's.

Remark 7.1.1 One way to increase the number of observations N (and hence to curtail the problem of dimensionality) is to use the CEX and ACCRA data for multiple quarters. One then needs, however, to make adjustments in the data to take into account inflation. This can be done, for example, by using CPI (Consumer Price Indexes) data. See Nicol [76] for more details.

Remark 7.1.2 One may ask the question whether by considering z's of a lower dimension, we indeed avoid problems of estimation of the local rank discussed earlier. It may seem that averaging of the coordinates of z to obtain a one-dimensional proxy (as we did for the ACCRA data) is equivalent to taking a large h when localizing at a multi-dimensional z. The advantage of the former approach, we feel, is that the one-dimensional z has a clear meaning and that, by being able to choose a small h then, we indeed achieve a localization. On the other hand, the latter approach based on keeping a multi-dimensional z and taking a large h, we feel, is likely to become a black box for moderate sample sizes with hard to interpret results.

We conclude this section with a few illustrative plots of shares of some goods against the logarithm of total income in the constructed data set. (In applications, one typically uses the logarithm of total income instead of the total income itself.) The share of a group of goods, say a share of food, is defined as the ratio of expenditures for that group of goods and the total expenditures. Figure 7·1 above shows the share of health care against the logarithm of total income while Figure 7·2 shows that of miscellaneous goods against the logarithm of total income. Observe that the two plots have quite different shapes. (If one ignores the price variable z, then the rank of the constructed data set of a demand system is the smallest number of functions whose linear combinations fit well the data in the plots of Figures 7·1, 7·2 and also the other figures that one would get by considering the rest of the shares. Since the two plots in Figures 7·1 and 7·2 have quite different shapes, one is inclined to deduce now that the rank is probably at least 2.)

7.2 Application of semi-parametric factor model

In this section, we model the economic data described above through the semi-parametric factor demand system $Y_i = \theta(Z_i)V(X_i) + \epsilon_i$, $i = 1, \dots, N$. Here, $N = 755$ is the number of households in our data set, Y_i is a vector of shares of six categories of goods, Z_i is the corresponding composite price index and X_i is the total income of a household. Recall that the categories of goods in Y_i are food, household, health care, transportation, apparel and miscellaneous goods. For notational convenience, we consider hereafter price indices Z_i divided by 100. So, for example, the fixed price regime $z = 1.2$ corresponds to prices 20 percent higher than the average and $z = 0.9$ corresponds to prices 10 percent lower that the average (see Section 7.1). Finally, the vector $V(x)$ above has the size 4×1 and is given by

$$V(x)' = (1 \quad c^{-1}\log x \quad (c^{-1}\log x)^2 \quad (c^{-1}\log x)^3)$$

where $c = \log(150000)$ is the normalization constant, and the 5×4 matrix $\theta(z)$ is unknown. Normalization constants are sometimes used in modeling of demand systems and their rank

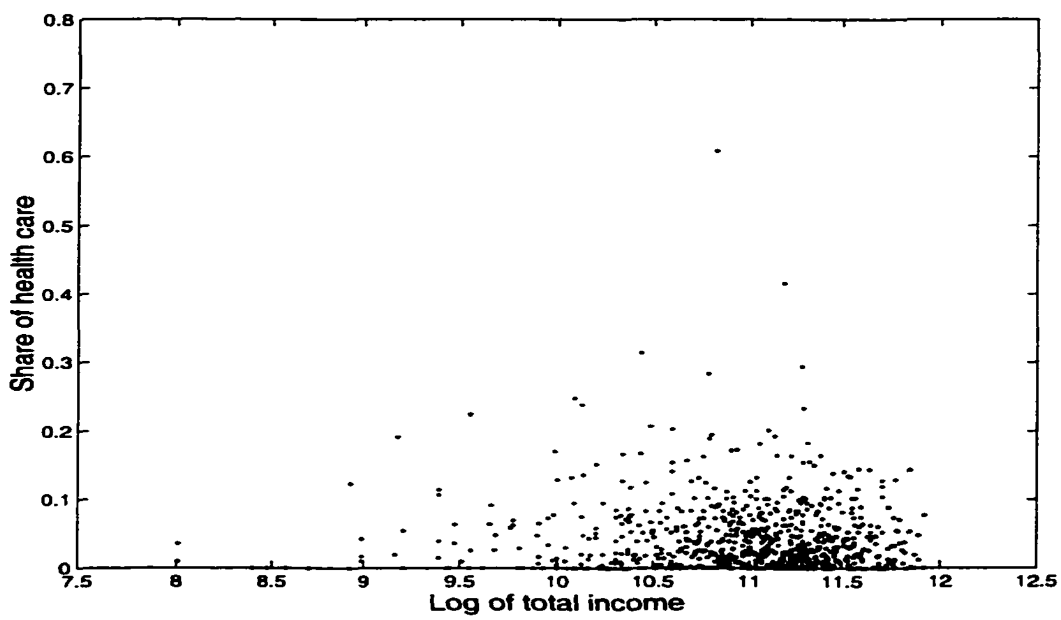

Figure 7·1: Share of health care versus logarithm of total income

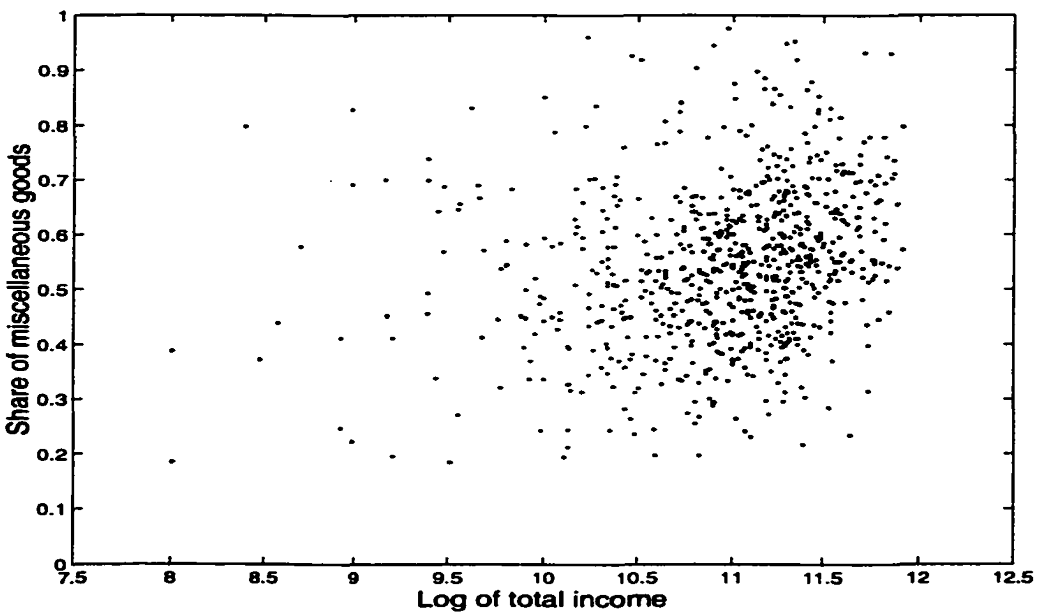

Figure 7·2: Share of miscellaneous goods versus logarithm of total income

estimation (see, for example, Cragg and Donald [18]) to make variables in the model of a comparable size. Recall that 150000 in the constant c is the maximum of total income considered. Hence, $c^{-1}\log(X_i) \in (0,1)$ and obviously $Y_i \in (0,1)$ as well. Observe also that this semi-parametric factor demand system nests the popular PIGLOG and QAID demand systems (see (2.18) and (2.19)). Our goal is to determine its rank.

As explained in Section 3.3.1, the rank can be estimated as the maximum of the ranks of all possible reduced demand systems, namely, demand systems where one share of goods is dropped from the analysis. Moreover, the rank of a reduced system is the rank of the corresponding reduced matrix $\Theta(z)$ and it can be estimated by using rank tests described in Section 5.1. Since these tests involve the kernel based estimator of $\Theta(z)$, we first need to choose a bandwidth h to estimate $\Theta(z)$ and also decide on what fixed values of price regimes z to consider. We will work with the Epanechnikov kernel (see Section 4.1) throughout.

We will look at a few price regime values, namely, $z = 1$, $z = 1.2$ and $z = 0.9$. As for the bandwidth h, the statistical literature offers a number of ways to select it in various contexts. One can use, for example, a cross-validation (see, e.g. Härdle [49] for basic ideas on cross-validation in the context of a non-parametric regression), the kth nearest neighbor method (see, e.g. Pagan and Ullah [77] and Härdle [49]) or employ some rule-of-thumb formula like $h = \widehat{\sigma}N^{-1/5}$ where N is the sample size and $\widehat{\sigma} > 0$ is the sample standard deviation of Z_i's. Cross-validation in our context chooses h which minimizes the sum

$$\frac{1}{N}\sum_{i=1}^{N}|Y_i - \widehat{\Theta}_i(Z_i)V(X_i)|^2,$$

where $\widehat{\Theta}_i(z)$ is defined in the same way as $\widehat{\Theta}(z)$ but by omitting the ith terms in the sums of its definition (the so-called "leave-one-out" estimator). For example, for a demand system with the food share eliminated, the cross-validation selected $h = 0.3$. The same value of h was also selected for other reduced demand systems except for the demand system without the miscellaneous goods in which case h was selected as 0.5. The cross-validation procedure is global in the sense that h is selected as optimal for a range of z's. The

kth nearest neighbor method, on the other hand, chooses h for each value of z separately (locally), namely, as two times the distance of z to its kth nearest neighbor (observation). The idea is that, when there are fewer observations around the variable z, then the selected h is larger. In practice, one typically takes $k \approx \sqrt{N}$ (see Pagan and Ullah [77]).

