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

By HERMAN J. BIERENS

Foundation for Economic Research, University of Amsterdam

SUMMARY. In this paper we propose a new class of kernel estimators of multivariate
density and regression functions. We shall construct a normal kernel density function 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.

1. INTRODUCTION

The problem of nonparametric estimation of a density function has got
extensive attention in the statistical literature (see Fryer, 1977 and Turi
and Thompson, 1978 for reviews) and since the pioneering papers of
Nadaraya (1964) and Watson (1964) there is now a growing extent of literature
on the related problem of nonparametric estimation of a regression function
(see Collomb, 1981 for a review). Most of the authors focus their attention
on asymptotic properties, i.e., pointwise and uniform consistency and
asymptotic normality. Empirical applications of these methods on large
scale multivariate data sets are, however, scarce. This is probably due to
some practical and computational difficulties of fitting these estimators.
For a particular class of nonparametric estimators of density and regression
functions, the so-called class of kernel estimators, these difficulties involve
the choice of an appropriate kernel and of an appropriate window width.
Some work has already been done on these problems, but only from an asympto-
tic point of view. We mention here the work of Epanechnikov (1962), who
established an optimal form of a product kernel in the case of density estima-
tion, and Greblicki and Krzyzak (1980) who confirmed the result of Epane-
chnikov for the regression case. The result of Epanechnikov suggests that
there exists an optimal kernel which is independent of the distribution of
the data. But in his approach the class of kernels taken into account is
reduced in size by normalization, by which the impact of the distribution of
the data has been moved out from the kernel to the window width parameters.
So in practical applications the problem remains how to find appropriate
empirical estimates of the window width parameters. Silverman (1978) has
developed a graphical method for establishing the window width for which
the asymptotic rate of convergence of a kernel density estimator (with fixed
kernel) is optimal, but his approach appears only to be practical in the uni-
variate and, to some extent, the bivariate case.

In this paper we shall propose a practical approach for specifying kernel
estimators of multivariate density and regression functions. Moreover, we
shall propose alternative methods for choosing appropriate window width
parameters, for the density as well as for the regression case.

In Section 2 we consider a kernel estimator \( \hat{h}_n(x|\gamma) \) of a \( k \)-variate density
\( h(x) \), say, where \( \gamma \) is a window width parameter, such that for every \( \gamma \in (0, 1] \),
\[
\bar{h}_n(x|\gamma) \geq 0, \quad \int_{\mathbb{R}^k} \bar{h}_n(x|\gamma) dx = 1 \text{ a.s.,}
\]
\[
\int_{\mathbb{R}^k} x \bar{h}_n(x|\gamma) dx = \frac{1}{n} \sum_{j=1}^{n} x_j \text{ a.s.}
\]
and
\[
\int_{\mathbb{R}^k} xx' \bar{h}_n(x|\gamma) dx = \frac{1}{n} \sum_{j=1}^{n} x_j x_j' \text{ a.s.,} \quad \ldots \quad (1.1)
\]
where \( \{x_1, \ldots, x_n\} \) is the data set involved. Given the window width para-
meter \( \gamma \) and the functional type of the kernel, the density estimator turns out
to be completely and explicitly determined by the conditions (1.1). In this
paper we shall only consider kernel estimators based on kernels of the multivi-
ariate normal density type. For obvious reasons, these density estimators
satisfying condition (1.1) will be called Sample Moments Integrating Normal
Kernel (SMNK) estimators.

As an alternative to Silverman's approach we shall propose to estimate
the window width \( \gamma \) by minimizing an estimate of the integrated square error
\[
\int_{\mathbb{R}^k} (\hat{h}_n(x|\gamma) - h(x))^2 dx.
\]

Section 3 is devoted to the regression case. We consider a data set
\( \{(y_j, x_j)\}_{j=1}^{n} \), where the \( (y_j, x_j) \)'s are random vectors in \( \mathbb{R}^k \times \mathbb{R}^k \) with common
density \( f(y, x) \), say, and marginal density \( h(x) = \int_{-\infty}^{\infty} f(y, x) dy \). We shall esti-
mate the regression function
\[
g(x) = E(y_j | x_j = x), \quad x \in \mathbb{R}^k, \quad h(x) > 0 \quad \ldots \quad (1.2)
\]
by
\[ \tilde{g}_n(x | \gamma_1, \gamma_2) = \frac{\int_{-\infty}^{+\infty} \tilde{f}_n(y, x | \gamma_1) dy}{\tilde{h}_n(x | \gamma_2)}, \quad \text{... (1.3)} \]

where \( \tilde{f}_n(y, x | \gamma) \) and \( \tilde{h}_n(x | \gamma) \) are SMINK estimators of the densities \( f(y, x) \) and \( h(x) \), respectively, and \( \gamma_1 \) and \( \gamma_2 \) are window width parameters. A neat property of this estimate is that contains the linear regression model as a special case, namely \( \tilde{g}_n(x | 1, 1) = \tilde{\beta} + \beta \cdot x \), where \( \tilde{\beta} \) and \( \beta \) are the OLS estimates of the constant term and the vector of regression coefficients, respectively, of the linear regression of \( y_i \) on \( x_i \).

