SERIE RESEARCH MEMORANDA

SAMPLE MOMENTS INTEGRATING NORMAL KERNEL ESTIMATORS

Herman J. Bierens

Research Memorandum 1988-60 Dec. '88

VRIJE UNIVERSITEIT
FACULTEIT DER ECONOMISCHE WETENSCHAPPEN
EN ECONOMETRIE
AMSTERDAM

SAMPLE MOMENTS INTEGRATING NORMAL KERNEL ESTIMATORS *

Herman J.Bierens
Department of Econometrics
Free University, Amsterdam

December 1988

* Chapter 7 of:
PRINCIPLES OF NONLINEAR AND NONPARAMETRIC REGRESSION ANALYSIS
To be published by Marcel Dekker Inc., New York.

PRINCIPLES OF NONLINEAR AND NONPARAMETRIC REGRESSION ANALYSIS

TABLE OF CONTENTS:

PREFACE

This book deals with statistical inference of nonlinear regression models from two opposite points of view, namely the case where the functional form of the model is completely specified as a known function of regressors and unknown parameters, and the opposite case where the functional form of the model is completely unknown. First it is assumed that the response function of the regression model under review belongs to *a certain well-specified parametric family of functional forms*, by which estimation of the model merely amounts to estimation of the unknown parameters. For this class of models we review the asymptotic properties of the nonlinear least squares estimator for independent data as well as for time series.

In practice assumptions on the functional form are often made on the basis of computational convenience rather than on the basis of precise a priori knowledge of the empirical phenomenon under review. Therefore the linear regression model is still the most popular model specification in applied research. However, even if the specification of the functional form is based on sound theoretical considerations there is quite often a large range of functional forms that are theoretically admissible, so that there is no guarantee that the actually chosen functional form is true. Functional specification of a parametric nonlinear regression model should therefore always be verified by conducting model misspecification tests. Various model misspecification tests will therefore be discussed, in particular consistent tests which have asymptotic power 1 against all deviations from the null hypothesis that the model is correct.

The opposite case of parametric regression is nonparametric regression. Nonparametric regression analysis is concerned with estimation of a regression model without specifying in advance its functional form. Thus the only source of information about the functional form of the model is the data set itself. In this book we shall review various nonparametric regression approaches, with special emphasis on the kernel method, under various distributional assumptions.

This book is divided into three parts. In the first part we review the elements of abstract probability theory we need in part 2. Part 2 is devoted to the asymptotic theory of parametric and nonparametric regression analysis in the case of independent data generating processes. In part 3 we extend the analysis involved to time series.

The selection of the topics mainly reflexes my own interest in the subject. Instead of providing an encyclopedic survey of the literature, I have chosen for a setup which aims to fill the gap between intermediate statistics (including linear time series analysis) and the level necessary to get access to the recent literature on nonlinear and nonparametric regression analysis, with emphasis on my own contributions. The ultimate goal is to provide the student with the tools for his own independent research in this area, by showing what tools I and others have used and what they have been used for. Thus, this book may be viewed as an account of my own struggle with the material involved. I think this book is particularly suitable for self-tuition (at least it aims to be), and may prove useful in a graduate course in mathematical statistics and advanced econometrics.

Acknowledgements:

The first five chapters of this book have been disseminated in draft form as working papers. I am grateful to Anil Bera, Alexander Georgiev and Jan Magnus for suggesting additional references, and in particular to Lourens Broersma, Johan Smits and Ton Steerneman who suggested various improvements.

A large body of the material in chapter 6 has been published earlier in Truman F.Bewley (ed.), *Advances in Econometrics, Fifth World Congress*, Cambridge University Press. I am indebted to Cambridge University Press for granting permission to reprint it.

7. SAMPLE MOMENTS INTEGRATING NORMAL KERNEL ESTIMATORS OF MULTIVARIATE DENSITY AND REGRESSION FUNCTIONS

In this chapter we discuss an alternative class of kernel estimators of multivariate density and regression functions proposed by Bierens (1983). We shall construct a normal kernel density functions estimator which is a density itself and satisfies the condition that its first and second moment integrals equal the first and second sample moments, respectively. This class of density estimators will be used for constructing a class of regression function estimators. Moreover, we shall propose a new approach for estimating appropriate window width parameters. The virtue of this class of kernel estimators is that the kernel density estimator involved contains the maximum likelihood estimator of the multivariate normal density as a special case, whereas the kernel regression estimator contains the linear regression model as a special case.

As far as asymptotic theory is concerned we confine our attention to uniform weak consistency.

7.1 *Kernel density estimators*

7.1.1 *Integral conditions*

Let $X_1, X_2, \ldots, X_n, \ldots$ be an i.i.d. stochastic process in R^k and suppose that the distribution of the X_j's is absolutely continuous with density $h(x)$. We recall that a kernel estimator of $h(x)$ is a random function of the form

$$\hat{h}(x) = (1/n)\Sigma_{j=1}^{n} K((x-X_j)/\gamma_n)\gamma_n^{-k} \tag{7.1.1}$$

where K is a real function on R^k, called the kernel, satisfying $\int K(x)dx=1$, and (γ_n) is a sequence of positive numbers, called window width parameters, converging to zero as $n\to\infty$.

For a good estimator of a function we should require that the estimator has similar properties as the function to be estimated. An obvious property of a density is that it is nonnegative and integrates to 1. Therefore we let the kernel K be a density itself, so that we then have

Condition a: $\hat{h}(x) \geq 0$, $\int\hat{h}(x)dx = 1$ a.s.

A further reasonable requirement is that the integral $\int x\hat{h}(x)dx$ is an appropriate estimator of the mathematical expectation $\int xh(x)dx$. Since the usual estimator of the latter integral is the sample mean

$$\bar{X} = (1/n)\Sigma_{j=1}^{n} X_j ,$$

we therefore require:

Condition b: $\int x\hat{h}(x)dx = \bar{X}$ a.s.

Since

$$\int x\hat{h}(x)dx = (1/n)\Sigma_{j=1}^{n}\int xK((x-X_j)/\gamma_n)\gamma^{-k}dx = \bar{X} + \gamma_n\int xK(x)dx,$$

we see that the kernel estimator satisfies conditions (a) and (b) if the kernel is a density with zero mean vector. We may go on this way by requiring that also the matrix $\int xx'h(x)dx$ is an appropriate estimator of $\int xx'h(x)dx$. The usual estimator of the latter matrix is the matrix

$$(1/n)\Sigma_{j=1}^{n} X_j X_j' .$$

Therefore we require:

Condition c: $\int xx'\hat{h}(x)dx = (1/n)\Sigma_{j=1}^{n} X_j X_j'$ a.s.

But if the conditions (a) and (b) hold, condition (c) cannot be satisfied, for

$$\int xx'\hat{h}(x)dx = (1/n)\Sigma_{j=1}^{n}\int xx'K((x-x_j)/\gamma_n)\gamma_n^{-k}dx =$$

$$= (1/n)\Sigma_{j=1}^{n} X_j X_j' + \gamma_n\bar{X}\int x'K(x)dx$$

$$+ \gamma_n(\int xK(x)dx)\bar{X}' + \gamma_n^2\int xx'K(x)dx$$

$$= (1/n)\Sigma_{j=1}^{n} X_j X_j' + \gamma_n^2\int xx'K(x)dx,$$

which dominates

$$(1/n)\Sigma_{j=1}^{n} X_j X_j'$$

by a positive definite matrix $\gamma_n^2\int xx'K(x)dx$.

7.1.2 *Sample moments integrating kernel density estimators*

We now shall construct a kernel density estimator which is (uniformly) consistent and satisfies the conditions (a), (b) and (c). The trick is to replace the X_j's in the formula (7.1.1) by $(1-\beta_n)X_j$, where (β_n) is a sequence of positive numbers converging to zero for $n\to\infty$. Thus we consider

$$\hat{h}^*(x) = (1/n)\Sigma_{j=1}^n K_*[(x-(1-\beta_n)X_j)/\gamma_n]\gamma_n^{-k} \tag{7.1.2}$$

where K_* is a density (possibly random). Clearly condition (a) is satisfied. Moreover, condition (b) now becomes:

$$\overline{X} = \int x\hat{h}^*(x)dx = (1/n)\Sigma_{j=1}^n\int[(1-\beta_n)X_j + \gamma_n x]K_*(x)dx$$

$$= (1-\beta_n)\overline{X} + \gamma_n\int xK_*(x)dx,$$

hence

$$\int xK_*(x)dx = (\beta_n/\gamma_n)\overline{X}. \tag{7.1.3}$$

Consequently, $K_*(x)$ is indeed a random function. Furthermore, condition (c) now becomes

$$(1/n)\Sigma_{j=1}^n X_jX_j' = \int xx'\hat{h}^*(x)dx$$

$$= (1/n)\Sigma_{j=1}^n\int[(1-\beta_n)X_j+\gamma_n x][(1-\beta_n)X_j+\gamma_n x]'K_*(x)dx$$

$$= (1-\beta_n)^2(1/n)\Sigma_{j=1}^n X_jX_j' + \gamma_n(1-\beta_n)\overline{X}\int x'K_*(x)dx$$

$$+ \gamma_n(1-\beta_n)(\int xK_*(x)dx)\overline{X}' + \gamma_n^2\int xx'K_*(x)dx$$

$$= (1-\beta_n)^2(1/n)\Sigma_{j=1}^n X_jX_j' + 2\beta_n(1-\beta_n)\overline{X}\,\overline{X}' + \gamma_n^2\int xx'K_*(x)dx, \tag{7.1.4}$$

where the latter equality follows from (7.1.3). Obviously, (7.1.3) and (7.1.4) imply

$$\int xx'K_*(x)dx = \gamma_n^{-2}\{(2\beta_n-\beta_n^2)(1/n)\Sigma_{j=1}^n X_jX_j' - (2\beta_n-2\beta_n^2)\overline{X}\,\overline{X}'\}$$

$$= \{(2\beta_n-\beta_n^2)/\gamma_n^2\}\{(1/n)\Sigma_{j=1}^n X_jX_j' - \overline{X}\,\overline{X}'\} + \int xK_*(x)dx\int x'K_*(x)dx.$$

Thus in order that $\hat{h}^*(x)$ satisfies the conditions (a), (b) and (c), $K_*(.)$ should be a density with mean vector and variance matrix

$$(\beta_n/\gamma_n)\overline{X}, \quad \{(2\beta_n-\beta_n^2)/\gamma_n^2\}\hat{V},$$

respectively, where

$$\hat{V} = (1/n)\Sigma_{j=1}^n (X_j-\overline{X})(X_j-\overline{X})'$$

is the sample variance matrix. Therefore, if we put

$$K_*(x) = K[(((2\beta_n-\beta_n^2)/\gamma_n^2)\hat{V})^{-\frac{1}{2}}(x-(\beta_n/\gamma_n)\overline{X})]$$
$$/ \ [\det\{((2\beta_n-\beta_n^2)/\gamma_n^2)\hat{V})\}^{\frac{1}{2}}] \qquad (7.1.5)$$

where K is a density with zero mean vector and variance equal to the unit matrix:

$$K(x) \geq 0, \ \int K(x)dx = 1, \ \int xK(x)dx = 0, \ \int xx'K(x)dx = I, \quad (7.1.6)$$

then clearly K_* has all the required properties. Now substitute (7.1.5) in (7.1.2). We then obtain

$$\hat{h}^*(x) = (1/n)\Sigma_{j=1}^n K\{\hat{V}^{-\frac{1}{2}}[x-(1-\beta_n)X_j \ - \ \beta_n\overline{X}]/\surd(2\beta_n-\beta_n^2)\}$$
$$/ \ \{[\surd(2\beta_n-\beta_n^2)]^k\surd\det(\hat{V})\} \qquad (7.1.7)$$

We now see that the window width parameter γ_n has been disappeared and that β_n has taken over its role. However, for convenience we shall introduce a new γ_n by putting

$$\gamma_n = (2\beta_n-\beta_n^2)^{\frac{1}{2}},$$

so that (7.1.7) now becomes

$$\hat{h}^*(x) = (1/n)\Sigma_{j=1}^n K[\hat{V}^{-\frac{1}{2}}\{x-\surd(1-\gamma_n^2)X_j-(1-\surd(1-\gamma_n^2))\overline{X}\}/\gamma_n]$$
$$/[\gamma_n^k \ \surd\det(\hat{V})]$$

Of course, γ_n should be chosen in the interval (0,1], as otherwise (7.1.7) will no longer be a *real* random function. We have shown by now:

Theorem 7.1.1. Let $X_1,\dots,X_n$ be random vectors in R^k. Define

$$\hat{h}(x|\gamma) = (1/n)\Sigma_{j=1}^{n}K[\hat{V}^{-\frac{1}{2}}(x-\surd(1-\gamma^2)X_j - (1-\surd(1-\gamma^2))\bar{X})/\gamma]$$

$$/[\gamma^k\surd\det(\hat{V})]$$

where K is a real function on R^k satisfying condition (7.1.6), $\bar{X}$ is the sample mean and V is the sample variance matrix. Then for every $\gamma\in(0,1]$, $\hat{h}(x|\gamma)$ satisfies the conditions (a), (b) and (c).