Which of these or other available methods are then appropriate to our situation? Since we are interested now in local tests at a few fixed values of z, then probably a cross-validation method is not that suitable because it is global in nature. For this reason, we will focus here on local selection methods. More precisely, we will look at a wide range of the values of h for the considered fixed price regimes z. We will include h in the range based on the number of observations Z_i in the neighborhood of the size h at a fixed price regime z. For example, when $z = 1$, there are 40, 88, 199, 282, 362, 541 and 678 observations Z_i in the $h = 0.01$, 0.015, 0.02, 0.04, 0.05, 0.08 and 0.2 neighborhoods of $z = 1$, respectively. We will consider $h = 0.015$, 0.02, 0.05 and 0.08 for $z = 1$. For $z = 1.2$, we will take $h = 0.05$, 0.07, 0.15 and 0.2 (with 65, 135, 282 and 513 observations, respectively) and, for $z \approx 0.9$, we will consider $h = 0.04$, 0.08, 0.11 and 0.17 (with 90, 175, 273 and 538 observations). Observe that the four values of h taken for each of the three fixed price regimes have a comparable number of corresponding observations. Note also that if, for example, h is chosen by the kth nearest neighbor method with $k \approx \sqrt{N} = \sqrt{755} \approx 27.47$, then this h is comparable with the smallest value of h chosen above since it is twice the distance to its 27th nearest neighbor (or approximately the distance to its $27(2) = 54$th neighbor). Other values of k can be considered in the same way. Finally, for the sake of completeness and comparison, in addition to the values of h chosen above, we will also report the results for h's selected by cross-validations.

Remark 7.2.1 Although we decided not to focus on global selection methods for the bandwidth h, there is one situation in local tests where such selection might be appropriate or just necessary. Observe that the variance-covariance matrix estimate $\widehat{\Sigma}$ in (4.22) used in either of the local tests described in Section 5.1, is defined by using the available range of Z_i's. Thus, one no longer focuses on a fixed value of z and, for this reason, we will compute $\widehat{\Sigma}$ by taking h selected by cross-validation.

Rank of reduced demand				
Elim. share	Method	1	2	3
Food	min-χ^2	0.0000	0.7198	0.8435
	LDU	0.0000	0.9999	0.9999
	ALS	0.0000	0.0000	0.8423
House	min-χ^2	0.0000	0.0508	0.5950
	LDU	0.0000	0.4999	0.9999
	ALS	0.0000	0.0000	0.5830
Health	min-χ^2	0.0000	0.0226	0.8800
	LDU	0.0000	0.9999	0.9999
	ALS	0.0000	0.0006	0.8073
Transport	min-χ^2	0.0000	0.0480	0.3557
	LDU	0.0000	0.7159	0.9997
	ALS	0.0000	0.0000	0.1898
Apparel	min-χ^2	0.0000	0.0065	0.3630
	LDU	0.0000	0.0000	0.9999
	ALS	0.0000	0.0000	0.3017
Miscell.	min-χ^2	0.0003	0.5508	0.3549
	LDU	0.0000	0.9998	0.9999
	ALS	0.0000	0.3803	0.1997

Table 7.2: P-values for hypothesis tests in a semi-parametric factor model ignoring z

Before we present the rank estimation results for the values of h and z chosen above, let us first examine what happens when the variable z is ignored altogether. As we will see below, results obtained ignoring z turn out to be interesting to compare with those when z is taken into account and also informative on what might be expected then. We hence suppose for a moment that the model is $Y_i = \theta V(X_i) + \epsilon_i$, estimate Θ in each of the reduced demand systems by using regression and deduce the rank of Θ by using some rank estimation method described in Section 5.1. The rank of θ is then taken as the maximum of the ranks of the corresponding reduced systems. The results are presented in Table 7.2 where the first column indicates the share dropped from the analysis, the second one shows which rank estimation method is used (the minimum-χ^2 test of Section 5.1.2, the asymptotic least squares of Section 5.1.4 or that based on the LDU decomposition of Section 5.1.1) and the rest of the columns contain the p-values for the hypothesis tests $\text{rk}\{\Theta\} \leq L$ with $L = 1, 2$ and 3, respectively. The rank 4 or higher is not reported since 4 is the highest possible rank for the matrix Θ. The p-values are computed by using the asymptotic results for test statistics described in Section 5.1. If a p-value is small, say less than 0.05 corresponding to a 5 percent significance level, then the corresponding null hypothesis test is rejected at that significance level. So, for example, according to Table 7.2, all three test statistics reject the rank 1 hypothesis at negligible significance levels.

Observe from Table 7.2 that the rank of the demand system with the apparel share eliminated is estimated as 3 by using any of the three estimation methods and by considering a significance level as low as 1 percent. Since none of the other reduced demand systems have their estimated rank higher than 3, we can conclude that the estimated rank of the *full* matrix θ and hence the rank of the demand system $y = \theta V(x)$ is 3. Observe also that rank estimation results for *reduced* demand systems are quite different, depending on the estimation method used. For example, at a significance level of 5 percent, the rank of the demand system without the food share is estimated as 2 by the LDU and the minimum-χ^2 tests, and as 3 by using the ALS test. The reader who is unfamiliar with rank estimation in practice, may be somewhat perplexed at the disparity of these results. They are, however, typical. It is known that even though the minimum-χ^2, the ALS and the LDU tests follow the same chi-square distribution in the limit, they have quite different properties for small or moderate sample sizes. We will come back to some of these points later.

Rank estimation results for the values of h and z chosen earlier are presented in Tables 7.4, 7.5 and 7.6 dealing with $z = 1$, $z = 1.2$ and $z = 0.9$, respectively. The first column of these tables indicates the share dropped from the analysis, the second column lists the values of h considered (the values of h between two horizontal lines correspond to those chosen by cross-validation) and the rest of the columns consist of the p-values for the hypothesis tests $\text{rk}\{\Theta(z)\} \leq L$ with $L = 1, 2$ and 3, based on the minimum-χ^2 (denoted simply by χ^2), the LDU and the ALS methods. Based on the results reported in these tables, we make the following observations. They are further discussed below.

1. *Dependence of rank tests on choices of h.* Observe that the p-values increase in most cases when the minimum-χ^2 test is applied with smaller values of h. One is thus inclined not to reject a lower rank when applying the minimum-χ^2 test with smaller h. This behavior seems to be characteristic to the LDU test as well (see, in particular, the LDU test results for rank 2 in Table 7.4 with the transportation or the apparel share eliminated) but it is less pronounced and more difficult to judge as most of the p-values are either close to 0 or to 1. The same conclusion cannot be drawn for the ALS test where the p-values do not seem to obey any particular rule with respect to changes in h. Observe, however, that in a number of cases, the p-values even decrease as h becomes smaller (see, for example, the results for rank 3 in Table 7.4 with the house share eliminated or in Table 7.6 with the miscellaneous goods share eliminated).

2. *Comparison to tests where z is ignored.* Observe that the p-values for the minimum-χ^2 test in Tables 7.4–7.6 are higher or comparable to the corresponding p-values in Table 7.2 where z is ignored altogether. Thus, when using the minimum-χ^2 test, one tends not to reject a local rank which is lower than that obtained when z is ignored. Moreover, observe that, as h becomes larger and in particular when h is chosen by cross-validation, the p-values for the minimum-χ^2 test come close to the corresponding ones in Table 7.2. For small h, however, the p-values are considerably larger and hence lead to the conclusions different from those resulting from Table 7.2. The same observation seems to hold for the LDU test in Table 7.4 (see, in particular, the results for rank 2 when the house or the apparel share is eliminated) and less so in Tables 7.5 and 7.6. As for the ALS test, it is difficult to draw any tangible conclusions. It seems, however, that for many reduced demand systems, in fact the opposite is true, namely, the p-values are comparable or smaller than the

corresponding ones in Table 7.2. One is inclined then not to reject a local rank which is the same or higher than that obtained ignoring z altogether.

3. *Comparison of different methods.* Observe that, when testing for rank 1, the p-values for the minimum-χ^2 test are higher than the corresponding ones for the ALS and the LDU tests. When testing for ranks 2 and 3, however, the situation changes. The p-values are now the smallest for the ALS test and the highest for the LDU test, while those for the minimum-χ^2 test are intermediate.

4. *Estimates of local ranks in the full and reduced demand systems.* In Table 7.3 below, we summarize the local rank estimates in the reduced demand systems which are obtained from the results of Tables 7.4–7.6 at a 5 percent significance level. In many cases, we put more than one entry because the estimates are different for different values of h. Moreover, the estimates in the parenthesis correspond to the values of h chosen by cross-validation. By taking the maximum of the ranks obtained in Table 7.3, we can conclude that the local ranks at $z = 1$, 1.2 and 0.9 are 2, independently of the values of h considered (but excluding the value of h chosen by cross-validation) and when either the minimum-χ^2 or the LDU test is used. Thus, in this case, the estimated local rank is smaller than rank 3 obtained from the results in Table 7.2 where z is ignored. If one uses h selected by cross-validation, then the estimated local rank is typically higher, namely, it is 3 at all three z's by using the minimum-χ^2 test, and it is 2 at $z = 1.2$, and 3 at $z = 1$ and 0.9 by using the LDU test. If one uses the ALS test, then the estimated local ranks of a full system is 4 in most cases. It is thus higher than the rank obtained when z is ignored.

   Finally, let us observe that, although the estimated local ranks of a full demand system are the same for three different price values of z and when the same method is used, the estimated local ranks for *reduced* systems are different in some situations. See, for example, the estimation of rank by using the minimum-χ^2 test in the system with the miscellaneous goods share eliminated for small values of h. In this case, the local rank is 1 at $z = 1$ and 1.2, and it is 2 at $z = 0.9$.