Since the regression estimator (1.3) is based on SMINK density estimators we shall call it too a SMINK estimator of a regression function. In deviation to the usual approach in nonparametric regression we shall treat the two window width parameters \( \gamma_1 \) and \( \gamma_2 \) apart, i.e., we first estimate \( \gamma_1 \) by using the minimum integrated square error criterion above and then, given \( \gamma_2 \), we propose to estimate \( \gamma_1 \) by minimizing the weighted mean square error
\[ \frac{1}{n} \sum_{j=1}^{n} \{(y_j - \tilde{g}_n(x_j | \gamma_1, \gamma_2)) \tilde{h}_n(x_j | \gamma_2)\}^2. \]

Finally, in Section 4 we shall demonstrate the performance of the SMINK density and regression estimators by a numerical example.

2. Kernel Estimators of a Density Function

2.1. Integral conditions. Let \( x_1, x_2, ..., x_n \) be an i.i.d. stochastic process in \( \mathbb{R}^k \) and suppose that the distribution of the \( x_j \)'s is absolutely continuous with density \( h(x) \). A kernel estimator of \( h(x) \) is a random function of the form
\[ \hat{h}_n(x) = \frac{1}{n} \sum_{j=1}^{n} \psi \left( \frac{x - x_j}{\gamma_n} \right) \frac{1}{\gamma_n^n}, \quad \text{... (2.1)} \]

where \( \psi \) is a real function on \( \mathbb{R}^k \), called the kernel, satisfying
\[ \int_{\mathbb{R}^k} \psi(x) dx = 1, \quad \text{... (2.2)} \]
and \((\gamma_n)\) is a sequence of positive numbers, called window width parameters, converging to zero for \(n \to \infty\). As is shown by Parzen (1962), Cacoullos (1966) and others (see the reviews by Fryer, 1977 and Tapia and Thompson, 1978), the kernel \(\psi\) and the window width parameter \(\gamma_n\) can be chosen such that \(\text{plim } \hat{h}_n(x) = h(x)\) for every continuity point \(x\) of \(h(x)\) or even

\[
\text{plim } \sup_{x \in \mathbb{R}^d} |\hat{h}_n(x) - h(x)| = 0
\]

if \(h(x)\) is everywhere continuous.

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 \(\psi\) be a density itself, so that we then have

**Condition a:** \(\hat{h}_n(x) \geq 0, \int_{\mathbb{R}^d} \hat{h}_n(x) dx = 1\) a.s.

A further reasonable requirement is that the integral \(\int_{\mathbb{R}^d} x \hat{h}_n(x) dx\) is an appropriate estimator of the mathematical expectation \(\int_{\mathbb{R}^d} x h(x) dx\). Since the usual estimator of the latter integral is the sample mean

\[
\bar{x} = \frac{1}{n} \sum_{j=1}^{n} x_j,
\]

we therefore require

**Condition b:** \(\int_{\mathbb{R}^d} x \hat{h}_n(x) dx = \bar{x}\) a.s.

But from (2.1) and (2.2) it follows

\[
\int_{\mathbb{R}^d} x \hat{h}_n(x) dx = \frac{1}{n} \sum_{j=1}^{n} x_j \left( x_j - \frac{x_j}{\gamma_n} \right) \frac{1}{\gamma_n} dx = \bar{x} + \gamma_n \int_{\mathbb{R}^d} x \psi(x) dx, \quad \cdots \ (2.3)
\]

so that the kernel estimator (2.1) satisfies condition (b) if the kernel is a density with zero mean vector. We may go on this way by requiring that also the matrix \(\int_{\mathbb{R}^d} xx' \hat{h}_n(x) dx\) is an appropriate estimator of \(\int_{\mathbb{R}^d} xx' h(x) dx\). The usual
estimator of the latter matrix is the matrix $\frac{1}{n} \sum_{j=1}^{n} x_j \hat{e}_j$. Therefore we require

$$\text{Condition c: } \int \frac{x x' \hat{h}_n(x) dx}{\gamma_n} = \frac{1}{n} \sum_{j=1}^{n} x_j \hat{e}_j \text{ a.s.}$$

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

$$\int \frac{x x' \hat{h}_n(x) dx}{\gamma_n} = \frac{1}{n} \sum_{j=1}^{n} x_j \hat{e}_j + \gamma_n \int \frac{x x' \psi(x) dx}{\gamma_n} + \gamma_n^2 \int \frac{x x' \psi(x) dx}{\gamma_n} \Rightarrow \int \frac{x x' \psi(x) dx}{\gamma_n}$$

which dominates $\frac{1}{n} \sum_{j=1}^{n} x_j \hat{e}_j$ by a positive definite matrix $\gamma_n^2 \int \frac{x x' \psi(x) dx}{\gamma_n}$.