However, can we select a sequence (γ_n) with $\gamma_n\in(0,1]$ such that $\hat{h}(x|\gamma_n)$ is a (uniformly) consistent estimator of the density $h(x)$ of the common distribution of the X_j's? The answer depends on the distribution of the X_j's and on the kernel K. For the case that K is chosen to be the density of the k-variate standard normal distribution, i.e.,

$$K(x) = \exp[-\tfrac{1}{2}x'x]/(\surd(2\pi))^k, \tag{7.1.8}$$

we have the following result.

Theorem 7.1.2. Let $X_1,X_2,\dots,X_n,\dots$ be an i.i.d. stochastic process in R^k. Let the distribution of the X_j's be absolutely continuous with continuous density $h(x)$ and finite second moments. Let (ζ_n) be any sequence satisfying

$$\zeta_n\in(0,1],\ \lim_{n\to\infty}\zeta_n = 0 \text{ and } \lim_{n\to\infty}\surd(n\zeta_n^k) = \infty,$$

and let (γ_n) be any (random) sequence satisfying

$$\gamma_n\in[\zeta_n,1] \text{ (a.s.)},\ (p)\lim_{n\to\infty}\gamma_n = 0.$$

If

$$\text{plim}_{n\to\infty}\bar{X} = \mu \text{ and } \text{plim}_{n\to\infty}\hat{V} = V \text{ exist},$$

where V is positive definite, then

$$\text{plim}_{n\to\infty}\sup_{x\in R^k}|\hat{h}(x|\gamma_n) - h(x)| = 0$$

where $\hat{h}(x|\gamma)$ is defined as in theorem 7.1.1 with K defined by (7.1.8).

Proof: Section 7.4.

In the sequel of this chapter we shall only consider kernel density estimators based on normal kernels. The main reason for this is that normal kernels are easy to work with. For example the proof of theorem 2 becomes rather messy if we employ other kernels than (7.1.8). A further advantage is that normal kernels are everywhere positive, hence so is a normal kernel density estimator and therefore ratio's of normal kernel density estimators are always defined. These ratio's will play a role in estimating regression functions, as we shall see in section 7.2. Moreover, we recall that a normal kernel is nearly optimal in the sense of Epanechnikov (1969).

The kernel density estimator considered in theorem 7.1.2, i.e.,

$$\hat{h}(x|\gamma) = (1/n)\Sigma_{j=1}^{n}\hat{h}_j(x|\gamma), \tag{7.1.9}$$

with

$$\hat{h}_j(x|\gamma) = \exp(-\tfrac{1}{2}(x-\surd(1-\gamma^2)X_j-(1-\surd(1-\gamma^2))\bar{X})'\hat{V}^{-1}$$
$$\times\ (x-\surd(1-\gamma^2)X_j-(1-\surd(1-\gamma^2))\bar{X})/\gamma^2)/\ (\surd(2\pi))^k\gamma^k\surd\det(\hat{V}) \tag{7.1.10}$$

will be called the *Sample Moments Integrating Normal Kernel* (*SMINK*) estimator of a density function, because it satisfies the integral conditions (a), (b) and (c) for every $\gamma\in(0,1]$ and it is based on a normal kernel.

The reader is invited to verify that the marginal densities of a SMINK density estimator are SMINK density estimators themselves. Cf. exercise 1. Moreover, observe that

$$\hat{h}(x|1) = \exp[-\tfrac{1}{2}(x-\bar{X})'\hat{V}^{-1}(x-\bar{X})]/(\surd 2\pi)^k\surd\det(\hat{V})$$

is just the maximum likelihood estimator of a k-variate normal Density.

7.1.3 *The Dirac-catastrophe and the modified SMINK density estimator*

It is easy to verify from (7.1.9) and (7.1.10) that

$$\lim_{\gamma\downarrow 0}\hat{h}(x|\gamma) = \infty \text{ if } x \in \{X_1,\ldots,X_n\}$$

$$\lim_{\gamma\downarrow 0}\hat{h}(x|\gamma) = 0 \text{ if } x \notin \{X_1,\ldots,X_n\}$$

[Cf. exercise 2]. This is the so-called Dirac-catastrophe (see Good and Gaskins (1972)). Thus if γ is too small then the estimate $\hat{h}(x|\gamma)$ will go wild in that its shape then takes the form of a collection of high peaks each corresponding with an observation X_j. A possible way to get around the Dirac-catastrophe is to modify (7.1.9) by leaving out the $h_j(x|\gamma)$ for which the corresponding observation X_j equals x, i.e., to modify (7.1.9) to

$$(1/n)\Sigma_{j=1}^{n} I(x_j \neq x)\hat{h}_j(x|\gamma), \qquad (7.1.11)$$

where I(.) is the well-known indicator function. However, we have seen that the marginal densities of a SMINK density estimator are SMINK density estimators themselves. This neat property does not carry over to estimators of the type (7.1.11), but it does for the following class of density estimators

$$\hat{h}^*(x|\gamma) = (1/n)\Sigma_{j=1}^{n}\{\Pi_{i=1}^{k} I(x^{(i)} \neq X_{ij})\}\hat{h}_j(x|\gamma), \qquad (7.1.12)$$

where the $x^{(i)}$ are the components of x and the X_{ij} are the components of X_j. Since $\hat{h}^*(x|\gamma)$ satisfies the conditions (a), (b) and (c), as is not hard to verify, we shall call (7.1.12) the *modified* SMINK density estimator. Obviously we now have

$$\lim_{\gamma\downarrow 0}\hat{h}^*(x|\gamma) = 0 \quad \text{for all } \ x \in R^k.$$

Moreover, it is not hard to verify that under the conditions of theorem 7.1.2,

$$\text{plim}_{n\to\infty}\sup_{\gamma\in[\zeta_n,1]}\sup_x |\hat{h}(x|\gamma) - \hat{h}^*(x|\gamma)| = 0, \qquad (7.1.13)$$

hence:

Theorem 7.1.3. Theorem 7.1.2 carries over for the modified SMINK density estimator.

7.1.4 *How to choose the window width of the modified SMINK density estimator*

The problem is to choose γ in $[\zeta_n,1]$ such that the 'distance' between the modified SMINK estimate $\hat{h}^*(x|\gamma)$ and the true density $h(x)$ is minimal. Thus we have to choose γ_n in $[\zeta_n,1]$ such that

$$\rho(\hat{h}^*(x|\gamma_n),h(x)) = \inf_{\gamma\in[\zeta_n,1]}\rho(\hat{h}^*(x|\gamma),h(x)),$$

where ρ is some metric on the space of k-variate densities. The best metric would be $\rho(h_1,h_2) = \sup_x|h_1(x)-h_2(x)|$, since this metric corresponds to uniform convergence. But $h(x)$ is unknown so that this metric is difficult to apply, though not impossible (see Silverman (1978)). As a second best we therefore propose the integrated square error metric

$$\rho(h_1,h_2) = \int(h_1(x)-h_2(x))^2dx.$$

Thus we propose to minimize

$$\rho(\hat{h}^*(x|\gamma),h(x)) = \int(\hat{h}^*(x|\gamma)-h(x))^2dx$$

$$= \int\hat{h}^*(x|\gamma)^2dx - 2\int\hat{h}^*(x|\gamma)h(x)dx + h(x)^2dx.$$

Also now the terms involving $h(x)$ are unobservable. But:

Theorem 7.1.4. Under the conditions of theorem 7.1.2,

$$\text{plim}_{n\to\infty}\sup_{\gamma\in[\zeta_n,1]}|\int\hat{h}^*(x|\gamma)h(x)dx-(1/n)\Sigma_{j=1}^n\hat{h}^*(X_j|\gamma)| = 0.$$

Proof: Section 7.4.

Thus we have

$$\text{plim}_{n\to\infty}\sup_{\gamma\in[\zeta_n,1]}\Big|\int(\hat{h}^*(x|\gamma) - h(x))^2dx$$

$$- \{\int\hat{h}^*(x|\gamma)^2dx - 2(1/n)\Sigma_{j=1}^n\hat{h}^*(X_j|\gamma) + \int h(x)^2dx\}\Big| = 0. \tag{7.1.14}$$

Putting

$$q_n(\gamma) = \int\hat{h}^*(x|\gamma)^2 - 2(1/n)\Sigma_{j=1}^n\hat{h}^*(X_j|\gamma), \tag{7.1.15}$$

the result (7.1.14) now suggests to choose $\gamma_n \in [\zeta_n,1]$ such that

$$q_n(\gamma_n) = \inf_{\gamma\in[\zeta_n,1]}q_n(\gamma). \tag{7.1.16}$$

Note that

$$\int\hat{h}^*(x|\gamma)^2dx = \int\hat{h}(x|\gamma)^2dx$$

$$= (1/n^2)\Sigma_{j_1=1}^n\Sigma_{j_2=1}^n\exp[-(1/4)(X_{j_1}-X_{j_2})'\hat{V}^{-1}(X_{j_1}-X_{j_2})((1-\gamma^2)/\gamma^2)]$$

$$/ \ [(\surd(2\pi))^k\gamma^k(\surd 2)^k\surd\det(\hat{V})]$$

so that

$$\lim_{\gamma\downarrow 0}\int\hat{h}^*(x|\gamma)^2dx = \infty. \tag{7.1.17}$$

From (7.1.17) and (7.1.13) it thus follows

$$\lim_{\gamma\downarrow 0}q_n(\gamma) = \infty.$$

However, if we would employ in (7.1.15) the common SMINK density estimator (7.1.10) then the procedure (7.1.16) would break down, due to the fact that

$$\lim_{\gamma\downarrow 0}\int\hat{h}(x|\gamma)^2dx - 2\ (1/n)\Sigma_{j=1}^n\hat{h}(X_j|\gamma) = -\infty.$$

This is the main reason for working with the modified SMINK estimator instead of the common SMINK estimator (7.1.10).

This approach for choosing γ may serve as an alternative for the method proposed by Silverman (1978). For the usual kernel estimator of a density function Silverman has derived a method for establishing the window width (the γ_n in our case) corresponding to the optimal asymptotic rate of uniform convergence. In practice this optimal window width has to be observed from the fluctuations of the graph of the second derivative of the density estimate involved. But as Silverman admitted, this might be difficult in the multivariate case. Moreover, the resulting window width is only optimal from an asymptotic point of view, while our approach is closer related to finite sample fitting.

We will now pay attention to the question whether the sequence (γ_n) derived from (7.1.16) is such that

$$\mathrm{plim}_{n\to\infty}\sup_x|\hat{h}^*(x|\gamma_n)-h(x)|=0.$$

Let V be the probability limit of $\hat{V}$ and let μ be the probability limit of $\bar{X}$. Put

$$h(x|\gamma)$$

$$= \int[\exp(-½[x-\surd(1-\gamma^2)z-(1-\gamma^2))\mu]'V^{-1}[x-\surd(1-\gamma^2)z-(1-\surd(1-\gamma^2))\mu]/\gamma^2)$$

$$/((\surd 2\pi)^k\gamma^k\surd\det(V))]h(z)dz \quad (7.1.18)$$

Then

Lemma 7.1.1. Under the conditions of theorem 7.1.2,

(I) $\mathrm{plim}_{n\to\infty}\sup_{\gamma\in[\zeta_n,1]}\sup_x|\hat{h}^*(x|\gamma)-h(x|\gamma)| = 0$,

(II) $h(x|\gamma)$ is uniformly continuous on $R^k\times[0,1]$, and

(III) $\lim_{\gamma\downarrow 0}\sup_x|h(x|\gamma)-h(x)| = 0$.