Estimated local rank at $\alpha = 5\%$ significance level				
Elim. share	Test	$z = 1$	$z = 1.2$	$z = 0.9$
Food	min-χ^2	1 or 2 (2)	1 or 2 (2)	1 or 2 (2)
	LDU	2 (2)	2 (2)	2 (2)
	ALS	2, 3 or 4 (3)	3 or 4 (3)	2, 3 or 4 (3)
House	min-χ^2	1 or 2 (3)	1 or 2 (3)	2 (2)
	LDU	2 (2)	2 (2)	2 or 3 (3)
	ALS	2 or 3 (4)	4 (4)	2, 3 or 4 (4)
Health	min-χ^2	1 or 2 (3)	2 or 3 (3)	2 (2)
	LDU	2 (2)	2 (2)	2 or 3 (2)
	ALS	2 or 3 (3)	4 (3)	4 (3)
Transp.	min-χ^2	1 or 2 (3)	2 (2)	2 (2)
	LDU	2 (2)	2 (2)	2 (2)
	ALS	3 or 4 (3)	4 (3)	2,3 or 4 (4)
Apparel	min-χ^2	1 or 2 (3)	1 or 2 (3)	2 (3)
	LDU	2 (3)	2 (2)	2 or 3 (2)
	ALS	3 or 4 (4)	3 (4)	3 or 4 (4)
Miscell.	min-χ^2	1 (2)	1 (1)	1 or 2 (2)
	LDU	2 (2)	1 or 2 (2)	2 (2)
	ALS	1 (2)	1 (2)	2, 3 or 4 (1)

Table 7.3: Estimated local ranks in a semi-parametric factor model at $z = 1$, 1.2 and 0.9

These observations summarize what we expect to find in other applications as well. Some of these observations are also supported by the results of Monte-Carlo simulations presented in Section 7.4. Our main finding is that, when using smaller values of h and either the minimum-χ^2 or the LDU test, the estimated local rank is smaller than the rank obtained when z is ignored. If larger h is considered, for example h selected by cross-validation, then the estimated local rank is essentially the same as that obtained ignoring z. Since we argued earlier in the section that the cross-validation and hence larger h may not be appropriate to our context, we suggest working with smaller h and therefore not rejecting lower local ranks. Our findings should not be surprising. To understand why

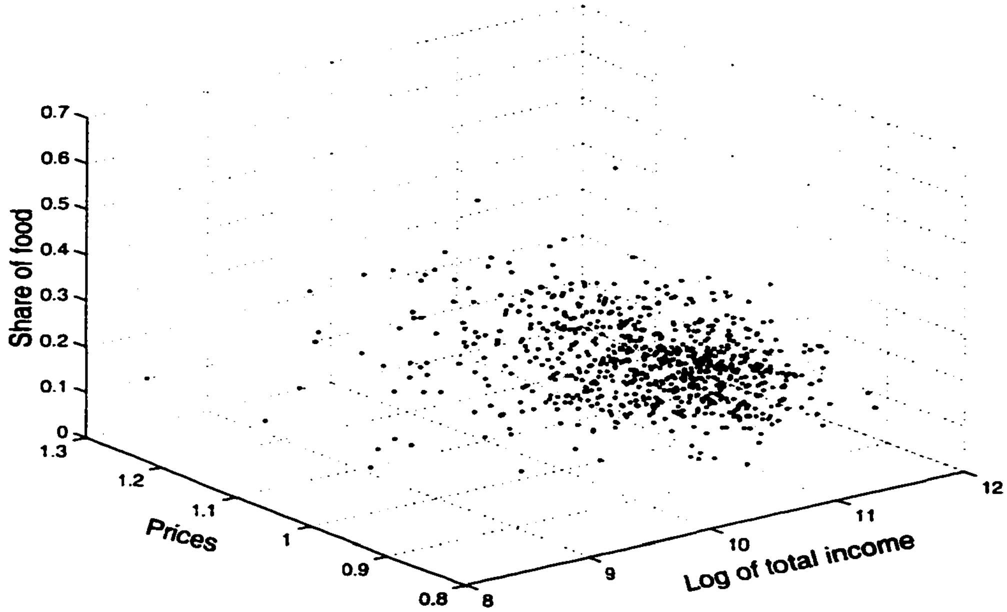

Figure 7·3: Share of food versus prices and logarithm of total income

not, imagine 3-dimensional plots of shares, prices and total income on the three axis as, for example, in Figure 7·3. If z is fixed, then the local rank at z (small h) is essentially the smallest number of functions needed to fit the shape of these graphs around a cross section at that fixed value of z. If z is now ignored, that is all of the cross sections are combined into a single two-dimensional plot, then there is clearly a greater variety of shapes in the newly obtained plot and hence the rank is higher. The results based on the ALS test are somewhat counterintuitive in the above sense because they point to local ranks higher than the rank obtained ignoring z. This may be explained by the fact that the ALS test is known to overestimate the rank (see, for example, Robin and Smith [89] and also Section 7.4 below).

Rank of reduced demand at $z = 1$										
Elimin. share	h	1			2			3		
		χ^2	LDU	ALS	χ^2	LDU	ALS	χ^2	LDU	ALS
Food	0.015	0.502	0.000	0.000	0.948	0.999	0.248	0.916	0.999	0.337
	0.02	0.155	0.000	0.000	0.763	0.999	0.000	0.909	0.999	0.011
	0.05	0.004	0.000	0.000	0.318	0.991	0.000	0.837	0.999	0.648
	0.08	0.001	0.000	0.000	0.498	0.998	0.000	0.912	0.999	0.865
	0.3	0.000	0.000	0.000	0.538	0.999	0.000	0.627	0.999	0.402
House	0.015	0.149	0.000	0.000	0.957	0.997	0.142	0.880	0.999	0.012
	0.02	0.062	0.000	0.000	0.709	0.999	0.000	0.910	0.999	0.150
	0.05	0.000	0.000	0.000	0.258	0.905	0.000	0.609	0.999	0.793
	0.08	0.000	0.000	0.000	0.340	0.542	0.000	0.521	0.999	0.624
	0.3	0.000	0.000	0.000	0.023	0.991	0.000	0.417	0.999	0.000
Health	0.015	0.124	0.000	0.000	0.813	0.939	0.000	0.889	0.999	0.292
	0.02	0.051	0.000	0.000	0.955	0.999	0.253	0.945	0.999	0.094
	0.05	0.000	0.000	0.000	0.810	0.999	0.425	0.773	0.999	0.513
	0.08	0.000	0.000	0.000	0.637	0.999	0.365	0.761	0.999	0.507
	0.3	0.000	0.000	0.000	0.017	0.999	0.000	0.845	0.999	0.673
Transport	0.015	0.060	0.000	0.000	0.705	0.893	0.000	0.931	0.999	0.823
	0.02	0.020	0.000	0.000	0.699	0.998	0.000	0.936	0.999	0.455
	0.05	0.000	0.000	0.000	0.222	0.576	0.000	0.496	0.999	0.006
	0.08	0.000	0.000	0.000	0.232	0.462	0.000	0.265	0.999	0.453
	0.3	0.000	0.000	0.000	0.029	0.597	0.000	0.266	0.999	0.741
Apparel	0.015	0.073	0.000	0.000	0.711	0.878	0.000	0.928	0.999	0.000
	0.02	0.029	0.000	0.000	0.805	0.973	0.000	0.980	0.999	0.584
	0.05	0.000	0.000	0.000	0.246	0.908	0.000	0.465	0.999	0.000
	0.08	0.000	0.000	0.000	0.196	0.628	0.000	0.177	0.999	0.000
	0.3	0.000	0.000	0.000	0.003	0.002	0.000	0.244	0.999	0.000
Miscell.	0.015	0.542	0.006	0.350	0.853	0.999	0.313	0.930	0.999	0.863
	0.02	0.784	0.006	0.751	0.988	0.999	0.985	0.914	0.999	0.894
	0.05	0.222	0.000	0.226	0.728	0.999	0.652	0.797	0.999	0.787
	0.08	0.089	0.000	0.075	0.480	0.999	0.198	0.571	0.999	0.526
	0.5	0.000	0.000	0.000	0.418	0.999	0.110	0.273	0.999	0.147

Table 7.4: P-values for hypothesis tests in a semi-parametric factor model at $z = 1$