2.2. 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 (2.1) by $(1 - \beta_n) x_j$, where $(\beta_n)$ is a sequence of positive numbers converging to zero for $n \to \infty$. Thus we consider

$$\hat{h}_n(x) = \frac{1}{n} \sum_{j=1}^{n} \psi_n \left( \frac{x - (1 - \beta_n) x_j}{\gamma_n} \right) \frac{1}{\gamma_n}, \quad \ldots \quad (2.6)$$

where $\psi_n$ is a density. Clearly condition (a) is satisfied. Moreover, condition (b) now becomes

$$\mathbb{E} = \int \frac{x \hat{h}_n(x) dx}{\gamma_n} = \frac{1}{n} \sum_{j=1}^{n} \int (1 - \beta_n) x_j + \gamma_n x \psi_n(x) dx \Rightarrow \int \frac{x \psi_n(x) dx}{\gamma_n}, \quad \ldots \quad (2.7)$$

hence

$$\int \frac{x \psi_n(x) dx}{\gamma_n} = \frac{\beta_n}{\gamma_n} \mathbb{E}, \quad \ldots \quad (2.8)$$
Furthermore, condition (c) now becomes
\[
\frac{1}{n} \sum_{j=1}^{n} x y_j = \int_{\mathbb{R}^2} x x' \mathcal{H}_n(x) dx
\]
\[
= \frac{1}{n} \sum_{j=1}^{n} [(1-\beta_n) x_j + \gamma_n z_j] [(1-\beta_n) x_j + \gamma_n z_j] \psi_n(x) dx
\]
\[
= (1-\beta_n)^2 \frac{1}{n} \sum_{j=1}^{n} x y_j + \gamma_n (1-\beta_n) \int_{\mathbb{R}^2} x' \psi_n(x) dx
\]
\[
+ \gamma_n (1-\beta_n) \int_{\mathbb{R}^2} x' \psi_n(x) dx \mathcal{E} + \gamma_n^2 \int_{\mathbb{R}^2} x x' \psi_n(x) dx,
\]
where the latter equality follows from (2.8). Obviously, (2.8) and (2.9) imply
\[
\int_{\mathbb{R}^2} x x' \psi_n(x) dx = \frac{1}{\gamma_n^2} \left\{ (2\beta_n - \beta_n^2) \frac{1}{n} \sum_{j=1}^{n} x y_j - (2\beta_n - \beta_n^2) \mathcal{E} \right\}
\]
\[
= \frac{2\beta_n - \beta_n^2}{\gamma_n^2} \left\{ \frac{1}{n} \sum_{j=1}^{n} x y_j - \mathcal{E} \right\} + \int_{\mathbb{R}^2} x \psi_n(x) dx \int_{\mathbb{R}^2} x' \psi_n(x) dx,
\]
Thus in order that \( \mathcal{H}_n(x) \) satisfies the conditions (a), (b) and (c), \( \psi_n \) should be a density with mean vector \( \beta_n \bar{z} \) and variance matrix \( \frac{2\beta_n - \beta_n^2}{\gamma_n^2} \mathcal{E} \), where
\[
\mathcal{E} = \frac{1}{n} \sum_{j=1}^{n} (x_j - \bar{x})(x_j - \bar{x})',
\]
which is the usual sample variance matrix. Therefore, if we put
\[
\psi_n(x) = \psi \left( \frac{2\beta_n - \beta_n^2}{\gamma_n^2} \mathcal{E} \right)^{-1} \left( x - \beta_n \bar{z} \right)
\]
\[
\psi(x) \geq 0, \quad \int_{\mathbb{R}^2} \psi(x) dx = 1, \quad \int_{\mathbb{R}^2} x \psi(x) dx = 0, \quad \int_{\mathbb{R}^2} x x' \psi(x) dx = I,
\]
where \( \psi \) is a density with zero mean vector and variance equal to the unit-matrix
\[
(2.13)
\]
then clearly, $\psi^*$ has all the required properties. Now substitute (2.12) in (2.8). We then obtain

$$
\tilde{h}_n^*(x) = \frac{1}{n} \sum_{j=1}^{n} \frac{\psi\left(V^{-1}x - \frac{(1 - \beta_n)x_j - \beta_n\bar{x}}{\sqrt{\beta_n - \beta_n^2}}\right)}{(\sqrt{2\beta_n - \beta_n^2})^n |V_x|^{\frac{1}{2}}}.
$$

(2.14)

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

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

(2.15)

so that (2.14) now becomes

$$
\tilde{h}_n^*(x) = \frac{1}{n} \sum_{j=1}^{n} \frac{\psi(V^{-1}x - \sqrt{1 - \gamma_n^2}x_j + (1 - \sqrt{1 - \gamma_n^2})\bar{x})/\gamma_n)}{\gamma_n |V_x|^\frac{1}{2}}
$$

(2.16)

Of course, $\gamma_n$ should be chosen in the interval $(0, 1]$, otherwise (2.16) may no longer be a real random function.