Proof: Section 7.4.

From part (I) of this lemma we may conclude without further discussion:

Lemma 7.1.2. Under the conditions of theorem 7.1.2

$$\mathrm{plim}_{n\to\infty}\sup_{\gamma\in[\zeta_n,1]}|q_n(\gamma)-q(\gamma)| = 0,$$

where

$$q(\gamma) = \int h(x|\gamma)^2 dx - 2\int h(x|\gamma)h(x)dx. \qquad (7.1.19)$$

Obviously we see from (7.1.19) that $q(\gamma)$ takes a global minimum

$$-\int h(x)^2 dx$$

for $\gamma \in [0,1]$ such that $\int (h(x|\gamma) - h(x))^2 dx = 0$. Moreover, by the continuity of $h(x|\gamma)$ and $h(x)$ on $R^k \times [0,1]$ and R^k, respectively, it follows that the latter condition holds if and only if $h(x|\gamma) = h(x)$ for every $x \in R^k$. Thus we have

$$\{\gamma \in [0,1]: q(\gamma) = \inf_{\gamma_0 \in [0,1]} q(\gamma_0)\}$$

$$= \{\gamma \in [0,1]: \int (h(x|\gamma) - h(x))^2 dx\}$$

$$= \{\gamma \in [0,1]: h(x|\gamma) = h(x) \text{ for every } x \in R^k\} = \Gamma \qquad (7.1.20)$$

say. From part (III) of lemma 7.1.1 it follows that Γ contains at least 0. But Γ may contain more than one point. For example, if $h(x)$ is a normal density:

$$h(x) = \exp\{-\tfrac{1}{2}(x-\mu)'V^{-1}(x-\mu)\} \;/\; [(\surd(2\pi))^k \surd\det(V)],$$

then $h(x|1) = h(x)$ and consequently also $1 \in \Gamma$.

Next we observe from the lemmas 7.1.1 and 7.1.2 and from (7.1.16) and (7.1.19) that

$$\text{plim}_{n\to\infty}(q_n(\gamma_n) - q(\gamma_n)) = 0,$$

$$\text{plim}_{n\to\infty}(q_n(\zeta_n) - q(\zeta_n) = 0,$$

$$\lim_{n\to\infty} q(\zeta_n) = -\int h(x)^2 dx,$$

$$q(\gamma) \geq -\int h(x)^2 dx \quad \text{for} \quad \gamma \in [0,1]$$

and

$$q_n(\gamma_n) \geq q_n(\zeta_n) \quad \text{a.s.}$$

Using these results it is not hard to prove:

Lemma 7.1.3. Under the conditions of theorem 7.1.3

$$\text{plim}_{n\to\infty}\int(h(x|\gamma_n) - h(x))^2 dx = 0,$$

where (γ_n) is defined by (7.1.16).

We show now that this lemma implies

$$\text{plim}_{n\to\infty}\sup_x |h(x|\gamma_n) - h(x)| = 0. \tag{7.1.21}$$

First, consider the case that γ_n is non-random. We then have to show that

$$\lim_{n\to\infty}\int(h(x|\gamma_n)-h(x))^2 dx=0 \text{ implies } \lim_{n\to\infty}\sup_x |h(x|\gamma_n)-h(x)|=0.$$

The sequence (γ_n) is a sequence of numbers in a compact interval $[0,1]$ and it therefore has at least one limit point γ_*, say, in $[0,1]$. This limit point must be a point of Γ, as otherwise

$$\lim_{n\to\infty}\int((h(x|\gamma_n)-h(x)))^2 dx \neq 0,$$

and the same applies to other limit points of (γ_n). Thus all the limit points of (γ_n) are contained in Γ. This implies that every subsequence of (γ_n) contains a further subsequence (γ_{n_j}) such that for some point $\gamma_*\in\Gamma$,

$$\lim_{j\to\infty}|\gamma_{n_j} - \gamma_*| = 0$$

and consequently by the uniform continuity of $h(x|\gamma)$ on $R^k\times[0,1]$, and by (7.1.20),

$$\lim_{j\to\infty}\sup_x |h(x|\gamma_{n_j})-h(x)| = \lim_{j\to\infty}\sup_x |h(x|\gamma_{n_j})-h(x|\gamma_*)| = 0.$$

This result implies that $\lim_{n\to\infty}\sup_x |h(x|\gamma_n)-h(x)| = 0$. The proof for random γ_n follows from the above argument and theorem 2.1.6.

Combining (7.1.21) with lemma 7.1.1 we now conclude:

Theorem 7.1.5. Let the conditions of theorem 7.1.3 be satisfied. The sequence (γ_n) obtained from (7.1.16) is such that

$$\text{plim}_{n\to\infty}\sup_x|\hat{h}^*(x|\gamma_n)-h(x)| = 0.$$

Exercises:

1. Verify that marginal densities of SMINK density estimators are SMINK density estimators themselves.
2. Verify the Dirac-catastrophe.
3. Prove Lemma 7.1.3.

7.2 *Regression*

7.2.1 *SMINK estimators of a regression function*

Consider an i.i.d. stochastic process $(Y_1,X_1),\ldots,(Y_n,X_n),\ldots$ in $R\times R^k$, where the distribution of the (Y_j,X_j)'s is absolutely continuous with continuous density f(y,x), say. Of course, then the distribution of the X_j's is continuous too with continuous density

$$h(x) = \int f(y,x)dy,$$

say. As is well know, the conditional expectation of Y_j relative to the event $X_j=x$ $(x\in R^k)$ is

$$E(Y_j|X_j=x) = \int yf(y,x)dy/h(x) = g(x)$$

say, provided h(x) > 0. This suggests to estimate the regression function g(x) by an estimator of the type

$$\hat{g}(x) = \int y\hat{f}(y,x)dy/\hat{h}(x)$$

where

$$\hat{f}(y,x) \text{ and } \hat{h}(x)$$

are estimators of the densities f(y,x) and h(x), respectively. In the case where these density estimators are (modified) SMINK estimators we get

$$\hat{g}(x|\gamma_1,\gamma_2) = \int y\hat{f}(y,x|\gamma_1)dy/\hat{h}(x|\gamma_2),$$

where γ_1 and γ_2 are window width parameters, as an estimator of $g(x)$. This estimator will be called a (modified) SMINK estimator of a regression function.

Now let us focus on the SMINK estimator

$$\hat{f}(y,x|\gamma)$$

of $f(y,x)$. Similarly to (7.1.9) and (7.1.10) we have

$$\hat{f}(y,x|\gamma) = (1/n)\Sigma_{j=1}^{n}\hat{f}_j(y,x|\gamma) \tag{7.2.1}$$

where

$$\hat{f}_j(y,x|\gamma) = \exp(-\tfrac{1}{2}[(y,x')-\surd(1-\gamma^2)(Y_j,X_j')-(1-\surd(1-\gamma^2))(\overline{Y},\overline{X}')]\hat{V}^{-1}$$
$$\times [(y,x')'-\surd(1-\gamma^2)(Y_j,X_j')'-(1-\surd(1-\gamma^2))(\overline{Y},\overline{X}')]/\gamma^2\}$$
$$/\{\surd(2\pi))^{k+1}\gamma^{k+1}\surd\det(\hat{V}_{y,x})\} \tag{7.2.2}$$

with

$\overline{Y}$ the sample mean of the Y_j's,

$\overline{X}$ the sample mean of the X_j's and

$\hat{V}_{y,x}$ the sample variance matrix of the vectors (Y_j,X_j).

Using formula (2) at page 28 of Anderson (1958) it is not too hard to verify [cf. exercise 1] that

$$\int y\hat{f}_j(y,x|\gamma)dy = \{\surd(1-\gamma^2)Y_j + (1-\surd(1-\gamma^2))\overline{Y}+\hat{\beta}'(x-\surd(1-\gamma^2))X_j$$
$$+ (1-\surd(1-\gamma^2))\overline{X})\}\cdot\hat{h}_j(x|\gamma), \tag{7.2.3}$$

where $\hat{\beta}$ is the usual OLS estimator of the regression coefficients of a linear model:

$$\hat{\beta} = [(1/n)\Sigma_{j=1}^{n}(X_j - \bar{X})(X_j - \bar{X})']^{-1}(1/n)\Sigma_{j=1}^{n}(X_j - \bar{X})(Y_j - \bar{Y}),$$

and $\hat{h}_j(x|\gamma)$ is defined by (7.1.10). Combining (7.2.1) and (7.2.3) we obtain

$$\int y\hat{f}(y,x|\gamma)dy$$
$$= \sqrt{(1-\gamma^2)}(1/n)\Sigma_{j=1}^{n}(Y_j - \hat{\beta}'X_j - \hat{\alpha})\hat{h}_j(x|\gamma) + (\hat{\beta}'x+\hat{\alpha})\hat{h}(x|\gamma), \tag{7.2.4}$$

where $\hat{\alpha}$ is the OLS estimator of the constant term of a linear regression model:

$$\hat{\alpha} = \bar{Y} - \hat{\beta}'\bar{X}.$$

Dividing (7.2.4) (with γ replaced by γ_1) by $\hat{h}(x|\gamma_2)$ yields the following explicit expression of the SMINK estimator of a regression function:

$$\hat{g}(x|\gamma_1,\gamma_2) = \{(\hat{\beta}'x+\hat{\alpha})\hat{h}(x|\gamma_1)$$
$$+\sqrt{(1-\gamma_1^2)}(1/n)\Sigma_{j=1}^{n}(Y_j - \hat{\beta}'X_j - \hat{\alpha})\hat{h}_j(x|\gamma_1)\}/\hat{h}(x|\gamma_2). \tag{7.2.5}$$

If $\hat{f}^*(y,x|\gamma)$ and $\hat{h}^*(x|\gamma)$ are modified SMINK regression estimators then

$$\hat{g}^*(x|\gamma_1,\gamma_2) = \int y\hat{f}^*(y,x|\gamma_1)dy/\hat{h}^*(x|\gamma_2) = [(\hat{\beta}'x+\hat{\alpha})\hat{h}^*(x|\gamma_1)$$

$$+ \sqrt{(1-\gamma_1^2)}(1/n)\Sigma_{j=1}^{n}(Y_j - \hat{\beta}'X_j - \hat{\alpha})\{\Pi_{i=1}^{k}I(x^{(i)} \neq X_{ij})\}\hat{h}_j(x|\gamma_1)]$$
$$/\hat{h}^*(x|\gamma_2) \tag{7.2.6}$$

which will be called the *modified* SMINK estimator of a regression function. Note that these regression function estimators contain the OLS estimator of a linear regression model as a special case, for

$$\hat{g}^*(x|1,1) = \hat{g}(x|1,1) = \hat{\beta}'x+\hat{\alpha}.$$

This corresponds to the fact that the SMINK estimator of a density equals the maximum likelihood estimator of a normal density if γ=1.

7.2.2 *How to choose the window width parameters*

In section 7.1.4 we have proposed an integrated square error approach for choosing the window width parameter γ. The argument involved also applies to the problem of choosing γ_2, for the denumerator of (7.2.5) is a SMINK density estimator which should be fitted as close as possible to h(x). Therefore we propose to choose γ_2 by the procedure described in section 7.1.4 and the resulting optimal window width will be denoted by $\gamma_n^{(2)}$.

For choosing γ_1 we propose a weighted least squares procedure, motivated by the following theorem.