Rank of reduced demand at $z = 1.2$										
Elimin. share	h	1 χ^2	1 LDU	1 ALS	2 χ^2	2 LDU	2 ALS	3 χ^2	3 LDU	3 ALS
Food	0.05	0.308	0.000	0.000	0.683	0.999	0.023	0.579	0.999	0.265
	0.09	0.086	0.000	0.000	0.519	0.999	0.000	0.326	0.999	0.409
	0.15	0.001	0.000	0.000	0.597	0.999	0.000	0.372	0.999	0.011
	0.2	0.001	0.000	0.000	0.827	0.988	0.000	0.654	0.999	0.377
	0.3	0.000	0.000	0.000	0.713	0.999	0.000	0.719	0.999	0.528
House	0.05	0.085	0.000	0.000	0.372	0.999	0.000	0.347	0.999	0.000
	0.09	0.030	0.000	0.000	0.258	0.999	0.000	0.251	0.999	0.000
	0.15	0.000	0.000	0.000	0.142	0.715	0.000	0.343	0.999	0.000
	0.2	0.000	0.000	0.000	0.102	0.994	0.000	0.505	0.999	0.000
	0.3	0.000	0.000	0.000	0.043	0.992	0.000	0.487	0.999	0.000
Health	0.05	0.000	0.000	0.000	0.043	0.999	0.000	0.487	0.999	0.000
	0.09	0.030	0.000	0.000	0.258	0.999	0.000	0.251	0.999	0.000
	0.15	0.000	0.000	0.000	0.142	0.319	0.000	0.343	0.999	0.000
	0.2	0.000	0.000	0.000	0.102	0.999	0.000	0.505	0.999	0.000
	0.3	0.000	0.000	0.000	0.043	0.999	0.002	0.487	0.999	0.430
Transport	0.05	0.036	0.000	0.000	0.216	0.999	0.000	0.287	0.999	0.000
	0.09	0.008	0.000	0.000	0.125	0.999	0.000	0.273	0.999	0.000
	0.15	0.000	0.000	0.000	0.157	0.998	0.000	0.383	0.999	0.000
	0.2	0.000	0.000	0.000	0.142	0.990	0.000	0.507	0.999	0.000
	0.3	0.000	0.000	0.000	0.064	0.348	0.000	0.431	0.999	0.289
Apparel	0.05	0.101	0.000	0.000	0.588	0.999	0.000	0.878	0.999	0.602
	0.09	0.019	0.000	0.000	0.307	0.999	0.000	0.889	0.999	0.631
	0.15	0.000	0.000	0.000	0.208	0.999	0.000	0.982	0.999	0.970
	0.2	0.000	0.000	0.000	0.054	0.999	0.000	0.948	0.999	0.864
	0.3	0.000	0.000	0.000	0.013	0.953	0.000	0.569	0.999	0.000
Miscell.	0.05	0.540	0.219	0.567	0.743	0.999	0.754	0.615	0.999	0.623
	0.09	0.479	0.316	0.500	0.710	0.999	0.729	0.616	0.999	0.601
	0.15	0.394	0.000	0.377	0.714	0.999	0.584	0.776	0.999	0.747
	0.2	0.102	0.000	0.377	0.758	0.999	0.584	0.795	0.999	0.747
	0.5	0.102	0.000	0.000	0.758	0.999	0.066	0.795	0.999	0.181

Table 7.5: P-values for hypothesis tests in a semi-parametric factor model at $z = 1.2$

Rank of reduced demand at $z = 0.9$										
Elimin.		1			2			3		
share	h	χ^2	LDU	ALS	χ^2	LDU	ALS	χ^2	LDU	ALS
Food	0.04	0.268	0.000	0.000	0.514	0.999	0.062	0.608	0.999	0.350
	0.08	0.322	0.000	0.000	0.439	0.999	0.014	0.309	0.999	0.027
	0.11	0.349	0.000	0.000	0.701	0.999	0.351	0.703	0.999	0.387
	0.17	0.037	0.000	0.000	0.741	0.958	0.000	0.733	0.999	0.622
	0.3	0.000	0.000	0.000	0.510	0.999	0.000	0.607	0.999	0.390
House	0.04	0.000	0.000	0.000	0.144	0.826	0.000	0.204	0.999	0.126
	0.08	0.000	0.000	0.000	0.205	0.000	0.111	0.175	0.999	0.000
	0.11	0.000	0.000	0.000	0.752	0.986	0.000	0.514	0.999	0.002
	0.17	0.000	0.000	0.000	0.659	0.619	0.000	0.648	0.999	0.039
	0.3	0.000	0.000	0.000	0.071	0.000	0.000	0.341	0.999	0.000
Health	0.04	0.000	0.000	0.000	0.133	0.962	0.000	0.168	0.999	0.000
	0.08	0.000	0.000	0.000	0.281	0.999	0.000	0.262	0.999	0.000
	0.11	0.000	0.000	0.000	0.526	0.999	0.000	0.446	0.999	0.000
	0.17	0.000	0.000	0.000	0.510	0.002	0.001	0.623	0.999	0.000
	0.3	0.000	0.000	0.000	0.075	0.999	0.000	0.545	0.999	0.059
Transport	0.04	0.000	0.000	0.000	0.296	0.999	0.003	0.642	0.999	0.289
	0.08	0.000	0.000	0.000	0.480	0.999	0.119	0.705	0.999	0.567
	0.11	0.000	0.000	0.000	0.666	0.999	0.004	0.717	0.999	0.352
	0.17	0.000	0.000	0.000	0.417	0.983	0.014	0.305	0.999	0.001
	0.3	0.000	0.000	0.000	0.082	0.999	0.000	0.202	0.999	0.000
Apparel	0.04	0.000	0.000	0.000	0.644	0.999	0.007	0.949	0.999	0.925
	0.08	0.000	0.000	0.000	0.696	0.999	0.000	0.992	0.999	0.991
	0.11	0.000	0.000	0.000	0.692	0.999	0.000	0.461	0.999	0.000
	0.17	0.000	0.000	0.000	0.353	0.000	0.000	0.262	0.999	0.000
	0.3	0.000	0.000	0.000	0.024	0.995	0.000	0.238	0.999	0.000
Miscell.	0.04	0.003	0.000	0.000	0.128	0.999	0.000	0.129	0.999	0.000
	0.08	0.016	0.000	0.000	0.327	0.999	0.013	0.247	0.999	0.031
	0.11	0.141	0.000	0.000	0.686	0.999	0.000	0.630	0.999	0.438
	0.17	0.053	0.000	0.001	0.565	0.999	0.154	0.498	0.999	0.375
	0.5	0.000	0.000	0.950	0.445	0.999	0.929	0.291	0.999	0.836

Table 7.6: P-values for hypothesis tests in a semi-parametric factor model at $z = 0.9$

7.3 Application of non-parametric model

In this section, we turn to applications of a non-parametric model to determine a local rank in a demand system. The data set is that constructed in Section 7.1 and already used in Section 7.2. Following the method described in Section 3.3.2, we will deduce the local rank by dropping one share of goods from the analysis, estimating the adjusted local rank in the reduced demand system by using tests developed in Section 5.2 and then adding to it 1.

But before we proceed with local tests, let us examine again what happens when the variable z is ignored altogether. In other words, we suppose that the model is now $Y_i = f(X_i) + \epsilon_i$, drop one share of goods from the analysis, use Donald's [28] method to estimate the adjusted rank of the reduced demand system (see also Section 2.4 and, in particular, the corresponding test statistic (2.37)) and then add to it 1. By following Donald [28], we take X_i above to be the logarithm of total income and not the total income itself as in Section 7.2. The share that we drop is health care. (The results were invariant to which share of goods is eliminated. Only when the share of miscellaneous goods was dropped, the conclusion that we will make, was not as pronounced.) In order to use Donald's estimation method, we first need to select a bandwidth h. We do so by using a cross-validation which yields $h = 0.7$. Results of applications of Donald's test are presented in Table 7.7. Column 1 lists the value of h selected by cross-validation, as well as other values of h that we consider. Columns 2 through 6 consist of the p-values for the hypothesis tests $\mathrm{rk}\{f\} \leq L$ (or, equivalently, $\mathrm{adrk}\{F\} + 1 \leq L$) with $L = 1, \cdots, 5$, based on the test statistic (2.37). According to the results in Table 7.7, the rank of the full demand system is estimated as 3 for all considered values of h. This finding should not be surprising because rank 3 has been found in the CEX data sets obtained from many other survey years (see, for example, Lewbel [63] and Donald [28]).

We now turn back to estimation of a local rank. In order to carry out the estimation,

Rank of demand					
h	1	2	3	4	5
0.1	0.0000	0.0002	0.8071	0.9526	0.9563
0.3	0.0000	0.0000	0.9399	0.9724	0.9271
0.5	0.0000	0.0000	0.8917	0.9697	0.9199
0.7	0.0000	0.0000	0.8219	0.9378	0.8749

Table 7.7: P-values for hypothesis tests in a non-parametric model ignoring z

we need to choose a share of goods to drop, select a bandwidth for local test statistics and also decide on what values of price regime z to consider. Since we want to compare our results to those in Table 7.7, we will drop the share of health care. We will consider the same price regimes as already used in Section 7.2, namely, $z = 1$, $z = 1.2$ and $z = 0.9$. For a bandwidth h, as argued in Remark 5.2.5, we will choose in fact two bandwidths h_x and h_z corresponding to the variables X_i and Z_i, respectively. For h_z, since we are interested in a fixed z, we will take the same values as in Section 7.2, namely, $h_z = 0.015$, 0.02, 0.05 and 0.08 when $z = 1$, $h_z = 0.05$, 0.09, 0.15 and 0.2 when $z = 1.2$, and $h_z = 0.04$, 0.08, 0.11 and 0.17 when $z = 0.9$. For h_x, we will take three values from those appearing in Table 7.7, namely, $h_x = 0.1$, 0.3 and 0.5 (the results with $h_x = 0.7$ were similar to those with $h_x = 0.5$). In addition, we will also report the results for $h_z = 0.3$ which, together with $h_x = 0.7$, was selected by cross-validation. Finally, as in Section 7.2, we will compute the variance-covariance estimate $\widehat{\Sigma}$ in (4.9) with $h_x = 0.7$ and $h_z = 0.3$ chosen by cross-validation.

Results on local rank estimation for the selected values of h_x, h_z and z are presented in Tables 7.8, 7.9 and 7.10 dealing with $z = 1$, $z = 1.2$ and $z = 0.9$, respectively. The first two columns in these tables indicate the chosen bandwidths h_x and h_z. The other five columns consist of the p-values for the hypothesis tests $\mathrm{rk}\{f(\cdot, z)\} \leq L$ (or, equivalently, $\mathrm{adrk}\{F(\cdot, z)\} + 1 \leq L$) with $L = 1, \cdots, 5$. The p-values are computed by using the test statistic $\widehat{T}_1(L, z)$ in (5.49) and its asymptotic behavior established in Theorem 5.2.2. (Results based on the test statistic $\widehat{T}_2(L, z)$ in (5.51) were of the same nature and hence are not reported.) Based on the results of Tables 7.8, 7.9 and 7.10, we make the following observations.