We have shown by now

**Theorem 1:** Let $x_1, ..., x_n$ be random vectors in $\mathbb{R}^k$. Define

$$
\tilde{h}_n(x | \gamma) = \frac{1}{n} \sum_{j=1}^{n} \frac{\psi(V^{-1}x - \sqrt{1 - \gamma^2}x_j + (1 - \sqrt{1 - \gamma^2})\bar{x})/\gamma)}{\gamma |V_x|^\frac{1}{2}}
$$

(2.17)

where $\psi$ is a real function on $\mathbb{R}^k$ satisfying condition (2.13), $\bar{x}$ is the sample mean and $V_x$ is the sample variance matrix. Then for every $\gamma \in (0, 1]$, $\tilde{h}_n(x | \gamma)$ satisfies the conditions (a), (b) and (c).

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

$$
\psi(x) = \frac{e^{-\frac{1}{2}x^2}}{(\sqrt{2\pi})^k},
$$

(3.18)

we have the following result.
Density and regression functions

Theorem 2: Let \( x_1, x_2, \ldots, x_n, \ldots \) be an i.i.d. stochastic process in \( \mathbb{R}^d \). Let the distribution of the \( x_j \)'s be absolutely continuous with continuous density \( h(x) \) and finite second moments. Let \( (\zeta_n) \) be any sequence satisfying

\[
\zeta_n \in (0, 1], \quad \lim \zeta_n = 0 \quad \text{and} \quad \lim \sqrt{n} \zeta_n^2 = \infty \quad \ldots \quad (2.19)
\]

and let \( (\gamma_n) \) be any (random) sequence satisfying

\[
\gamma_n \in [\zeta_n, 1] \quad \text{(a.s.),} \quad (p) \lim \gamma_n = 0 \quad \ldots \quad (2.20)
\]

If

\[
\operatorname{plim} z = \mu \quad \text{and} \quad \operatorname{plim} V_z = V \quad \text{exist,} \quad \ldots \quad (2.21)
\]

where \( V \) is positive definite, then

\[
\operatorname{plim} \sup_{x \in \mathbb{R}^d} | \tilde{h}_n(x|\gamma_n) - h(x) | = 0, \quad \ldots \quad (2.22)
\]

where \( \tilde{h}_n(x|\gamma) \) is defined as in Theorem 1 with \( \psi \) defined by (2.18).

Proof: Appendix.

In the following 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 (2.18). 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 3. Moreover, we note that a normal kernel is nearly optimal in the sense of Epanechnikov (1969).

The kernel density estimator considered in Theorem 2, i.e.,

\[
\tilde{h}_n(x|\gamma) = \frac{1}{n} \sum_{j=1}^{n} \tilde{h}_{n,j}(x|\gamma), \quad \ldots \quad (2.23)
\]

where

\[
\tilde{h}_{n,j}(x|\gamma) = \exp \left\{ -\frac{1}{2} (x - \sqrt{1 - \gamma^2} x_j - (1 - \sqrt{1 - \gamma^2}) z)^\top \left( \frac{1}{n} \sum_{i=1}^{n} \tilde{h}_{n,i}(x|\gamma) \right) \right\} \quad \ldots \quad (2.24)
\]
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.

2.3. The Dirac-catastrophe and the modified SMINK density estimator.

It is easy to verify from (2.23) and (2.24) that

\[
\lim_{\gamma \to 0} \tilde{h}_n(x|\gamma) = \begin{cases} 
\infty & \text{if } x \in \{x_1, \ldots, x_n\} \\
0 & \text{if } x \notin \{x_1, \ldots, x_n\}.
\end{cases} \quad \ldots (2.25)
\]

This is the so-called Dirac-catastrophe (see Good and Gaskins, 1972). Thus if \( \gamma \) is too small then the estimate \( \tilde{h}_n(x|\gamma) \) 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 (2.23) by leaving out the \( \tilde{h}_{n,j}(x|\gamma) \) for which the corresponding observation \( x_j \) equals \( x \), i.e. to modify (2.23) to

\[
\frac{1}{n} \sum_{j=1}^{n} I(x_j \neq x) \tilde{h}_{n,j}(x|\gamma), \quad \ldots (2.26)
\]

where \( I(\cdot) \) is the well-known indicator function.

We have seen that the marginal densities of a SMINK density estimator are SMINK density estimators themselves. This neat property, however, does not carry over to estimators of the type (2.26), but it does for the following class of density estimators

\[
\tilde{h}_n(x|\gamma) = \frac{1}{n} \sum_{j=1}^{n} \prod_{i=1}^{k} I(x^{(i)} \neq x_{\gamma}) \tilde{h}_{n,j}(x|\gamma), \quad \ldots (2.27)
\]

where the \( x^{(i)} \) are the components of \( x \) and the \( x_{\gamma} \) are the components of \( x_j \). Since \( \tilde{h}_n(x|\gamma) \) satisfies the conditions (a), (b) and (c) as is not hard to verify, we shall call (2.27) the modified SMINK density estimator. Obviously, we now have

\[
\lim_{\gamma \to 0} \tilde{h}_n(x|\gamma) = 0 \quad \text{for all } x \in \mathbb{R}^k. \quad \ldots (2.28)
\]

Moreover it is not hard to verify that under the conditions of Theorem 2,

\[
\lim_{\gamma \to 0} \sup_{x \in \mathbb{R}^k} \sup_{\gamma \in (0,1]} |\tilde{h}_n(x|\gamma) - \tilde{h}_n(x|\gamma)| = 0, \quad \ldots (2.29)
\]

hence
Theorem 3: Theorem 2 carries over for the modified SMINK density estimator.