Theorem 7.2.1. Let $(Y_1,X_1),\ldots,(Y_n,X_n),\ldots$ be an i.i.d. stochastic process in $R\times R^k$. Assume in addition to the conditions of theorem 7.1.2 that the distribution of the (Y_j,X_j) is absolutely continuous, tnat g(x)h(x) is continuous and absolutely integrable, where

$$g(x) = E(Y_j|X_j=x)$$

and h(x) is the density of the X_j's, that the sample covariance of the Y_j's and the X_j's converges in probability and that $EY_j^2<\infty$. Let the sequence (ζ_n) be defined as in theorem 7.1.2. Put

$$Q_n(\gamma) = (1/n)\Sigma_{j=1}^n(Y_j-\hat{g}(X_j|\gamma,\gamma_n^{(2)})^2\hat{h}(X_j|\gamma_n^{(2)})^2.$$

If we choose $\gamma_n^{(1)}$ in $[\zeta_n,1]$ such that

$$Q_n(\gamma_n^{(1)}) = \inf_{\gamma\in[\zeta_n,1]}Q_n(\gamma)$$

then

$$\text{plim}_{n\to\infty}\sup_{x\in\{x\in R^k:h(x)\geq\varepsilon\}}|\hat{g}(x|\gamma_n^{(1)},\gamma_n^{(2)})-g(x)| = 0$$

for every $\varepsilon\in(0,\sup_x h(x)]$. Moreover, the same result holds for the modified SMINK regression estimator.

Proof: Section 7.4.

Exercise:
1. Prove (7.2.3) for the bivariate case.

7.3 *A numerical example*

We now demonstrate the performance of the SMINK estimators on an artificial data set of size 100. In the regression case to be considered we employed the model.

$$Y_j = X_{1,j}^2 + X_{2,j}^2 + U_j,\ j=1,2,\ldots,100,$$

so that the regression function to be estimated is:

$$g(x_1,x_2) = x_1^2 + x_2^2. \tag{7.3.1}$$

We have drawn the $X_{1,j}$ and $X_{2,j}$ independently from a mixture of the normal distributions N(-2,1) and N(2,1) with equal weights, and the U_j have been drawn independently from the normal distribution N(0,1). So the density of the distribution of the pairs $(X_{1,j},X_{2,j})$ is:

$$h(x_1,x_2) = \{\tfrac{1}{2}\exp[-\tfrac{1}{2}(x_1+2)^2]+\tfrac{1}{2}\exp[-\tfrac{1}{2}(x_1-2)^2]/\surd(2\pi)\}$$

$$\times\ \{\tfrac{1}{2}\exp[-\tfrac{1}{2}(x_2+2)^2]+\tfrac{1}{2}\exp[-\tfrac{1}{2}(x_2-2)^2]/\surd(2\pi)\} \tag{7.3.2}$$

Obviously this density has four modes, namely at

$$(-2,-2),\ (-2,2),\ (2,-2) \text{ and } (2,2).$$

Applying the integrated square error method (section 7.1.4) and the weighted least squares method (section 7.2.2) we found the following window width parameters:

$$\gamma_n^{(1)} = .31,\ \gamma_n^{(2)} = .33.$$

In the figures 1 and 2 we show the true density (7.3.2) and its SMINK estimate, respectively, and similarly in the figures 3 and 4 we show the true regression function (7.3.1) and its

SMINK estimate, where in all cases x_1 and x_2 go from -3.5 to 3.5.

Comparing the figures 1 and 2 we see that the SMINK density estimator is flatter than the true density. Nevertheless the density estimate is smooth and clearly has four modes at the right places. The SMINK regression estimate in figure 4 is also flatter than the true regression function in figure 3, especially at the bottom and at the edges. These spots, however, have relatively low values of the density h, as one can observe from figure 1. Since the convergence of the SMINK regression estimators is only uniform on sets with sufficient probability mass, one may expect relative poor fit at x with low value of h(x).