1. *Dependence of rank tests on choices of h_x and h_z.* Observe that in almost all cases (except for $z = 0.9$ and $h_x = 0.1$), for a fixed h_x, the p-values increase as h_z becomes smaller. Observe also that in most cases (except when smallest values of h_z are considered), for a fixed h_z, the p-values also increase as h_x becomes smaller. Hence, the p-values are likely to increase as h_x and h_z become smaller simultaneously. This shows that one would not reject a lower rank when smaller h_x and h_z are considered.

2. *Comparison to tests where z is ignored.* Observe that the p-values in Tables 7.8, 7.9 and 7.10 are higher than the corresponding ones in Table 7.7. Thus, one tends not to reject a local rank which is lower than the rank obtained ignoring z. Observe also that, for a fixed h_x, as h_z increases, the p-values do not reach the corresponding p-values reported in Table 7.7. This behavior is in contrast to that observed in Section 7.2 for the semi-parametric factor model.

3. *Estimates of local rank.* Observe that the estimated local ranks are clearly less than 3 for all three values of z. In particular, they are lower than the estimated rank 3 obtained from Table 7.7 where z is ignored. The estimated local rank comes closest to 2 at $z = 0.9$ and maybe at $z = 1.2$. For smaller values of h_x and h_z, however, the local rank is estimated as 1 throughout.

According to these observations, the local rank estimates are lower than the rank estimate obtained ignoring z. This phenomenon is in the spirit of that observed in connection to the semi-parametric factor model considered in Section 7.2. What is perhaps surprising is that the local rank estimates turned out to be so unequivocally smaller than rank 3 obtained ignoring z. This raises an important question whether local ranks are smaller than 3 not only for the three values of z considered here, but for all z's. The answer to this question is only to come in the future after statistical tests for global ranks are developed.

Remark 7.3.1 As discussed in Section 5.2.3, there are potentially other ways to estimate a local rank in a non-parametric relation, namely, the LDU and the ALS methods for symmetric matrices (the minimum-χ^2 method corresponding to the test statistic $\widehat{T}_2(L, z)$ in (5.51) yields results analogous to those in Tables 7.7–7.10). These alternatives are interesting and important because, as noted at the end of Section 5.1.4, they provide a better grip on an estimation object (in this case, a local rank of a non-parametric relation). In fact, the author has applied the LDU and the ALS methods with a covariance-variance matrix $W(z)$ in these methods replaced by $\Sigma \otimes \Sigma$ and by comparing the obtained statistics to a chi-square distribution with $(G - L)(G - L + 1)/2$ degrees of freedom (see Section 5.2.3). Interestingly, the results were similar to those obtained in Tables 7.7–7.10. We decided not to include them as there is no theoretical justification yet for the use of these methods.

Rank of demand at $z = 1$						
h_x	h_z	1	2	3	4	5
	0.015	0.8143	0.9791	0.9892	0.9963	0.9949
0.1	0.02	0.9526	0.9989	0.9984	0.9979	0.9970
	0.05	0.9222	0.9999	0.9999	0.9999	0.9999
	0.08	0.4968	0.9999	0.9999	0.9999	0.9999
	0.3	0.0010	0.9507	0.9944	0.9969	0.9923
	0.015	0.9555	0.9949	0.9981	0.9989	0.9984
0.3	0.02	0.9713	0.9993	0.9997	0.9996	0.9982
	0.05	0.4295	0.9997	0.9999	0.9999	0.9997
	0.08	0.0107	0.9996	0.9999	0.9999	0.9988
	0.3	0.0000	0.9884	0.9994	0.9981	0.9839
	0.015	0.9424	0.9940	0.9967	0.9970	0.9969
0.5	0.02	0.9061	0.9896	0.9939	0.9949	0.9852
	0.05	0.0633	0.9911	0.9999	0.9998	0.9953
	0.08	0.0000	0.9920	0.9997	0.9993	0.9914
	0.3	0.0000	0.9196	0.9964	0.9900	0.9640

Table 7.8: P-values for hypothesis tests in a non-parametric model at $z = 1$

Rank of demand at $z = 1.2$						
h_x	h_z	1	2	3	4	5
0.1	0.05	0.6699	0.8691	0.9271	0.9549	0.9638
	0.09	0.6244	0.8314	0.9098	0.9450	0.9569
	0.15	0.2138	0.7563	0.9367	0.9308	0.8895
	0.2	0.0462	0.7044	0.9255	0.9179	0.9065
	0.3	0.0035	0.7754	0.9478	0.9530	0.9348
0.3	0.05	0.7723	0.9682	0.9839	0.9839	0.9722
	0.09	0.5779	0.9082	0.9412	0.9390	0.8829
	0.15	0.1470	0.9157	0.9249	0.9177	0.9006
	0.2	0.0113	0.9009	0.9318	0.9156	0.8917
	0.3	0.0000	0.9404	0.9761	0.9615	0.9175
0.5	0.05	0.5185	0.8980	0.9311	0.9107	0.8581
	0.09	0.4227	0.8798	0.9234	0.8937	0.8434
	0.15	0.0694	0.8893	0.9099	0.8853	0.8180
	0.2	0.0011	0.8358	0.8899	0.8864	0.8253
	0.3	0.0000	0.8469	0.9518	0.9322	0.8889

Table 7.9: P-values for hypothesis tests in a non-parametric model at $z = 1.2$

Rank of demand at $z = 0.9$						
h_x	h_z	1	2	3	4	5
0.1	0.04	0.1197	0.4977	0.7672	0.8831	0.8402
	0.08	0.2631	0.6509	0.9263	0.9923	0.9799
	0.11	0.4451	0.7383	0.9305	0.9925	0.9736
	0.17	0.4265	0.9216	0.9865	0.9975	0.9851
	0.3	0.0586	0.9296	0.9798	0.9946	0.9853
0.3	0.04	0.2271	0.9196	0.9753	0.9854	0.9799
	0.08	0.2050	0.8532	0.9537	0.9712	0.9389
	0.11	0.2314	0.8614	0.9663	0.9579	0.9131
	0.17	0.0532	0.9007	0.9881	0.9730	0.9345
	0.3	0.0003	0.9615	0.9961	0.9911	0.9635
0.5	0.04	0.2297	0.9407	0.9621	0.9617	0.9427
	0.08	0.2511	0.9219	0.9527	0.9454	0.9031
	0.11	0.1632	0.8819	0.9553	0.9408	0.8822
	0.17	0.0070	0.8476	0.9728	0.9469	0.8890
	0.3	0.0000	0.8661	0.9837	0.9659	0.9160

Table 7.10: P-values for hypothesis tests in a non-parametric model at $z = 0.9$

7.4 Simulation results

In this section, we use Monte Carlo simulations to examine small sample properties of tests for local ranks. We will focus on two such properties of a test, called a size and a power. The size of a test is defined as a frequency of rejecting the null hypothesis H_0 when H_0 is true. This frequency is computed at a given significance level from repeated test results in small or moderate size samples (hence the term "small sample property"). The power of a test is a frequency of rejecting the null hypothesis H_0 when, in fact, the alternative H_1 is true. Size and power provide information on two types of possible errors when making an inference concerning a test from a finite sample of observations and thus serve as a guiding tool in real life applications.

We will present here sizes and powers for local rank tests in the semi-parametric factor model only. Those for the non-parametric model will appear elsewhere. By the semi-parametric factor model, we mean here the (SPF) model defined in Section 3.1 and satisfying the assumptions of that section. The hypothesis tests are then those for the rank of the corresponding matrix $\Theta(z)$.

Remark 7.4.1 Alternatively, one may start with a semi-parametric factor system where the coordinate functions sum up to 1 and compute the small sample properties based on the maximum of the test statistics in all possible reduced systems (see Section 3.3.1). Although this approach is of a particular interest in the context of demand systems, we do not pursue it here for the sake of simplicity. The results of this section provide information on small sample properties of tests in each of the reduced systems and we expect that most of our observations apply to the situations where one looks at the maximum over all reduced demand systems as well.

The experimental set-up in the case of the semi-parametric factor model is as follows. We let $Y_i = \Theta(Z_i)V(X_i) + U_i$, $i = 1, \ldots, N$, where N is the sample size, Z_i and X_i are independent random variables uniformly distributed on the intervals $[-1, 1]$ and $[0, 1]$, respectively, and U_i, $i = 1, \ldots, N$, are independent $\mathcal{N}(0,1)$ random variables. The matrix $\Theta(z)$ and the vector $V(x)$ are given by

$$\Theta(z) = \begin{pmatrix} 1 & 1+z^2 & 1-2z^2 \\ 0 & \theta_{22}(z) & \theta_{23}(z) \\ 0 & 0 & \theta_{33}(z) \end{pmatrix} D(z), \quad V(x) = \begin{pmatrix} 1 \\ x \\ x^2 \end{pmatrix},$$

with three different choices for the entries $\widetilde{\Theta}(z) = (\theta_{22}(z), \theta_{23}(z), \theta_{33}(z))$, namely, $\widetilde{\Theta}_1(z) = (z, 0, z^2)$, $\widetilde{\Theta}_2(z) = (1/4 + z, 1/4, z)$ and $\widetilde{\Theta}_3(z) = (1/4 + z, 1/4, 1/6 + z)$, and two different choices for the 3×3 non-singular matrix $D(z)$, namely, $D(z) = I_3$ where I_3 is a 3×3 identity matrix, and $D(z)$ satisfying $D(z)Q(z)D(z)' = I_3$ where $Q(z) = p(z)E(V(X_1)V(X_1)'|Z_1 = z)$ is the matrix appearing in (3.2). These choices of $\widetilde{\Theta}(z)$ are motivated by the fact that we will consider local ranks for the semi-parametric factor model above at $z = 0$. In this case, the matrices $\widetilde{\Theta}_1(z)$, $\widetilde{\Theta}_2(z)$ and $\widetilde{\Theta}_3(z)$ correspond to the local ranks $L = 1$, $L = 2$ and $L = 3$, respectively, and hence cover all possible non-zero rank values for a 3×3 matrix. The size and power computations of a local rank test will then be based upon the results of the simulations for these three different choices of $\widetilde{\Theta}_i(z)$ and we will refer to them by saying that the (local) rank of the matrix $\Theta(z)$ is $L = 1$, $L = 2$ or $L = 3$.