2.4. How to choose the window width of the modified SMINK density estimator? The problem is to choose $\gamma$ in $[\xi_n, 1]$ such that the “distance” between the modified SMINK estimate $\tilde{h}_n(x | \gamma)$ and the true density $h(x)$ is minimal. Thus we have to choose $\gamma_n$ in $[\xi_n, 1]$ such that

$$
\rho(\tilde{h}_n(x | \gamma_n), h(x)) = \inf_{\gamma \in [\xi_n, 1]} \rho(\tilde{h}_n(x | \gamma), h(x)),
$$

where $\rho$ is some metric on the space of $k$-variate densities. The best metric would be

$$
\rho(h_1, h_2) = \sup_{x \in \mathbb{R}^k} |h_1(x) - h_2(x)|,
$$

since this metric corresponds with 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

$$
\rho(h_1, h_2) = \int_{\mathbb{R}^k} (h_1(x) - h_2(x))^2 dx.
$$

Thus we propose to minimize

$$
\rho(\tilde{h}_n(x | \gamma), h(x)) = \int_{\mathbb{R}^k} (\tilde{h}_n(x | \gamma) - h(x))^2 dx
$$

$$
= \int_{\mathbb{R}^k} \tilde{h}_n(x | \gamma)^2 dx - 2 \int_{\mathbb{R}^k} \tilde{h}_n(x | \gamma) h(x) dx + \int_{\mathbb{R}^k} h(x)^2 dx.
$$

Also now the term involving $h(x)$ are unobservable. But,

Theorem 4: Under the conditions of Theorem 2

$$
\text{plim} \sup_{\gamma \in [\xi_n, 1]} \left| \int_{\mathbb{R}^k} \tilde{h}_n(x | \gamma) h(x) dx - \frac{1}{n} \sum_{j=1}^{n} \tilde{h}_n(x_j | \gamma) \right| = 0.
$$

Proof: Appendix.

Thus we have

$$
\text{plim} \sup_{\gamma \in [\xi_n, 1]} \left| \int_{\mathbb{R}^k} (\tilde{h}_n(x | \gamma) - h(x))^2 dx - \left\{ \int_{\mathbb{R}^k} \tilde{h}_n(x | \gamma)^2 dx - \frac{1}{n} \sum_{j=1}^{n} \tilde{h}_n(x_j | \gamma) \right\} \right| = 0.
$$

... (2.30)
Putting
\[ \tilde{g}_a(\gamma) = \frac{1}{n} \sum_{j=1}^{n} \tilde{h}_a(x_j | \gamma), \quad \cdots \tag{2.31} \]
the result (2.30) now suggests to choose \( \gamma_n \in [\gamma_a, 1] \) such that
\[ \tilde{g}_a(\gamma_n) = \inf_{\gamma \in [\gamma_a, 1]} \tilde{g}_a(\gamma). \quad \cdots \tag{2.32} \]

Note that
\[ \frac{1}{n} \sum_{j=1}^{n} \tilde{h}_a(x_j | \gamma)^2 \, dx = \frac{1}{n} \sum_{j=1}^{n} \tilde{h}_a(x_j | \gamma)^2 \, dx \]
\[ = \frac{1}{n} \sum_{j=1}^{n} \frac{\exp \left[ -\frac{1}{4} \left( x_{j1} - x_{j2} \right) \right]}{V^{-1}(x_{j1} - x_{j2}) \gamma^2} \]
so that
\[ \lim_{\gamma \downarrow 0} \frac{1}{n} \sum_{j=1}^{n} \tilde{h}_a(x_j | \gamma)^2 \, dx = \infty. \quad \cdots \tag{2.34} \]

From (2.34) and (2.28) it thus follows
\[ \lim_{\gamma \downarrow 0} \tilde{g}_a(\gamma) = \infty. \quad \cdots \tag{2.35} \]

However, if we would employ in (2.31) the common SMINK density estimator (2.24) then the procedure (2.32) would break down, due to the fact that
\[ \lim_{\gamma \downarrow 0} \frac{1}{n} \sum_{j=1}^{n} \tilde{h}_a(x_j | \gamma)^2 \, dx - 2 \frac{1}{n} \sum_{j=1}^{n} \tilde{h}_a(x_j | \gamma) = \infty. \]

This is the main reason for working with the modified SMINK estimator instead of the common SMINK estimator (2.24).