Figure 1: True Density

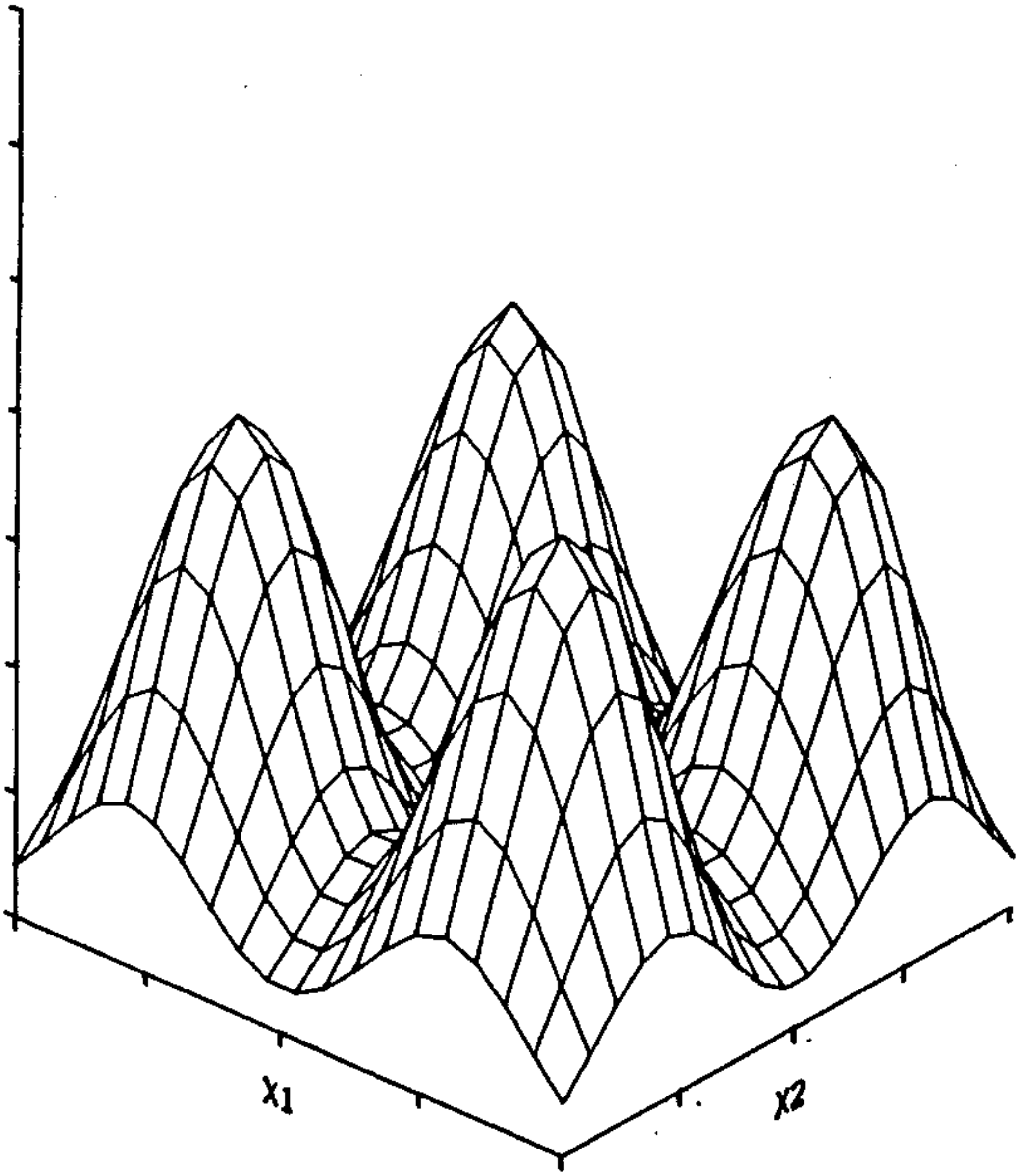


Figure 2: SMINK Density Estimator

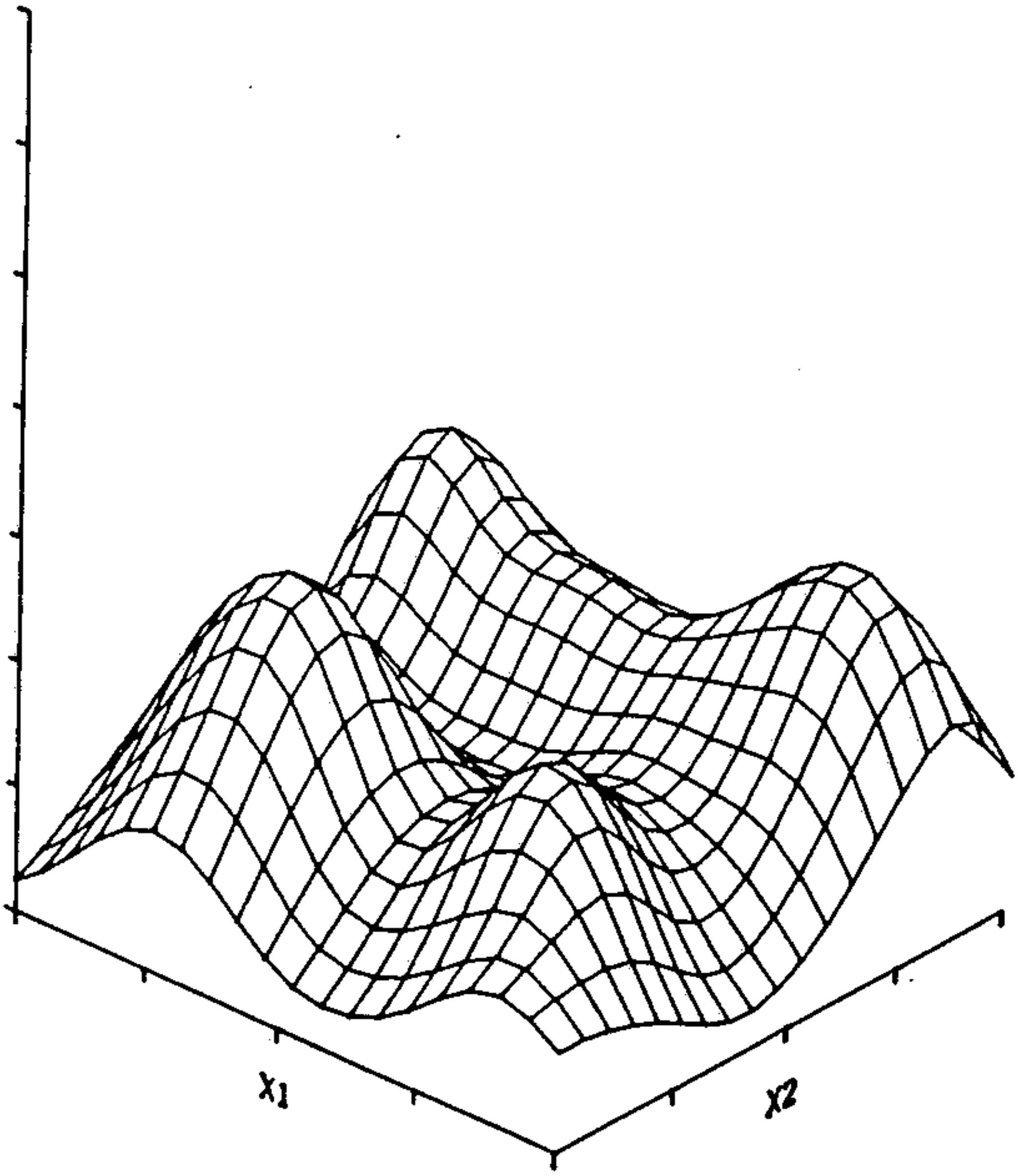


Figure 3: True Regression Function

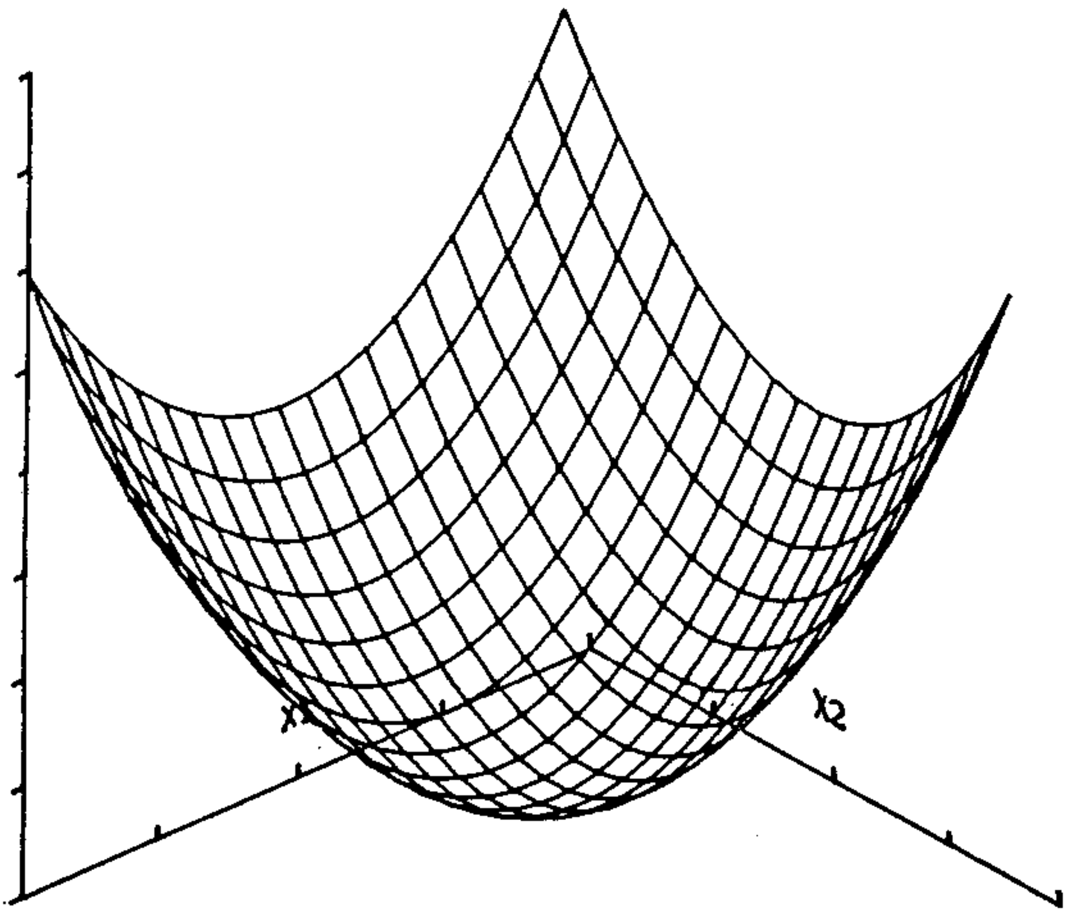
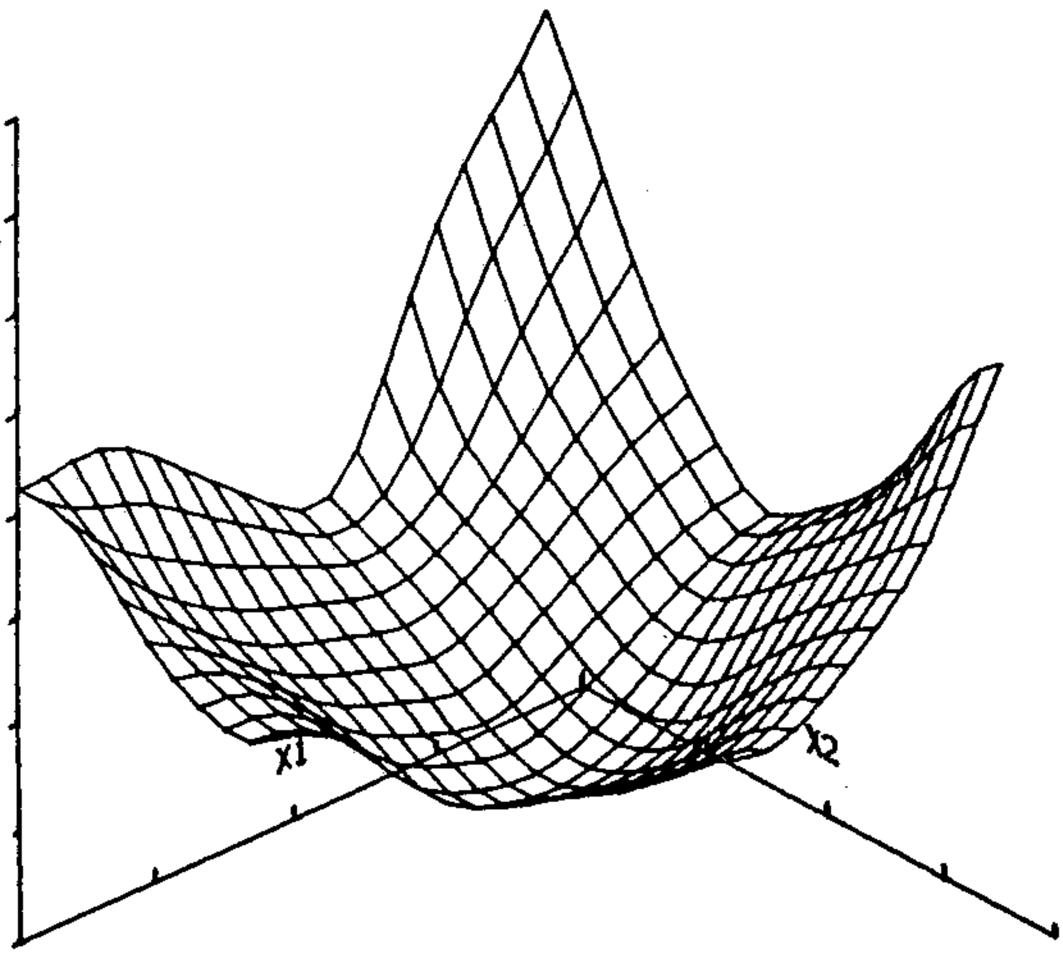


Figure 4: SMINK Regression Function Estimator



7.4 *Proofs*

For proving the previous asymptotic results we need the following lemmas.

Lemma 7.4.1. Let $(Y_1, X_1), \ldots, (Y_n, X_n), \ldots$ be an i.i.d. stochastic process in $R \times R^k$. Assume that the distribution of the X_j is absolutely continuous with continuous density $h(x)$, say, and that $g(x)h(x)$ is continuous and absolutely integrable on R^k, where $g(x)=E(Y_j|X_j=x)$. Furthermore we assume

$$E\ Y_j^2 < \infty \tag{7.4.1}$$

$$E\ X_j = \mu \text{ and } E\ X_j X_j' = W \text{ exist} \tag{7.4.2}$$

$$\text{plim}_{n\to\infty}\overline{X} = \mu \quad \text{and} \quad \text{plim}_{n\to\infty}\hat{V} = V \ (=W-\mu\mu') \text{ exist,} \tag{7.4.3}$$

where $\overline{X}$ is the sample mean, $\hat{V}$ is the sample variance of the X_j's and

$$V \text{ is positive definite.} \tag{7.4.4}$$

Denote

$$a_n(x|\gamma) = (1/n)\Sigma_{j=1}^n Y_j \hat{h}_j(x|\gamma), \tag{7.4.5}$$

where $\hat{h}_j(x|\gamma)$ is defined by (7.1.10) and let

$$a(x|\gamma) = \int g(z)\exp(-\tfrac{1}{2}[x-\surd(1-\gamma^2)z-(1-\surd(1-\gamma^2))\mu]'V^{-1}$$

$$\times [x-\surd(1-\gamma^2)z-(1-\surd(1-\gamma^2))\mu]/\gamma^2\}h(z)dz/\{(\surd(2\pi))^k\gamma^k\surd\det(V)\} \tag{7.4.6}$$

Let (ζ_n) be a sequence of numbers in (0,1) such that

$$\lim_{n\to\infty}\zeta_n^k\surd n = \infty \quad \text{and} \quad \lim_{n\to\infty}\zeta_n = 0. \tag{7.4.7}$$

Then

$$\text{plim}_{n\to\infty}\sup_{\gamma\in[\zeta_n,1]}\sup_x|a_n(x|\gamma) - a(x|\gamma)| = 0. \tag{7.4.8}$$

Moreover, $a(x|\gamma)$ is a uniformly bounded and uniformly continuous function on $R \times [0,1]$ satisfying

$$\lim_{\gamma\downarrow 0}\sup_x |a(x|\gamma) - g(x)h(x)| = 0. \qquad (7.4.9)$$

Proof. From the inversion formula for characteristic functions (cf. theorem 1.5.1) it follows that $h_j(x|\gamma)$ can be written as

$$\hat{h}_j(x|\gamma)=(1/(2\pi))^k\int\exp\{-i\cdot t'x+i\cdot t'X_j\surd(1-\gamma^2)+i\cdot t'\bar{X}(1-\surd(1-\gamma^2))\}$$
$$\times \exp(-\tfrac{1}{2}\gamma^2 t'\hat{V}t)dt, \qquad (7.4.10)$$

hence

$$a_n(x|\gamma)=(1/2\pi)^k\int\{(1/n)\Sigma_{j=1}^n y_j\exp[i\cdot t'X_j\surd(1-\gamma^2)]\}$$
$$\times \exp[-i\cdot t'x+i\cdot t'\bar{X}(1-\surd(1-\gamma^2))]\exp(-\tfrac{1}{2}\gamma^2 t'\hat{V}t)dt \qquad (7.4.11)$$

Put

$$a_n^*(x|\gamma)=(1/(2\pi))^k\int\{(1/n)\Sigma_{j=1}^n E\ Y_j\exp[i\cdot t'X_j\surd(1-\gamma^2)]\}$$
$$\times \exp[-i\cdot t'x+i\cdot t'\bar{X}(1-\surd(1-\gamma^2))]\exp(-\tfrac{1}{2}\gamma^2 t'\hat{V}t)dt \qquad (7.4.12)$$

We shall prove (7.4.8) now in four steps. Assuming that (ζ_n) is monotonic and chosing $\varepsilon\in(0,1-\zeta_n)$ for all n it will be shown that

$$\text{plim}_{n\to\infty}\sup_{\gamma\in[\zeta_n,1-\varepsilon]}\sup_x|a_n(x|\gamma) - a_n^*(x|\gamma)| = 0 \qquad (7.4.13)$$

$$\lim_{\varepsilon\downarrow 0}\lim_{n\to\infty}E\ \sup_{\gamma\in[1-\varepsilon,1]}\sup_x|a_n(x|\gamma) - a_n^*(x|\gamma)| = 0 \qquad (7.4.14)$$

$$\text{plim}_{n\to\infty}\sup_{\gamma\in[0,1-\varepsilon]}\sup_x|a_n^*(x|\gamma) - a(x|\gamma)| = 0 \qquad (7.4.15)$$

and finally

$$\text{plim}_{n\to\infty}\sup_{\gamma\in[1-\varepsilon,1]}\sup_x|a_n^*(x|\gamma) - a(x|\gamma)| = 0. \qquad (7.4.16)$$

Obviously (7.4.13) through (7.4.16) imply (7.4.8).