To understand our choices of the matrix $D(z)$, recall from Section 5.1.3 that the minimum-χ^2 statistic for a test of rank L is the sum of the smallest $G - L$ eigenvalues of the estimator $\widehat{\Gamma}(z)$ of the matrix $\Gamma(z)$ defined in (5.21). If the smallest *non-zero* eigenvalue is close to zero, then the corresponding estimated eigenvalue will be close to zero as well and hence ranks estimated by the minimum-χ^2 test will be lower than those obtained in the case when the non-zero eigenvalues are larger. (This fact is well known in the statistical literature on ranks. See, for example, Cragg and Donald [19, 20].) The matrix $D(z)$ satisfying $D(z)Q(z)D(z)' = I_3$ allows to make non-zero eigenvalues of the matrix $\Gamma(z)$ larger. In our experimental set-up, if $D(z) = I_3$, then the eigenvalues of the matrix $\Gamma(z)$ at $z = 0$ are 0, 0.00281 and 1.879 when $L = 2$, and 0.000016, 0.0032 and 1.8817 when $L = 3$.

For $D(z)$ satisfying $D(z)Q(z)D(z)' = I_3$, the eigenvalues are 0, 0.0405 and 3.084 when $L = 2$, and 0.012, 0.046 and 3.094 when $L = 3$. (Note that the non-zero eigenvalues are indeed larger for the latter choice of $D(z)$.) Thus, by considering these two choices of $D(z)$, we will examine small sample properties both in the situations where the smallest non-zero eigenvalue of the matrix $\Gamma(z)$ is close to zero and in the situations where its distance to zero is larger.

The rank tests that we will consider are the minimum-χ^2, the asymptotic least squares and the LDU tests of Section 5.1. For the sample size N, we will take $N = 400$ and $N = 1000$. As for the bandwidth h which enters into the statistics of the three considered rank tests, we will take $h = 0.15$, 0.25 and 0.5. Let us observe in the spirit of Sections 7.2 and 7.3 that, when $N = 400$ ($N = 1000$, respectively), there are on average 60, 100 and 200 (150, 250 and 500, respectively) observations at the three chosen neighborhoods of $z = 0$, respectively.

Size calculations for tests of rank $L = 1$ and $L = 2$ with the above choices of test statistics, matrix $D(z)$ and parameter N and h values are presented in Table 7.11. The choice of D found in the first column indicates that we work either with $D(z)$ satisfying $D(z)Q(z)D(z)' = I_3$ (that is, $D(z) = Q(z)^{-1/2}$) or with $D(z) = I_3$. To compute the sizes of tests in Table 7.11, we use 1000 Monte Carlo replications and a 5 percent significance level. So, for example, to find the actual size of the LDU test when $L = 2$, $N = 400$, $D(z) = I_3$ and $h = 0.15$, we first compute the LDU statistics (with $h = 0.15$) for rank $L = 2$ test in 1000 replications of the semi-parametric factor model of the sample size $N = 400$ and with $D(z) = I_3$, and then compare these obtained statistics to the critical value of the limiting $\chi^2((3-2)(3-2)) = \chi^2(1)$ distribution at a 5 percent significance level. The actual size is just the frequency of those statistics that exceed the critical value (that is, their total number divided by the sample size). The sizes in other entries are obtained in an analogous way.

Based on the results of Table 7.11, we can draw the following conclusions. Observe

Size of tests						
Rank test			$L=1$		$L=2$	
D	Method	$h \setminus N$	400	1000	400	1000
$D=Q^{-1/2}$	min-χ^2	0.15	0.0	0.0	0.1	0.2
		0.25	0.0	0.0	0.1	0
		0.5	0.0	0.6	0.5	0.6
	LDU	0.15	28.7	19.5	0.2	3
		0.25	20.4	12	1.3	4.1
		0.5	18.8	15.4	5.2	8.2
	ALS	0.15	0.7	0.1	7.1	11.5
		0.25	0.3	0.1	9.5	10.5
		0.5	0.7	1.5	17.8	13.5
$D=I_3$	min-χ^2	0.15	5.3	4	0.6	0.8
		0.25	3.9	5.8	0.5	0.6
		0.5	4.9	4.8	0.4	1
	LDU	0.15	49.5	55.9	1.5	1.7
		0.25	55.8	62.4	1.2	2.2
		0.5	57.3	60.2	1.3	3.5
	ALS	0.15	41.6	31.2	3.8	3.4
		0.25	35.5	23.6	4.7	6.1
		0.5	25.8	19.7	4.5	6.7

Table 7.11: Size of tests in a semi-parametric factor model

that, when testing for small rank (namely, rank $L = 1$), the minimum-χ^2 test is the most undersized (as compared to the nominal size of 5 percent) and the LDU test is the most oversized for all values of h and N, and for both choices of D. Undersizing (oversizing, respectively) means that the rank is underestimated (overestimated, respectively) by a test. Comparing two tests, we may also say that, when sizes for one of the tests are larger, the ranks estimated by this test will be higher. Now, when testing for higher rank (namely, rank $L = 2$), the sizes are largest for the ALS test. This suggest that the rank estimated by the ALS test will be the highest. Interestingly, this is what we observed in the applications of the semi-parametric factor model in Section 7.2. Results of Table 7.11 also suggest that one would not reject a lower rank for smaller values of h because the sizes of tests decrease in most cases as h becomes smaller. This is also what we found in our earlier applications.

Power of tests								
True rank			$L=2$		$L=3$			
Rank test			$L=1$		$L=1$		$L=2$	
D	Method	$h \setminus N$	400	1000	400	1000	400	1000
$D=Q^{-1/2}$	min-χ^2	0.15	7.40	64.70	22.20	89.70	2.50	29.40
		0.25	29.00	97.90	62.60	100.00	10.50	62.30
		0.5	86.00	100.00	98.40	100.00	47.50	93.30
	LDU	0.15	33.30	65.00	43.90	81.69	0.30	6.40
		0.25	47.90	92.90	64.70	97.70	1.50	11.30
		0.5	81.30	99.70	90.10	100.00	13.40	43.30
	ALS	0.15	14.70	69.00	38.50	95.50	28.00	79.80
		0.25	37.10	96.59	73.00	99.90	56.49	93.80
		0.5	88.40	100.00	99.59	100.00	84.00	99.40
$D=I_3$	min-χ^2	0.15	6.40	9.50	6.00	9.90	0.30	0.90
		0.25	7.20	9.10	7.60	12.60	0.30	1.10
		0.5	9.60	20.80	12.40	20.20	1.00	1.40
	LDU	0.15	46.20	54.00	43.00	53.10	0.90	1.20
		0.25	49.50	55.49	48.40	53.20	1.10	1.40
		0.5	55.89	65.50	53.10	63.70	1.20	2.70
	ALS	0.15	41.10	32.99	42.30	42.30	4.20	4.70
		0.25	38.00	32.40	38.70	38.70	3.10	6.10
		0.5	33.09	32.99	36.10	36.10	4.90	8.10

Table 7.12: Power of tests in a semi-parametric factor model

Power calculations can be found in Tables 7.12 and 7.13 where we present rejection frequencies when testing for rank $L = 1$ in the semi-parametric factor models of ranks $L = 2$ and $L = 3$, and for rank $L = 2$ in the semi-parametric factor model of rank $L = 3$. Table 7.12 contains putative powers, that is, rejection frequencies computed by using the critical value at a significance level of 5 percent corresponding to the limiting distribution used in a test. Table 7.13 consists of size-adjusted powers, that is, rejection frequencies computed by using the critical value at which the actual size of a test in Table 7.11 is 5 percent. So, for example, to compute the size-adjusted power for the LDU test of rank 2 when $h = 0.25$, $N = 400$, $D_3 = I_3$ and the matrix $\Theta(z)$ in the semi-parametric factor model has rank $L = 3$, we first go back to the results on size calculations in this case and take the critical value that one would have used to attain a 5 percent rejection frequency

Size-adjusted power of tests								
True rank			$L=2$		$L=3$			
Rank test			$L=1$		$L=1$		$L=2$	
D	Method	$h \setminus N$	400	1000	400	1000	400	1000
$D=Q^{-1/2}$	min-χ^2	0.15	76.00	99.40	91.10	100.00	40.10	84.70
		0.25	94.30	100.00	99.20	100.00	59.20	94.89
		0.5	99.59	100.00	100.00	100.00	80.30	98.90
	LDU	0.15	4.60	3.10	0.40	7.60	6.49	13.30
		0.25	5.60	21.70	2.30	59.59	9.90	14.00
		0.5	22.5	92.70	58.40	98.90	12.40	14.50
	ALS	0.15	58.60	99.40	82.60	100.00	22.20	55.20
		0.25	93.10	100.00	98.90	100.00	40.40	82.20
		0.5	98.20	100.00	99.90	100.00	50.40	90.80
$D=I_3$	min-χ^2	0.15	5.80	10.50	5.70	11.50	6.00	5.30
		0.25	8.70	8.60	8.90	11.60	6.29	5.30
		0.5	9.60	21.40	12.40	21.00	4.60	5.30
	LDU	0.15	3.40	3.80	2.80	3.40	3.90	4.50
		0.25	4.20	5.30	3.00	3.90	5.20	3.30
		0.5	3.90	3.90	3.10	3.40	5.40	4.80
	ALS	0.15	5.30	5.10	5.10	5.50	5.40	7.80
		0.25	3.70	6.20	3.20	5.30	3.20	4.80
		0.5	6.80	6.90	6.40	8.50	5.40	6.10

Table 7.13: Size-adjusted power of tests in a semi-parametric factor model

when testing for rank $L=2$ with $h=0.25$, $N=400$ and $D=I_3$. The size adjusted power is then the frequency of those LDU statistics for rank $L=2$ test in 1000 replications that were larger than this adjusted critical value. Both putative and size-adjusted powers can be found in the statistical literature (see, for example, Cragg and Donald [19], p. 1307, and Cragg and Donald [18], pp. 229-230). Size-adjusted powers are thought to allow to compare results across different tests and tests with different parameter values. Putative powers are used because they provide information and feeling on how misleading the results are at the nominal significance levels.