This approach for choosing \( \gamma \) 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 \( \gamma_n \) 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 $(\gamma_n)$ derived from (2.32) is such that \( \lim \sup_{x \in \mathcal{R}^k} |\tilde{h}_n(x|\gamma_n) - h(x)| = 0 \). Let \( V \) be the probability limit of \( V_x \) and let \( \mu \) be the probability limit of \( \varepsilon \). Put
\[
\tilde{h}(x|\gamma) = \int_{\mathcal{R}^k} \frac{\exp\left[-\frac{1}{2}(x - \sqrt{1-\gamma^2}(1-\sqrt{1-\gamma^2})\mu)^T V^{-1}(x - \sqrt{1-\gamma^2}(1-\sqrt{1-\gamma^2})\mu)\gamma^T\right]}{\sqrt{(2\pi)^k |V|}} |V|^{1/2} \times h(z)dz.
\]
Then

**Lemma 1:** Under the conditions of Theorem 2,

(i) \( \lim \sup_{\gamma \in [0,1]} \sup_{x \in \mathcal{R}^k} |\tilde{h}_n(x|\gamma) - h(x|\gamma)| = 0 \),

(ii) \( h(x|\gamma) \) is uniformly continuous on \( \mathcal{R}^k \times [0,1] \)

and

(iii) \( \lim_{\gamma \to \mu} \sup_{x \in \mathcal{R}^k} |h(x|\gamma) - h(x)| = 0 \).

**Proof:** Appendix.

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

**Lemma 2:** Under the conditions of Theorem 2,

\[ \lim_{\gamma \to \mu} \sup_{x \in \mathcal{R}^k} |\tilde{h}_n(x|\gamma) - q(\gamma)| = 0, \]

where
\[ q(\gamma) = \int_{\mathcal{R}^k} h(x|\gamma)^2dx - 2 \int_{\mathcal{R}^k} h(x|\gamma)h(x)dx. \]

Obviously we see from (2.37) that \( q(\gamma) \) takes a global minimum in \( \mathcal{R}^k \) only for \( \gamma \in [0,1] \) such that \( \int_{\mathcal{R}^k} (h(x|\gamma) - h(x))^2dx = 0 \). Moreover, by the continuity of \( h(x|\gamma) \) and \( h(x) \) on \( \mathcal{R}^k \times [0,1] \) and \( \mathcal{R}^k \), respectively, it follows that the latter condition holds if and only if \( h(x|\gamma) = h(x) \) for every \( x \in \mathcal{R}^k \). Thus we have
\[ \gamma \in [0,1] : q(\gamma) = \inf_{\gamma \in [0,1]} q(\gamma) \]
\[ = [\gamma \in [0,1] : \int_{\mathcal{R}^k} (h(x|\gamma) - h(x))^2dx] \]
\[ = [\gamma \in [0,1] : h(x|\gamma) = h(x) \text{ for every } x \in \mathcal{R}^k] = \Gamma \] \( \cdots \) (2.38)
say. From part (i) of Lemma 1 it follows that $\Gamma$ contains at least $0$. But $\Gamma$ may contain more than one point. For example, if $h(x)$ is a normal density:

$$h(x) = \frac{e^{-\frac{1}{2}(x-\mu)^2}}{(\sqrt{2\pi})^d}$$

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

Next we observe from the Lemma's 1 and 2 and from (2.33) and (2.37) that

$$\text{plim} \{q_2(y_n) - q_2(x_n)\} = 0,$$

$$\text{plim} \{q_2(y_n^*) - q_2(x_n^*)\} = 0,$$

$$\lim_{n \to \infty} q_2(x_n^*) = - \int_{\mathcal{X}^d} h(x)^2 dx,$$

$$q_2(y) \geq - \int_{\mathcal{X}^d} h(x)^2 dx \text{ for } \gamma \in [0, 1]$$

and

$$\tilde{g}_n(y_n) < \tilde{g}_n(x_n^*) \text{ a.s.}$$

Using these results it is not hard to prove

Lemma 3: Under the conditions of Theorem 3,

$$\text{plim} \int_{\mathcal{X}^d} \{h(x|y_n) - h(x)\}^2 dx = 0,$$

where $(y_n)$ is defined by (2.32).

We show now that this lemma implies

$$\text{plim} \sup_{x \in \mathcal{X}^d} |h(x|y_n) - h(x)| = 0. \quad \text{(2.39)}$$

The sequence $(y_n)$ is a sequence of numbers in a compact interval $[0, 1]$ and it therefore has at least one limit point $y_\ast$, say, in $[0, 1]$. This limit point must be a point of $\Gamma$ otherwise Lemma 3 cannot be true and the same applies to other limit points of $(y_n)$. Thus all the limit points of $(y_n)$ are contained in $\Gamma$. This implies that every subsequence $(y_{n_m})$ contains a further subsequence $(y_{n_{m_j}})$ such that for some point $y_\ast \in \Gamma$, $\lim_{j \to \infty} y_{n_{m_j}} - y_\ast = 0$ and consequently by the uniform continuity of $h(x|\gamma)$ on $\mathcal{X}^d \times [0, 1]$, and by (2.36),

$$\lim_{j \to \infty} \sup_{x \in \mathcal{X}^d} |h(x|y_{n_{m_j}}) - h(x|y_\ast)| = \lim_{j \to \infty} \sup_{x \in \mathcal{X}^d} |h(x|y_{n_{m_j}}) - h(x|y_\ast)| = 0. \quad \text{(2.40)}$$
In its turn (2.40) implies (2.39), as is not hard to verify. Combining (2.39) with Lemma 1 we now conclude