Step 1: Proof of (7.4.13).
From (7.4.11) and (7.4.12) we have

$$\sup_x |a_n(x|\gamma) - a_n^*(x|\gamma)$$

$$\leq (1/(2\pi))^k \int |(1/n)\Sigma_{j=1}^n (Y_j \exp[i\cdot t'X_j\surd(1-\gamma^2)]$$

$$- E\, Y_j \exp[i\cdot t'X_j\surd(1-\gamma^2)])| \exp(-\tfrac{1}{2}\gamma^2 t'\hat{V}t)dt$$

$$= (1/(2\pi))^k [1/\surd(1-\gamma^2)]^k \int |(1/n)\Sigma_{j=1}^n (Y_j \exp(i\cdot t'X_j)$$

$$-E\, Y_j \exp(i\cdot t'X_j))|$$

$$\times \exp[\tfrac{1}{2}\{\gamma^2/(1-\gamma^2)\}t'\hat{V}]dt = b_n(\gamma) \qquad (7.4.17)$$

say. Now let S be the set of all positive definite (k×k) matrices with minimum eigenvalue larger than $\sigma_0 > 0$, where σ_0 is such that $V \in S$. If $V \in S$ then $b_n(\gamma)$ is bounded from above by

$$\bar{b}_n(\gamma) = (1/(2\pi))^k [1/\surd(1-\gamma^2)]^k \int |(1/n)\Sigma_{j=1}^n (Y_j \exp(i\cdot t'X_j)$$

$$-E\, Y_j \exp(i\cdot t'X_j))|$$

$$\times \exp[-\tfrac{1}{2}\{\gamma^2/(1-\gamma^2)\}\sigma_0 t't]dt \qquad (7.4.18)$$

because then $t'\hat{V}t \geq \sigma_0 t't$. But

$$\sup_t E|(1/n)\Sigma_{j=1}^n (Y_j \exp(i\cdot t'X_j) - E\, Y_j \exp(i\cdot t'X_j))|$$

$$\leq \sup_t \{E[(1/n)\Sigma_{j=1}^n (Y_j \cos(t'X_j) - E\, Y_j \cos(t'X_j))]^2$$

$$+ E[\Sigma_{j=1}^n (Y_j \sin(t'X_j) - E\, Y_j \sin(t'X_j))]^2\}^{\frac{1}{2}}$$

$$= O(1/\surd n) \to 0, \qquad (7.4.19)$$

hence

$$E \sup_{\gamma\in[\zeta_n, 1-\varepsilon]} \bar{b}_n(\gamma)$$

$$\leq (1/(2\pi))^k [1/\surd(1-(1-\varepsilon)^2)]^k \int E|(1/n)\Sigma_{j=1}^n (Y_j \exp(i\cdot t'X_j)$$

$$- E\ Y_j \exp(i \cdot t'X_j))|\exp(-\tfrac{1}{2}\zeta_n{}^2\sigma_0 t't)dt$$

$$= O[(\zeta_n^{-k}\surd n)^{-1}] \to 0 \qquad (7.4.20)$$

where the last conclusion follows from condition (7.4.7). Consequently by Chebishev's inequality,

$$\text{plim}_{n\to\infty}\sup_{\gamma\in[\zeta_n,1-\varepsilon]}\bar{b}_n(\gamma) = 0 \qquad (7.4.21)$$

Since (7.4.3) and (7.4.4) imply $\lim_{n\to\infty}P(\hat{V} \in S) = 1$ and since

$$b_n(\gamma) \le \bar{b}_n(\gamma) \text{ if } V \in S,$$

(7.4.21) implies (7.4.13).

Step 2: Proof of (7.4.14).
Observe from (7.4.18) that

$$\bar{b}(\gamma) = (1/(2\pi))^k\int|(1/n)\Sigma_{j=1}^n\{(Y_j\exp[i\cdot t'X_j\surd(1-\gamma^2)]$$

$$- E\ Y_j\exp[i\cdot t'X_j\surd(1-\gamma^2)]\}|\exp(-\tfrac{1}{2}\sigma_0\gamma^2 t't)dt$$

$$\le (1/(2\pi))^k\int|(1/n)\Sigma_{j=1}^n\{Y_j(\exp[i\cdot t'X_j\surd(1-\gamma^2)]-1)$$

$$- E\ Y_j(\exp[i\cdot t'X_j\surd(1-\gamma^2)]-1)\}|\exp(-\tfrac{1}{2}\sigma_0\gamma^2 t't)dt$$

$$+ (1/(2\pi))^k|(1/n)\Sigma_{j=1}^n(Y_j - E\ Y_j)|\int\exp(-\tfrac{1}{2}\gamma^2\sigma_0 t't)dt$$

$$\le 2(1/(2\pi))^k(1/n)\Sigma_{j=1}^n|Y_j||X_j|\surd(1-\gamma^2)\int|t|\exp(-\tfrac{1}{2}\gamma^2\sigma_0 t't)dt$$

$$+ (1/2\pi)^k|(1/n)\Sigma_{j=1}^n(Y_j - E\ Y_j)|\int\exp(-\tfrac{1}{2}\gamma^2\sigma_0 t't)dt, \qquad (7.4.22)$$

where the last inequality follows from the mean value theorem. Thus

$$E\ \sup_{\gamma\in[1-\varepsilon,1]}\bar{b}_n(\gamma) \le$$

$$\le (1/(2\pi))^k 2E|Y_j||X_j|\surd(1-(1-\varepsilon)^2)\int|t|\exp[-\tfrac{1}{2}\sigma_0(1-\varepsilon)^2 t't]dt$$

$$+ (1/(2\pi))^k E|(1/n)\Sigma_{j=1}^n (Y_j - E\ Y_j)| \int \exp[-\tfrac{1}{2}\sigma_0 (1-\varepsilon)^2 t't]dt. \tag{7.4.23}$$

Since $E|Y_j||X_j| < \infty$ the first term of the upperbound above converges to zero if $\varepsilon \downarrow 0$. Moreover, by Liapounov's inequality,

$$E|(1/n)\Sigma_{j=1}^n (Y_j - E\ Y_j)| \le \{E(1/n)\Sigma_{j=1}^n (Y_j - E\ Y_j))^2\}^{\frac{1}{2}} \to 0 \tag{7.4.24}$$

hence the second term converges to zero for $n\to\infty$. Thus:

$$\lim_{\varepsilon\downarrow 0}\lim_{n\to\infty} E\ \sup_{\gamma\in[1-\varepsilon,1]}\tilde{b}_n(\gamma) = 0. \tag{7.4.25}$$

Using this result and the trivial equality

$$b_n(\gamma) = b_n(\gamma)I(\hat{V}\in S) + b_n(\gamma)I(\hat{V}\notin S),$$

where $I(.)$ is the well-known indicator function, (7.4.14) can easily be proved.

Step 3: Proof of (7.4.15).
From (7.4.12) and the inversion formula for characteristic functions it follows

$$\begin{aligned} a_n^*(x|\gamma) &= \int \{\exp\{-\tfrac{1}{2}[x-\surd(1-\gamma^2)z-(1-\surd(1-\gamma^2))X]\hat{V}^{-1} \\ &\qquad\times [x-\surd(1-\gamma^2)z-(1-\surd(1-\gamma^2))X]/\gamma^2\} \\ &\qquad /[(\surd(2\pi))^k\gamma^k\surd\det(\hat{V})]\}g(z)h(z)dz. \end{aligned} \tag{7.4.26}$$

Now put for convenience

$$f(x) = g(x)h(x). \tag{7.4.27}$$

Then $a_n^*(x|\gamma)$ can be written as

$$\begin{aligned} &a_n^*(x|\gamma) \\ &= (\surd(2\pi))^{-k}(\surd(1-\gamma^2))^{-k}\int f([x-(1-\surd(1-\gamma^2))\overline{X} - \gamma\hat{V}^{\frac{1}{2}}w]/\surd(1-\gamma^2)) \\ &\qquad \times \exp(-\tfrac{1}{2}w'w)dw \end{aligned} \tag{7.4.28}$$

and similarly we can write

$$a(x|\gamma)$$

$$= (\surd(2\pi))^{-k}(\surd(1-\gamma^2))^{-k}\int f([x-(1-\surd(1-\gamma^2))\mu-\gamma V^{\frac{1}{2}}]/\surd(1-\gamma^2))$$

$$\times\ w\cdot\exp(-\tfrac{1}{2}w'w)dw. \qquad (7.4.29)$$

Since $f(x)=g(x)h(x)$ is continuous and absolutely integrable on R^k it is uniformly continuous on R^k, hence

$$\sup_{\gamma\in[0,1-\varepsilon]}\sup_x|f([x-(1-\surd(1-\gamma^2))\overline{X}-\gamma\hat{V}^{\frac{1}{2}}w]/\surd(1-\gamma^2))$$

$$-\ f([x-(1-\surd(1-\gamma^2))\mu-\gamma V^{\frac{1}{2}}w]/\surd(1-\gamma^2))| \to 0 \qquad (7.4.30)$$

in prob. pointwise in w if

$$\sup_{\gamma\in[0,1-\varepsilon]}[(1-\surd(1-\gamma^2))/\surd(1-\gamma^2)]|\overline{X}-\mu|$$

$$+\ \sup_{\gamma\in[0,1-\varepsilon]}(\gamma/\surd(1-\gamma^2))|\hat{V}^{\frac{1}{2}}w-V^{\frac{1}{2}}w| \to 0 \qquad (7.4.31)$$

in prob. pointwise in w. However, condition (7.4.31) follows from condition (7.4.3). Therefore (7.4.15) follows from (7.4.28), (7.4.29) and (7.4.30) by bounded convergence.

Step 4: Proof of (7.4.16).
From the inversion formula for characteristic functions we have

$$a(x|\gamma) = \{1/(2\pi)\}^k\int\{E\ Y_j\exp[i\cdot t'X_j\surd(1-\gamma^2)]\}$$

$$\times\ \exp[-i\cdot t'+i\cdot t'\mu(1-\surd(1-\gamma^2))]\exp(-\tfrac{1}{2}\gamma^2t'Vt)dt, \quad (7.4.32)$$

hence from (7.4.12) and (7.4.13) it follows

$$\sup_x|a_n^*(x|\gamma)\ -\ a(x|\gamma)|$$

$$\leq \{1/(2\pi)\}^k(E|Y_j|)\int|\exp[i\cdot t'\overline{X}(1-\surd(1-\gamma^2))]\cdot\exp(-\tfrac{1}{2}\gamma^2t'\hat{V}t)$$

$$-\ \exp[i\cdot t'\mu(1-\surd(1-\gamma^2))]\cdot\exp(-\tfrac{1}{2}\gamma^2t'Vt)|dt$$

$$\leq \{1/(2\pi)\}^k (E|Y_j|)\int|\exp(-\tfrac{1}{2}\gamma^2 t'\hat{V}t) - \exp(-\tfrac{1}{2}\gamma^2 t'Vt)|dt$$

$$+ \{1/(2\pi)\}^k (E|Y_j|)\int|\exp[i\cdot t'\overline{X}(1-\surd(1-\gamma^2))]$$

$$- \exp[-i\cdot t'\mu(1-\surd(1-\gamma^2))]|\exp(-\tfrac{1}{2}\gamma^2 t'Vt)dt$$

$$\leq \{1/(2\pi)\}^k (E|Y_j|)\int|\exp[-(1/4)\gamma^2 t'\hat{V}t] - \exp[-(1/4)\gamma^2 t'Vt]|$$

$$\times |\exp[-(1/4)\gamma^2 t'\hat{V}t] + \exp[-(1/4)\gamma^2 t'Vt]|dt$$

$$+ \{1/(2\pi)\}^k (E|Y_j|)(1-\surd(1-\gamma^2))|\overline{X}-\mu|\int|t|\exp(-\tfrac{1}{2}\gamma^2 t'Vt)dt. \tag{7.4.33}$$

For $\gamma\in[1-\varepsilon,1]$ and $\hat{V}\in S$ the first term at the right side of (7.4.33) is bounded by

$$2\{1/(2\pi)\}^k (E|Y_j|)\int \sup_{\gamma\in[1-\varepsilon,1]}|\exp[-(1/4)\gamma^2 t'\hat{V}t]$$

$$- \exp[-(1/4)\gamma^2 t'Vt]|\exp[-\tfrac{1}{2}\sigma_0(1-\varepsilon)^2 t't]dt, \tag{7.4.34}$$

which converges to zero in probability by condition (7.4.3) and bounded convergence. The convergence in prob. to zero uniformly in $\gamma\in[1-\varepsilon,1]$ of the second term follows straightforwardly from condition (7.4.3). Similarly to step 1, this completes the proof of (7.4.16) and hence of (7.4.8).