Based on the results of Tables 7.12 and 7.13, we make the following observations. As already mentioned earlier, the results when $D(z)=Q(z)^{-1/2}$ are quite different from those

Size of tests ignoring z					
Rank test		$L=1$		$L=2$	
D	Method	$N=400$	$N=1000$	$N=400$	$N=1000$
$D=Q^{-1/2}$	min-χ^2	100.0	100.0	8.9	11.3
	LDU	100.0	100.0	11.8	11.5
	ALS	100.0	100.0	29.8	22.8
$D=I_3$	min-χ^2	13.1	26.7	1.4	1.2
	LDU	61.9	69.9	3.2	6.5
	ALS	24.2	33.1	6.7	11.4

Table 7.14: Size of tests in a semi-parametric factor model ignoring z

when $D(z) = I_3$. When $D(z) = Q(z)^{-1/2}$, the smallest and other eigenvalues of the matrix $\Gamma(z)$ at $z = 0$ are larger and hence the minimum-χ^2 and, interestingly, the other two tests reject a lower rank more often (larger power). Note also that, since the power decreases in most cases as h becomes smaller, the estimated rank will be lower for smaller h. This observation only confirms similar observations made from the results of Table 7.11 and in the applications of the semi-parametric factor model in Section 7.2. Let us also note from the results of Table 7.12 that, when testing in the semi-parametric factor model of rank $L = 3$, the rank estimated by the ALS test is the largest (since the powers of the ALS tests are the largest). This is also what we found in our applications in Section 7.2.

Remark 7.4.2 Observe from Tables 7.11, 7.12 and 7.13 that the power of tests improves as h becomes larger while, at the same time, the sizes of tests are typically comparable. (This is most notable for the minimum-χ^2 test.) This observations suggests that, in practice, one may prefer the values of h which are larger.

Finally, in Tables 7.14 and 7.15, we present size and (putative) power computations for rank tests in the semi-parametric factor model when the variable z is ignored. In other words, we generate the data according to the semi-parametric factor model chosen above but when testing for different ranks, we ignore the variables z in the model altogether. Observe from Tables 7.14 and 7.15 that both size and power of tests increase as compared

Power of tests ignoring z							
True rank		$L=2$		$L=3$			
Rank test		$L=1$		$L=1$		$L=2$	
D	Method	$N=400$	$N=1000$	$N=400$	$N=1000$	$N=400$	$N=1000$
$D=Q^{-1/2}$	min-χ^2	99.30	100.00	99.90	100.00	70.70	97.70
	LDU	90.50	99.09	99.20	100.00	30.70	89.30
	ALS	99.50	100.00	99.90	100.00	79.89	98.59
$D=I_3$	min-χ^2	16.10	35.70	16.90	40.50	1.00	2.40
	LDU	58.20	62.30	58.20	63.10	2.60	5.90
	ALS	29.10	42.30	31.10	46.40	6.40	10.40

Table 7.15: Power of tests in a semi-parametric factor model ignoring z

to those in Tables 7.11 and 7.12 for all three types of tests and all parameter N, h and D values. Hence, the rank estimated ignoring the variable z is higher than that obtained when z is taken into account. Moreover, it is likely to be higher than the true (local) rank because tests are clearly oversized in most cases according to Table 7.14. These observations agree with that made in the applications of the semi-parametric factor model in Section 7.2.

Bibliography

[1] U.S. Census Bureau, Census 2000 Redistricting Data (P.L. 94-171) Summary File and 1990 Census. Available at: http://www.census.gov.

[2] ACCRA Cost of Living Index Manual. ACCRA. Available at: http://www.accra.org.

[3] Metropolitan Areas 1999. Statistical Policy Office. Office of Management and Budget. Attachments to OMB Bulletin No. 99-04. Available at: http://www.whitehouse.gov/omb/bulletins/99-04att-1.pdf.

[4] U.S. Dept. of Labor, Bureau of Labor Statistics. Consumer Expenditure Survey, 1999: Interview Survey and Detailed Expenditure Files. 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.

[5] ACCRA Cost of Living data for First Quarter 2000. ACCRA, 2000.

[6] Hirotugu Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.

[7] T. W. Anderson. The asymptotic distribution of certain characteristic roots and vectors. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950*, pages 103–130, Berkeley and Los Angeles, 1951. University of California Press.

[8] T. W. Anderson. Estimating linear restrictions on regression coefficients for multivariate normal distributions. *The Annals of Mathematical Statistics*, 22(3):327–351, 1951.

[9] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons Inc., New York, 1958.

[10] James Banks, Richard Blundell, and Arthur Lewbel. Quadratic Engel curves and consumer demand. *The Review of Economics and Statistics*, 79(4):527–539, 1997.

[11] Peter Bauer, Benedikt M. Pötscher, and Peter Hackl. Model selection by multiple test procedures. *Statistics*, 19(1):39–44, 1988.

[12] Ernst R. Berndt and Eugine N. Savin. Estimation and hypothesis testing in singular equation systems with autoregressive disturbances. *Econometrica*, 43(5/6):937–958, 1975.

[13] P. J. Bickel and M. Rosenblatt. On some global measures of the deviations of density function estimates. *The Annals of Statistics*, 1(6):1071–1095, 1973.

[14] Patrick Billingsley. *Convergence of Probability Measures*. John Wiley & Sons Inc., New York, 1968.

[15] David Cass and Joseph E. Stiglitz. The structure of investor preferences and asset returns, and separability in portfolio allocation: a contribution to the pure theory of mutual funds. *Journal of Economic Theory*, 2(2):122–160, 1970.

[16] Gary Chamberlain. Multivariate regression models for panel data. *Journal of Econometrics*, 18(1):5–46, 1982.

[17] Gary Chamberlain. Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, 34(3):305–334, 1987.

[18] John G. Cragg and Stephen G. Donald. Testing identifiability and specification in instrumental variable models. *Econometric Theory*, 9(2):222–240, 1993.

[19] John G. Cragg and Stephen G. Donald. On the asymptotic properties of LDU-based tests of the rank of a matrix. *Journal of the American Statistical Association*, 91(435):1301–1309, 1996.

[20] John G. Cragg and Stephen G. Donald. Inferring the rank of a matrix. *Journal of Econometrics*, 76(1-2):223–250, 1997.

[21] Harald Cramér. *Mathematical Methods of Statistics.* Princeton University Press, Princeton, N. J., 1946.

[22] Peter de Jong. A central limit theorem for generalized quadratic forms. *Probability Theory and Related Fields*, 75(2):261–277, 1987.

[23] A. Deaton and J. Muellbauer. *Economics and Consumer Behavior.* Cambridge University Press, Cambridge, U. K., 1980.

[24] Angus Deaton and John Muellbauer. An almost ideal demand system. *The American Economic Review*, 70(3):312–326, 1980.

[25] Luc Devroye. *A Course in Density Estimation.* Birkhäuser Boston Inc., Boston, MA, 1987.

[26] Luc Devroye and László Györfi. *Nonparametric density estimation.* John Wiley & Sons Inc., New York, 1985. The L_1 view.

[27] W. E. Diewert. Generalized Slutsky conditions for aggregate consumer demand functions. *Journal of Economic Theory*, 15(2):353–362, 1977.

[28] Stephen G. Donald. Inference concerning the number of factors in a multivariate nonparametric relationship. *Econometrica*, 65(1):103–131, 1997.

[29] Paul Embrechts, Claudia Klüppelberg, and Thomas Mikosch. *Modelling Extremal Events.* Springer-Verlag, Berlin, 1997. For insurance and finance.

[30] E. Estes and B. Honoré. Partially linear regression using one nearest neighbor. Princeton University, Department of Economics, Preprint, 1995.

[31] J. Fan and I. Gijbels. *Local Polynomial Modelling and its Applications.* Chapman & Hall, London, 1996.

[32] Jianqing Fan and Wenyang Zhang. Statistical estimation in varying coefficient models. *The Annals of Statistics*, 27(5):1491–1518, 1999.

[33] Thomas Ferguson. A method of generating best asymptotically mormal estimates with application to the estimation of bacterial densities. *The Annals of Mathematical Statistics*, 29(4):1046–1062, 1958.

[34] David M. Frankel and Eric D. Gould. The retail price of inequality. *Journal of Urban Economics*, 49(2):219–239, 2001.

[35] Xavier Freixas and Andreu Mas-Colell. Engel curves leading to the weak axiom in the aggregate. *Econometrica*, 55(3):515–531, 1987.

[36] Vanessa Fry and Panos Pashardes. An almost ideal quadratic logarithmic demand analysis for the analysis of micro data. Applied Economics Discussion Paper Series 145, Institute of Economics and Statistics, University of Oxford, Manor Road, Oxford, 1992.

[37] Yasunori Fujikoshi. Asymptotic expansions for the distributions of some multivariate tests. In *Multivariate analysis, IV (Proc. Fourth Internat.. Sympos., Dayton, Ohio, 1975)*, pages 55–71. North-Holland, Amsterdam, 1977.