Theorem 5: Let the conditions of Theorem 3 be satisfied. The sequence \((\gamma_n)\) obtained from (2.32) is such that

\[
\lim_{n \to \infty} \sup_{x \in \mathcal{F}^k} |\tilde{h}_n(x|\gamma_n) - h(x)| = 0.
\]

3. REGRESSION

3.1. SMINK estimators of a regression function. Consider an i.i.d. stochastic process \((y_1, x_1), \ldots, (y_n, x_n), \ldots \) in \(\mathcal{F} \times \mathcal{F}^k\), where the distribution of the \((y_i)\)'s is absolutely continuous with continuous density \(f(y, x)\), say. Of course then the distribution of the \(x_i\)'s is continuous too with continuous density

\[
h(x) = \int_{-\infty}^{+\infty} f(y, x) dy,
\]

say. As is well known, the expectation of \(y_1\) conditional on the event \(x_1 = x (x \in \mathcal{F}^k)\) is

\[
E(y_1|x_1 = x) = \int_{-\infty}^{+\infty} yf(y, x) dy/h(x) = g(x)
\]

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

\[
\hat{g}_n(x) = \int_{-\infty}^{+\infty} y\hat{f}_n(y, x) dy/\hat{h}_n(x)
\]

where \(\hat{f}_n(y, x)\) and \(\hat{h}_n(x)\) 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

\[
\tilde{g}_n(x|\gamma_1, \gamma_2) = \int_{-\infty}^{+\infty} y\tilde{f}_n(y, x|\gamma_1) dy/\tilde{h}_n(x|\gamma_2)
\]

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 \( \tilde{f}_n(y, x | \gamma) \) of \( f(y, x) \). Similar to (2.23) and (2.24) we have
\[
\tilde{f}_n(y, x | \gamma) = \frac{1}{n} \sum_{i=1}^{n} f(y_i, x_i | \gamma) 
\]
\[
\text{where}
\]
\[
\tilde{f}_n(y, x | \gamma) = \exp \left\{ -\frac{1}{2} \left[ (y, x') - \sqrt{1 - \gamma} (y_i, x_i') - (1 - \sqrt{1 - \gamma})(\bar{y}, \bar{x}') \right] \right. 
\times V_{y,x}^{-1} \left[ \begin{pmatrix} y \\ x \end{pmatrix} - \sqrt{1 - \gamma} \begin{pmatrix} y_i \\ x_i \end{pmatrix} - (1 - \sqrt{1 - \gamma}) \begin{pmatrix} \bar{y} \\ \bar{x} \end{pmatrix} \right] / \gamma^2 \left. \right\}^{\frac{1}{2}} 
\times \left\{ (\sqrt{2n})^{k+1} \gamma^{k+1} \right\} V_{y,x} \left( y, x \right) 
\]
(3.4)
with \( \bar{y} \) the sample mean of the \( y_i \)'s, \( \bar{x} \) the sample mean of the \( x_i \)'s and \( V_{y,x} \) the sample variance matrix of the vectors \( \left( y_i \right) \). Using formula (2) at page 28 of Anderson (1958) it is not too hard to verify that
\[
\int_{-\infty}^{\infty} y \tilde{f}_n(y, x | \gamma) \, dy = 
\]
\[
(\sqrt{1 - \gamma} y + (1 - \sqrt{1 - \gamma}) \bar{y} + \beta_n(x - \bar{x} + \sqrt{1 - \gamma} x_i + (1 - \sqrt{1 - \gamma}) \bar{x})) \tilde{h}_n(y | \gamma) 
\]
(3.5)
where \( \beta_n \) is the usual OLS estimator of the regression coefficients of a linear model : 
\[
\beta_n = \left[ \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \right]^{-1} \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}), 
\]
(3.6)
and \( \tilde{h}_n(y | \gamma) \) is defined by (2.24). Combining (3.3) and (3.5) we obtain
\[
\int_{-\infty}^{\infty} y \tilde{f}_n(y, x | \gamma) \, dy 
\]
\[
= \sqrt{1 - \gamma} \frac{1}{n} \sum_{i=1}^{n} (y_i - \beta_n x_i - \beta_n) \tilde{h}_n(x | \gamma) + (\beta_n x + \beta_n) \tilde{h}_n(x | \gamma) 
\]
(3.7)
where \( \beta_n \) is the usual OLS estimator of the constant term of a linear regression model : 
\[
\beta_n = \bar{y} - \beta_n \bar{x}.
\]
(3.8)
Dividing (3.7) (where $\gamma$ is replaced by $\gamma_1$) by $\hat{h}_m(x \mid \gamma_2)$ yields the following explicit expression of the SMINK estimator of a regression function

$$
\tilde{g}_m(x \mid \gamma_1, \gamma_2) = \frac{(\beta_n'x + \theta_m)\bar{h}_m(x \mid \gamma_1) + \sqrt{1 - \gamma_1^2} \frac{1}{n} \sum_{j=1}^{n} (y_j - \beta_n'x_j - \theta_m)\tilde{h}_n,j(x \mid \gamma_1)}{\tilde{h}_m(x \mid \gamma_2)} \quad ... (3.9)
$$