We now complete the proof of the lemma under review. First, the uniform boundedness of $a(x|\gamma)$ on $R^k\times[0,1]$ follows straightforwardly from (7.4.29) and (7.4.32). Next we show that $a(x|\gamma)$ is uniformly continuous on $R^k\times[0,1]$. Observe from (7.4.29) and the uniform continuity of $f(x) = g(x)h(x)$ that

$$\lim_{\alpha\downarrow 0}\sup_{x_1\in R^k, x_2\in R^k, |x_1-x_2|\leq\alpha}\sup_{\gamma\in[0,1-\varepsilon]}|a(x_1|\gamma)-a(x_2|\gamma)| = 0 \tag{7.4.35}$$

and from (7.4.33) we see that

$$\lim_{\alpha\downarrow 0}\sup_{x_1\in R^k, x_2\in R^k, |x_1-x_2|\leq\alpha}\sup_{\gamma\in[1-\varepsilon,1]}|a(x_1|\gamma)-a(x_2|\gamma)|=0 \tag{7.4.36}$$

hence

$$\lim_{\alpha\downarrow 0}\sup_{x_1\in R^k, x_2\in R^k, |x_1-x_2|\le\alpha}\sup_{\gamma\in[0,1]}|a(x_1|\gamma)-a(x_2|\gamma)|=0. \tag{7.4.37}$$

Moreover, from (7.4.33) we see that for $\gamma_1,\gamma_2\in[\varepsilon,1]$

$$\begin{aligned}
&\sup_x|a(x|\gamma_1)-a(x|\gamma_2)|\\
&\le (1/(2\pi))^k E|Y_j|\int|\exp[i\cdot t'\mu(1-\surd(1-\gamma_1^2))]\exp(-\tfrac{1}{2}\gamma_1^2 t'Vt)\\
&\qquad - \exp[i\cdot t'\mu(1-\surd(1-\gamma_2^2))]\exp(-\tfrac{1}{2}\gamma_2^2 t'Vt)|dt\\
&\le (1/(2\pi))^k E|Y_j|\int\exp[i\cdot t'\mu(1-\surd(1-\gamma_1^2))]\\
&\qquad - \exp[i\cdot t'\mu(1-\surd(1-\gamma_2^2))]|\exp(-\tfrac{1}{2}\varepsilon^2 t'Vt)dt\\
&\quad + (1/(2\pi))^k E|Y_j|\int|\exp(-\tfrac{1}{2}\gamma_1^2 t'Vt) - \exp(-\tfrac{1}{2}\gamma_2^2 t'Vt)|dt\\
&\le (1/(2\pi))^k E|Y_j||\surd(1-\gamma_1^2)-\surd(1-\gamma_2^2)||\mu|\int|t|\exp(-\tfrac{1}{2}\varepsilon^2 t'Vt)dt\\
&\quad + (1/(2\pi))^k E|Y_j|\tfrac{1}{2}|\gamma_1^2-\gamma_2^2|\int t'Vt.\exp[-(1/4)\varepsilon^2 t'Vt]dt,
\end{aligned} \tag{7.4.38}$$

hence

$$\lim_{\alpha\downarrow 0}\sup_{\gamma_1\in[\varepsilon,1],\gamma_2\in[\varepsilon,1],|\gamma_1-\gamma_2|\le\alpha}\sup_x|a(x|\gamma_1)-a(x|\gamma_2)|=0. \tag{7.4.39}$$

If (7.4.9) is true, then it follows from (7.4.9) and (7.4.39) that

$$\lim_{\alpha\downarrow 0}\sup_{\gamma_1\in[0,1],\gamma_2\in[0,1],|\gamma_1-\gamma_2|\le\alpha}\sup_x|a(x|\gamma_1)-a(x|\gamma_2)|=0. \tag{7.4.40}$$

Combining (7.4.37) and (7.4.40) we then see that $a(x|\gamma)$ is uniformly continuous on $R^k\times[0,1]$. So it remains to prove (7.4.9). Observe from (7.4.29) and (7.4.27) that for every $M>0$ and every $\gamma\in(0,1)$

$$\sup_x \left| a(x|\gamma) - (\surd(1-\gamma^2))^{-k} f([x-(1-\surd(1-\gamma^2))\mu]/\surd(1-\gamma^2)) \right|$$

$$\leq (\surd(1-\gamma^2))^{-k} \sup_{|z|\leq M} \sup_x |f(x-z)-f(x)|$$

$$\int_{|\gamma V^{\frac{1}{2}} w/\surd(1-\gamma^2)|\leq M} [\exp(-\tfrac{1}{2}w'w)/(\surd(2\pi))^k]dw$$

$$+ (\surd(1-\gamma^2))^{-k} 2\sup_x |f(x)|$$

$$\times \int_{|\gamma V^{\frac{1}{2}} w/\surd(1-\gamma^2)|\leq M} [\exp(-\tfrac{1}{2}w'w)/(\surd(2\pi))^k]dw$$

$$\leq (\surd(1-\gamma^2))^{-k} (\sup_{|z|\leq M} \sup_x |f(x-z)-f(x)|$$

$$+ 2\sup_x |f(x)| \int_{\{w'Vw > ((1-\gamma^2)/\gamma^2)M^2\}} [\exp(-\tfrac{1}{2}w'w)/(\surd(2\pi))^k]dw) \tag{7.4.41}$$

Choosing $M=\surd\gamma$ and letting $\gamma\to 0$ we now see that

$$\lim_{\gamma\downarrow 0} \sup_x |a(x|\gamma) - (1/\surd(1-\gamma^2))^k f([x-(1-\surd(1-\gamma^2))\mu]/\surd(1-\gamma^2))| = 0. \tag{7.4.42}$$

From the uniform continuity and boundedness of f(x) it easily follows

$$\lim_{\gamma\downarrow 0} \sup_x |(1/\surd(1-\gamma^2))^k f([x-(1-\surd(1-\gamma^2))\mu]/\surd(1-\gamma^2)) - f(x)| = 0. \tag{7.4.43}$$

Combining (7.4.42) and (7.4.43), (7.4.9) follows. This completes the proof lemma 7.4.1. Q.E.D.

Lemma 7.4.2. Let the conditions of lemma 7.4.1 be satisfied. Put

$$C_n(x|\gamma) = (1/n)\Sigma_{j=1}^n (x-\surd(1-\gamma^2)X_j - (1-\surd(1-\gamma^2))\overline{X})$$

$$\times \exp(-\tfrac{1}{2}[x-(\surd(1-\gamma^2))X_j - (1-\surd(1-\gamma^2))\overline{X}]\hat{V}^{-1}$$

$$\times [x-\surd(1-\gamma^2)X_j - (1-\gamma^2))\overline{X}]/\gamma^2) \quad / (\surd(2\pi))^k \gamma^k \surd\det(\hat{V}) \tag{7.4.44}$$

and

$$C(x|\gamma) = E(x-\surd(1-\gamma^2)X_j-(1-\surd(1-\gamma^2))\mu)$$

$$\times \exp\{-\tfrac{1}{2}[x-\surd(1-\gamma^2)X_j-(1-\surd(1-\gamma^2))\mu]'V^{-1}$$

$$\times [x-\surd(1-\gamma^2)X_j-(1-\surd(1-\gamma^2))\mu]/\gamma^2\} \;/\; (\surd(2\pi))^k\gamma^k\surd\det(V) \quad (7.4.45)$$

Then

$$\text{plim}_{n\to\infty}\sup_{\gamma\in[\zeta_n,1]}\sup_x|C_n(x|\gamma)-C(x|\gamma)| = 0. \qquad (7.4.46)$$

Moreover, $C(x|\gamma)$ is a uniformly continuous and bounded mapping from $R^k\times[0,1]$ into R^k satisfying

$$\lim_{\gamma\downarrow 0}\sup_x|C(x|\gamma)| = 0. \qquad (7.4.47)$$

Proof: If we differentiate (7.4.10) to x we see from (7.1.10) that

$$(\partial/\partial x')\hat{h}_j(x|\gamma)$$

$$= (1/(2\pi))^k \int(-i\cdot t)\exp[-i\cdot t'x+i\cdot t'X_j\surd(1-\gamma^2)+i\cdot t'\bar{X}(1-\surd(1-\gamma^2))]$$

$$\times \exp(-\tfrac{1}{2}\gamma^2t'\hat{V}t)\ dt$$

$$= -\ (1/\gamma^2)\hat{V}^{-1}(x-\surd(1-\gamma^2)X_j-(1-\surd(1-\gamma^2))\bar{X})$$

$$\times \exp\{-\tfrac{1}{2}[x-\surd(1-\gamma^2)X_j-(1-\surd(1-\gamma^2))\bar{X}]'\hat{V}^{-1}$$

$$\times [x-\surd(1-\gamma^2)X_j-(1-\surd(1-\gamma^2))\bar{X}]/\gamma^2\} \;/\; (\surd(2\pi))^k\surd\det(\hat{V}) \qquad (7.4.48)$$

hence

$$C_n(x|\gamma) = \gamma^2\hat{V}(1/n)\Sigma_{j=1}^n\Omega_{n,j}(x|\gamma), \qquad (7.4.49)$$

where

$$\Omega_{n,j}(x|\gamma)$$

$$= i(1/(2\pi))^k\int\exp[-i\cdot t'x+i\cdot t'X_j\surd(1-\gamma^2)+i\cdot t'\bar{X}(1-\surd(1-\gamma^2))]$$

$$\times t\cdot\exp(-\tfrac{1}{2}\gamma^2t'\hat{V}t)dt. \qquad (7.4.50)$$

Now put

$$\Omega_n(x|\gamma) = (1/n)\Sigma_{j=1}^n \Omega_{n,j}(x|\gamma) \tag{7.4.51}$$

and

$$\Omega_n^*(x|\gamma)$$

$$= i\{1/(2\pi)\}^k \int \{E \exp[i\cdot t'X_j\surd(1-\gamma^2)]\}\exp[-i\cdot t'x+i\cdot t'\overline{X}(1-\surd(1-\gamma^2))]$$
$$\times\ t\cdot\exp(-\tfrac{1}{2}\gamma^2 t'\hat{V}t)dt. \tag{7.4.52}$$

Similarly to (7.4.13) and (7.4.14) in the proof of lemma 7.4.1 it can be shown that

$$\text{plim}_{n\to\infty}\sup_{\gamma\in[\zeta_n,1-\varepsilon]}\sup_x|\Omega_n(x|\gamma)-\Omega_n^*(x|\gamma)| = 0 \tag{7.4.53}$$

and

$$\lim_{\varepsilon\downarrow 0}\lim_{n\to\infty}E\ \sup_{\gamma\in[1-\varepsilon,1]}\sup_x|\Omega_n(x|\gamma)-\Omega_n^*(x|\gamma)| = 0, \tag{7.4.54}$$

hence

$$\text{plim}_{n\to\infty}\sup_{\gamma\in[\zeta_n,1]}\sup_x|\Omega_n(x|\gamma)-\Omega_n^*(x|\gamma)| = 0 \tag{7.4.55}$$

and consequently

$$\text{plim}_{n\to\infty}\sup_{\gamma\in[\zeta_n,1]}\sup_x|C_n(x|\gamma)-C_n^*(x|\gamma)| = 0, \tag{7.4.56}$$

where

$$C_n^*(x|\gamma) = \gamma^2\hat{V}\ \Omega_n^*(x|\gamma). \tag{7.4.57}$$

Next observe from (7.4.48) and (7.4.52) that similar to (7.4.28),

$$\Omega_n^*(x|\gamma) = \int(1/\gamma^2)\hat{V}^{-1}(x-\surd(1-\gamma^2)z-(1-\surd(1-\gamma^2))\overline{X})$$

$$\times \exp\{-\tfrac{1}{2}[x-\surd(1-\gamma^2)z-(1-\surd(1-\gamma^2))\overline{X}]'\hat{V}^{-1}$$

$$\times [x-\surd(1-\gamma^2)z-(1-\surd(1-\gamma^2))\overline{X}]/\gamma^2\}$$

$$\times [(\surd(2\pi))^k\gamma^k\surd\det(\hat{V})]^{-1}h(z)dz$$

$$= (\surd(2\pi))^{-k}\gamma^{-1}(\surd(1-\gamma^2))^{-1}\hat{V}^{-\frac{1}{2}}\int h([x-(1-\surd(1-\gamma^2))\overline{X}-\gamma\hat{V}^{\frac{1}{2}}w]/\surd(1-\gamma^2))$$

$$\times w\cdot\exp(-\tfrac{1}{2}w'w)\cdot dw. \qquad (7.4.58)$$

Thus if we put

$$\Omega(x|\gamma)$$

$$= k(\surd(2\pi))^{-k}\gamma^{-1}(\surd(1-\gamma^2))^{-1}V^{-\frac{1}{2}}\int h([x-(1-\surd(1-\gamma^2))\mu-\gamma V^{\frac{1}{2}}w]/\surd(1-\gamma^2))$$

$$\times w\cdot\exp(-\tfrac{1}{2}w'w)dw \qquad (7.4.59)$$

then it follows similarly to the proof of (7.4.15) that

$$\text{plim}_{n\to\infty}\sup_{\gamma\in[0,1-\varepsilon]}\sup_x|\gamma\hat{V}^{\frac{1}{2}}\Omega_n^*(x|\gamma)-\gamma V^{\frac{1}{2}}\Omega(x|\gamma) = 0. \qquad (7.4.60)$$