[38] Len Gill and Arthur Lewbel. Testing the rank and definiteness of estimated matrices with applications to factor, state-space and ARMA models. *J. Amer. Statist. Assoc.*, 87(419):766–776, 1992.

[39] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, MD, second edition, 1989.

[40] W. M. Gorman. Separable utility and aggregation. *Econometrica*, 27(3):469–481, 1959.

[41] W. M. Gorman. Some Engel curves. In Angus Deaton, editor, *Essays in The theory and measurement of consumer behaviour in honour of Sir Richard Stone*, pages 7-29. Cambridge University Press, Cambridge, 1981.

[42] C. Gouriéroux and A. Monfort. A general framework for testing a null hypothesis in a "mixed" form. *Econometric Theory*, 5(1):63–82, 1989.

[43] Christian Gouriéroux, Alain Monfort, and Eric Renault. Tests sur le noyau, l'image et le rang de la matrice des coefficients d'un modèle linéaire multivarié. *Annales d'Économie et de Statistique*, 32:81–111, 1993.

[44] Christian Gouriéroux, Alain Monfort, and Alain Trognon. Moindres carrés asymptotiques. *Annales de l'I.N.S.É.É.*, 58:91–122, 1985.

[45] B. Grodal and Werner Hildenbrand. Cross-section Engel curves, expenditure distributions and the law of demand. In L. Philips and L. D. Taylor, editors, *Essays in Honor of H. S. Houthakker*. Kluwer Academic Publisher, Dordrecht, 1992.

[46] Peter Hall. Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of Multivariate Analysis*, 14(1):1–16, 1984.

[47] Lars Peter Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054, 1982.

[48] Wolfgang Härdle. *Applied Nonparametric Regression*. Cambridge University Press, Cambridge, 1990.

[49] Wolfgang Härdle. *Smoothing Techniques.* Springer-Verlag, New York, 1991. With implementation in S.

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

[51] Werner Hildenbrand. *Market Demand: Theory and Empirical Evidence.* Princeton University Press, Princeton, New Jersey, 1994.

[52] Howard Howe, Robert A. Pollak, and Terence J. Wales. Theory and time series estimation of the quadratic expenditure system. *Econometrica*, 47(5):1231–1247, 1979.

[53] P. L. Hsu. On the problem of rank and the limiting distribution of Fisher's test function. *Annals of Eugenics*, 11:39–41, 1941.

[54] Gordon J. Johnston. Smooth nonparametric regression analysis. Mimeo series 1253, Department of Statistics, University of North Carolina, Chapel Hill, North Carolina, 1979.

[55] Alois Kneip. Nonparametric estimation of common regressors for similar curve data. *The Annals of Statistics*, 22(3):1386–1427, 1994.

[56] J. Komlós, P. Major, and G. Tusnády. An approximation of partial sums of independent rv's and the sample df. I. *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete*, 32:111–131, 1975.

[57] M. R. Leadbetter, Georg Lindgren, and Holger Rootzén. *Extremes and Related Properties of Random Sequences and Processes.* Springer-Verlag, New York, 1983.

[58] C. E. V. Leser. Forms of Engel functions. *Econometrica*, 31(4):694–703, 1963.

[59] Arthur Lewbel. Characterizing some Gorman Engel curves. *Econometrica*, 55(6):1451–1459, 1987.

[60] Arthur Lewbel. Fractional demand systems. *Journal of Econometrics*, 36(3):311–337, 1987.

[61] Arthur Lewbel. A demand system rank theorem. *Econometrica*, 57(3):701–705, 1989.

[62] Arthur Lewbel. Full rank demand systems. *International Economic Review*, 31(2):289–300, 1990.

[63] Arthur Lewbel. The rank of demand systems: theory and nonparametric estimation. *Econometrica*, 59(3):711–730, 1991.

[64] Arthur Lewbel. Utility functions and global regularity of fractional demand systems. *International Economic Review*, 36(4):943–961, 1995.

[65] Arthur Lewbel. Consumer demand systems and household equivalence scales. In M. Hashem Pesaran and Peter Schmidt, editors, *Handbook of Applied Econometrics: Microeconomics*, volume II, pages 167–201. Blackwell Publishers, Oxford, 1997.

[66] Arthur Lewbel. Rank, separability, and conditional demands. Preprint, 2000.

[67] Arthur Lewbel. A rational rank four demand system. Preprint, 2000.

[68] Arthur Lewbel and William Perraudin. A theorem on portfolio separation with general preferences. *Journal of Economic Theory*, 65(2):624–626, 1995.

[69] Qi Li, Cliff J. Huang, Dong Li, and Tsu-Tan Fu. Semiparametric smooth coefficient models. Preprint, 2001.

[70] Keh Shin Lii. A global measure of a spline density estimate. *The Annals of Statistics*, 6(5):1138–1148, 1978.

[71] Panayiota Lyssiotou, Panos Pashardes, and Thanasis Stengos. Preference heterogeneity and the rank of demand systems. *Journal of Business and Economic Statistics*, 17(2):248–252, 1999.

[72] Panayiota Lyssiotou, Panos Pashardes, and Thanasis Stengos. Testing the rank of engel curves with endogenous expenditure. *Economics Letters*, 64(1):61–65, 1999.

[73] Jan R. Magnus and Heinz Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons Ltd., Chichester, 1999. Revised reprint of the 1988 original.

[74] John Muellbauer. Aggregation, income distribution and consumer demand. *The Review of Economic Studies*, 42(4):525–543, 1975.

[75] John Muellbauer. Community preferences and the representative consumer. *Econometrica*, 44(5):979–999, 1976.

[76] Christopher J. Nicol. The rank and model specification of demand systems: an empirical analysis using united states microdata. *Canadian Journal of Economics*, 34(1):259–289, 2001.

[77] Adrian Pagan and Aman Ullah. *Nonparametric Econometrics*. Cambridge University Press, Cambridge, 1999.

[78] Robert A. Pollak and Terence J. Wales. Estimation of complete demand systems from household budget data: the linear and quadratic expenditure systems. *The American Economic Review*, 68(3):348–359, 1978.

[79] Robert A. Pollak and Terence J. Wales. Comparison of the quadratic expenditure system and translog demand systems with alternative specifications of demographic effects. *Econometrica*, 48(3):595–612, 1980.

[80] Robert A. Pollak and Terence J. Wales. *Demand System Specification and Estimation*. Oxford University Press, New York, 1992.

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

[82] James L. Powell, James H. Stock, and Thomas M. Stoker. Semiparametric estimation of index coefficients. *Econometrica*, 57(6):1403–1430, 1989.

[83] B. L. S. Prakasa Rao. *Nonparametric Functional Estimation.* Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1983.

[84] C. Radhakrishna Rao. *Linear Statistical Inference and its Applications.* John Wiley & Sons, New York-London-Sydney, second edition, 1973. Wiley Series in Probability and Mathematical Statistics.

[85] C. Radhakrishna Rao and M. Bhaskara Rao. *Matrix Algebra and its Applications to Statistics and Econometrics.* World Scientific Publishing Co. Inc., River Edge, NJ, 1998.

[86] Zaka Ratsimalahelo. Rank test based on matrix perturbation theory. Preprint, 2000.

[87] Sidney I. Resnick. *Extreme Values, Regular Variation, and Point Processes.* Springer-Verlag, New York, 1987.

[88] J.-M. Robin and R. J. Smith. Tests of rank. DAE Working Paper 9521, Department of Applied Economics, University of Cambridge, Cambridge, U. K., 1995.

[89] Jean-Marc Robin and Richard J. Smith. Tests of rank. *Econometric Theory*, 16(2):151–175, 2000.

[90] Thomas J. Rothenberg. *Efficient Estimation with a Priori Information.* Yale University Press, New Haven, Conn., 1973. Clowes Foundation for Research in Economics at Yale University, Monograph 23.

[91] René Roy. *De l'Utilité: Contribution à la Théorie des Choix.* Hermann, Paris, 1942.

[92] Thomas Russell and Frank Farris. The geometric structure of some systems of demand equations. *Journal of Mathematical Economics*, 22(4):309–325, 1993.

[93] Peter Schmidt. *Econometrics.* Marcel Dekker, New York, 1976.

[94] Gideon Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

[95] Robert J. Serfling. *Approximation Theorems of Mathematical Statistics.* John Wiley & Sons Inc., New York, 1980. Wiley Series in Probability and Mathematical Statistics.

[96] Ronald W. Shephard. *Cost and Production Functions.* Princeton University Press, Princeton, New Jersey, 1953.

[97] B. W. Silverman. *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, New York, 1986.

[98] G. W. Stewart and Ji Guang Sun. *Matrix Perturbation Theory.* Academic Press Inc., Boston, MA, 1990.

[99] Thomas M. Stoker. Exact aggregation and generalized Slutsky conditions. *Journal of Economic Theory*, 33(2):368–377, 1984.

[100] Jerzy Szroeter. Generalized Wald methods for testing nonlinear implicit and overidentifying restrictions. *Econometrica*, 51(2):335–353, 1983.

[101] Halbert White and Yongmiao Hong. M-testing using finite and infinite dimensional parameter estimators. Discussion paper 93-01R, Department of Economics, University of California, San Diego, 1999.

[102] Holbrook Working. Statistical laws of family expenditure. *Journal of the American Statistical Association*, 38(221):43–56, 1943.

[103] A. Yatchew. An elementary estimator of the partial linear model. *Economics Letters*, 57(2):135–143, 1997.

Vita

Miss Natércia Fortuna was born near Porto city in Portugal. She graduated from Universidade Lusíada in 1994 and then joined the faculty of Economics at Universidade do Porto where she worked as a lecturer. Since 1997, Miss Fortuna has been at Boston University, first, in the department of Economics where she received a Master of Arts and later, in the department of Mathematics and Statistics. Her research and professional interests lie in Econometrics.