Realizing that (7.4.59) can also be written as

$$\Omega(x|\gamma)$$

$$=i(1/(2\pi))^k\int\{E\exp[i\cdot t'X_j\surd(1-\gamma^2)]\}\exp[-i\cdot t'x+i\cdot t'\mu(1-\surd(1-\gamma^2))]$$

$$\times t\cdot\exp(-\tfrac{1}{2}\gamma^2t'Vt)dt \qquad (7.4.61)$$

(compare (7.4.52)), we see from (7.4.52), (7.4.61) and the proof of (7.4.16) that

$$\sup_{\gamma\in[1-\varepsilon,1]}\sup_x|\Omega_n^*(x|\gamma)-\Omega(x|\gamma)| = 0. \qquad (7.4.62)$$

From (7.4.60) and (7.4.56) we obtain

$$\text{plim}_{n\to\infty}\sup_{\gamma\in[0,1-\varepsilon]}\sup_x|C_n^*(x|\gamma)-\gamma^2 V\Omega(x|\gamma)|$$

$$\leq \text{plim}_{n\to\infty}\sup_{\gamma\in[0,1-\varepsilon]}\sup_x|\gamma\hat{V}^{1/2}\Omega_n^*(x|\gamma)-\gamma V^{1/2}\Omega(x|\gamma))|$$

$$+\ \text{plim}_{n\to\infty}\sup_{\gamma\in[\zeta_n,1-\varepsilon]}\sup_x|\gamma^2(\hat{V}^{1/2}-V^{1/2})V^{1/2}\Omega(x|\gamma)| = 0 \qquad (7.4.63)$$

because

$$\text{plim}_{n\to\infty}\hat{V} = V$$

and $|\Omega(x|\gamma)|$ is bounded on $R^k\times[0,1]$. Similarly we conclude from (7.4.62) that

$$\text{plim}_{n\to\infty}\sup_{\gamma\in[1-\varepsilon,1]}\sup_x|C_n^*(x|\gamma)-\gamma^2 V\Omega(x|\gamma)| = 0 \qquad (7.4.64)$$

and consequently, combining (7.4.63) and (7.4.64),

$$\text{plim}_{n\to\infty}\sup_{\gamma\in[0,1]}\sup_x|C_n^*(x|\gamma)-\gamma^2 V\Omega(x|\gamma)| = 0.$$

Since $C(x|\gamma)=\gamma^2 V\Omega(x|\gamma)$, (7.4.46) follows from (7.4.55) and (7.4.65). The rest of the lemma can be proved in the same way as in the proof of lemma 7.4.1. Q.E.D.

Proof of theorem 7.1.2: Theorem 7.1.2 follows from lemma 7.4.1 by substituting $Y_j=1$. Q.E.D.

Proof of theorem 7.1.4: Theorem 7.1.4 follows straightforwardly from lemma 7.1.1 in section 7.1.4 if

$$\text{plim}_{n\to\infty}\sup_{\gamma\in[0,1]}|(1/n)\Sigma_{j=1}^n h(X_j|\gamma) - \int h(x|\gamma)h(x)dx| = 0, \qquad (7.4.66)$$

where $h(x|\gamma)$ is defined by (7.1.18). But $h(x|\gamma)$ is uniformly continuous and uniformly bounded on $R^k\times[0,1]$, as follows from lemma 7.4.1. So (7.4.66) follows from theorem 2.7.5 and hence theorem 7.1.4 follows from lemma 7.1.1. Q.E.D.

Proof of lemma 7.1.1: Lemma 7.1.1 follows straightforwardly from (7.1.13) and lemma 7.4.1 by substituting $Y_j=1$. Q.E.D.

Proof of theorem 7.2.1: From the conditions of theorem 7.2.1 it follows that the vector β of OLS parameter estimators and the OLS estimator $\hat{\alpha}$ of the constant term of a linear regression model converge in probability:

$$\text{plim}_{n\to\infty}\ \hat{\alpha} = \alpha,\ \text{plim}_{n\to\infty}\ \hat{\beta} = \beta,$$

say. Now observe from (7.2.4), (7.4.5) and (7.4.44) that

$$\hat{g}(x|\gamma,\gamma_n^{(2)})\hat{h}(x|\gamma_n^{(2)}) = \int y\hat{f}(y,x|\gamma)dy =$$

$$= \surd(1-\gamma^2)a_n(x|\gamma) + (1-\surd(1-\gamma^2))\overline{Y}\ \hat{h}(x|\gamma) + \hat{\beta}'C_n(x|\gamma)$$

$$= d_n(x|\gamma), \tag{7.4.67}$$

say. Put

$$d(x|\gamma) = \surd(1-\gamma^2)a(x|\gamma) + (1-\surd(1-\gamma^2))(\alpha+\beta'\mu)h(x|\gamma) + \beta'C(x|\gamma), \tag{7.4.68}$$

where $a(x|\gamma)$, $h(x|\gamma)$ and $C(x|\gamma)$ are defined by (7.4.6), (7.1.18) and (7.4.45), respectively, and

$$\mu = \text{plim}_{n\to\infty}\overline{X}.$$

Realizing that

$$\text{plim}_{n\to\infty}\overline{Y} = \alpha + \beta'\mu$$

and that $d(x|\gamma)$ is uniformly bounded on $R^k\times[0,1]$, we conclude from the lemma's 7.4.1, 7.1.1 and 7.4.2 that

$$\text{plim}_{n\to\infty}\sup_{\gamma\in[\zeta_n,1]}\sup_x|d_n(x|\gamma)-d(x|\gamma)| = 0. \tag{7.4.69}$$

Using this result and the fact that by theorem 7.1.4

$$\text{plim}_{n\to\infty}\sup_x|\hat{h}(x|\gamma_n^{(2)})-h(x)| = 0, \tag{7.4.70}$$

it is easy to show

$$\text{plim}_{n\to\infty}\sup_{\gamma\in[\zeta_n,1]}|(1/n)\Sigma_{j=1}^n[Y_j\hat{h}(X_j|\gamma_n^{(2)})-d_n(X_j|\gamma)]^2$$

$$-(1/n)\Sigma_{j=1}^{n}[Y_j h(X_j) - d(X_j|\gamma)]^2| = 0. \quad (7.4.71)$$

Moreover, from the conditions of the theorem under review and the fact that $d(x|\gamma)$ is uniformly bounded and continuous on $R^k\times[0,1]$ it follows that theorem 2.7.5 is applicable, hence

$$\mathrm{plim}_{n\to\infty}\sup_{\gamma\in[0,1]}|(1/n)\Sigma_{j=1}^{n}\{Y_j h(X_j)-d(X_j|\gamma)\}^2 - E\ U_j^2 h(X_j)^2$$
$$-\int\{g(x)h(x)-d(x|\gamma)\}^2 h(x)dx| = 0 \quad (7.4.72)$$

where $U_j=Y_j-g(X_j)$. Putting

$$Q(\gamma) = E\ U_j{}^2 h(X_j)^2 + \int\{g(x)h(x)-d(x|\gamma)\}^2 h(x)dx, \quad (7.4.73)$$

we have proved by now

$$\mathrm{plim}_{n\to\infty}\sup_{\gamma\in[\zeta_n,1]}|Q_n(\gamma)-Q(\gamma)| = 0. \quad (7.4.74)$$

Consequently we have

$$\mathrm{plim}_{n\to\infty}\int\{g(x)h(x)-d(x|\gamma_n^{(1)})\}^2 h(x)dx = 0. \quad (7.4.75)$$

Similarly to the proof of theorem 7.1.5 it can now be shown that (7.4.75) implies

$$\mathrm{plim}_{n\to\infty}\sup_{x\in\{x\in R^k:h(x)\geq\varepsilon\}}|d_n(x|\gamma_n^{(1)})-g(x)h(x)| = 0 \quad (7.4.76)$$

for $\varepsilon\in(0,\sup_x h(x)]$. The SMINK regression case considered in theorem 7.4.1 now follows straightforwardly from (7.4.76).

The modified SMINK regression case follows from

$$\mathrm{plim}_{n\to\infty}\sup_{\gamma\in[\zeta_n,1]}\sup_x|\hat{g}^*(x|\gamma,\gamma_2)\hat{h}^*(x|\gamma_2)$$
$$-\hat{g}(x|\gamma,\gamma_2)\hat{h}(x|\gamma_2)| = 0, \quad (7.4.77)$$

which is not hard to verify. Q.E.D.

References

Anderson,T.W. (1958), *Introduction to Multivariate Statistical Analysis*, New York-London: John Wiley.

Bierens,H.J. (1983), "Sample Moments Integrating Normal Kernel Estimators of Multivariate Density and Regression Functions", *Sankhya B* 45, 160-192.

Epanechnikov,V.A. (1969), "Non-parametric Estimation of a Multivariate Probability Density", *Theory of Probability and Its Applications*, 14, 153-158.

Good,I.J. and R.A.Gaskins (1972), "Global Nonparametric Estimation of Probability Densities", *Virginia Journal of Science*, 23, 171-193.

Silverman,B.W. (1978), "Choosing the Window Width when Estimating a Density", *Biometrica* 65, 1-11.

1988-1	H. Visser	Austrian thinking on international economics
1988-2	A.H.Q.M. Merkies T. van der Meer	Theoretical foundations for the 3-C model
1988-3	H.J. Bierens J. Hartog	Nonlinear regression with discrete explanatory variables, with an application to the earnings function
1988-4	N.M.van Dijk	On Jackson's product form with 'jump-over' blocking
1988-5	N.M.van Dijk M.Rumsewicz	Networks of queues with service anticipating routing
1988-6	H.Linneman C.P.van Beers	Commodity Composition of Trade in Manufactures and South-South Trade Potential
1988-7	N.M.van Dijk	A LCFS finite buffer model with batch input and non-exponential sevices
1988-8	J.C.W.van Ommeren	Simple approximations for the batch-arrival $M^X/G/1$ queue
1988-9	H.C.Tijms	Algorithms and approximations for bath-arrival queues
1988-10	J.P.de Groot H.Clemens	Export Agriculture and Labour Market in Nicaragua
1988-11	H.Verbruggen J.Wuijts	Patterns of South-South trade in manufactures
1988-12	H.C.Tijms J.C.W.van Ommeren	Asymptotic analysis for buffer behaviour in communication systems
1988-13	N.M.van Dijk E.Smeitink	A non-exponential queueing system with batch servicing
1988-14	J.Rouwendal	Existence and uniqueness of stochastic price equilibria in heterogeneous markets
1988-15	H.Verbruggen	GSTP, the structure of protection and South-South trade in manufactures
1988-16	Mevr.H.Weijland Mevr.R Herweijer J.de Groot	Female participation in agriculture in the Dominican Republic
1988-17	N.M.van Dijk	Product Forms for Random Access Schemes
1988-18	A.H.Q.M.Merkies I.J.Steyn	Adaptive Forecasting with Hyperfilters
1988-19	J.Rouwendal	Specification and Estimation of a Logit Model for Housing Choice in the Netherlands
1988-20	J.C.W.van Ommeren	An elementary proof of a basic result for the