If $\tilde{f}_m(y, x \mid \gamma)$ and $\tilde{h}_m(x \mid \gamma)$ are modified SMINK density estimators then it follows

$$
\tilde{g}_m(x \mid \gamma_1, \gamma_2) = \frac{\int yf_m(y, x \mid \gamma_1)dy}{\tilde{h}_m(x \mid \gamma_2)}
$$

$$
= \frac{(\beta_n'x + \theta_m)\bar{h}_m(x \mid \gamma_1) + \sqrt{1 - \gamma_2^2} \frac{1}{n} \sum_{j=1}^{n} (y_j - \beta_n'x_j - \theta_m)\tilde{h}_n,j(x \mid \gamma_1)}{\tilde{h}_m(x \mid \gamma_2)} \quad ... (3.10)
$$

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

$$
\tilde{g}_m(x \mid 1, 1) = \tilde{g}_m(x \mid 1, 1) = \beta_n'x + \theta_m. \quad ... (3.11)
$$

This corresponds with the fact that the SMINK estimator of a density equals the maximum likelihood estimator of a normal density if $\gamma = 1$.

3.2. How to choose the window width parameters. In Section 2.4 we have proposed an integrated square error approach for choosing the window width parameter $\gamma$. The argument involved also applies to the problem of choosing $\gamma_2$, for the denominator of (3.9) is a SMINK density estimator which should be fitted as close as possible to $h(x)$. Therefore we propose to choose $\gamma_2$ by the procedure described in Section 2.4 and the resulting optimal window width will be denoted by $\gamma_2^{O}$. For choosing $\gamma_1$ we propose a weighted least squares procedure, motivated by the following theorem.

Theorem 6: Let $(y_1, x_1), ..., (y_n, x_n), ...$ be an i.i.d. stochastic process in $\mathbb{R}^2$. Assume in addition to the conditions of Theorem 2 that the distribution of the $(y_1, x_1)$ is absolutely continuous, that $g(x)h(x)$ is continuous and absolutely integrable, where $g(x) = E[y_1 \mid x_1 = x]$ and $h(x)$ is the density of the
that the sample covariance of the \( y_j \)'s and the \( x_j \)'s converges in probability and that \( E y_j^2 < \infty \). Let the sequence \((\xi_n)\) be defined as in Theorem 2. Put

\[
\Tilde{Q}_n(\gamma) = \frac{1}{n} \sum_{j=1}^{n} \{ y_j - \Tilde{y}_n(x_j \mid \gamma; \gamma^{(2)}) \}^2 \delta_n(x_j \mid \gamma^{(2)}). \tag{3.12}
\]

If we choose \( \gamma^{(2)} \) in \([\xi_n, 1]\) such that

\[
\Tilde{Q}_n(\gamma^{(2)}) = \inf_{\gamma \in [\xi_n, 1]} \Tilde{Q}_n(\gamma) \tag{3.13}
\]

then

\[
\limsup_{\delta \to 0} \sup_{x \in \mathbb{R}^2 \setminus \mathbb{R}^2 \cap h(x) \geq \delta} | \Tilde{Q}_n(x \mid \gamma^{(1)}, \gamma^{(2)}) - g(x) | = 0
\]

for every \( \epsilon (0, \sup_{x \in \mathbb{R}^2} h(x)) \). Moreover, the same result holds for the modified \( \mathbf{SMINK} \) regression estimator.

**Proof:** Appendix.

4. A NUMERICAL EXAMPLE

We now will demonstrate the performance of the \( \mathbf{SMINK} \) estimators on an artificial data set of size 100. In the regression case to be considered we employed the model

\[
y_j = x_{1,j}^2 + x_{2,j}^2 + u_j, \quad j = 1, 2, \ldots, 100, \tag{4.1}
\]

so that the regression function to be estimated is

\[
g(x_1, x_2) = x_1^2 + x_2^2. \tag{4.2}
\]

We have drawn the \( x_{1,j} \) and \( x_{2,j} \) independently from a mixture of the normal distribution \( \mathcal{N}(-2, 1) \) and \( \mathcal{N}(2, 1) \) with equal weights, and the \( u_j \) have been drawn from the normal distribution \( \mathcal{N}(0, 1) \). So the density of the distribution of the pairs \((x_{1,j}, x_{2,j})\) is

\[
h(x_1, x_2) = \frac{1}{\sqrt{2\pi}} \left\{ \frac{1}{2} \exp \left[ -\frac{1}{2} (x_1-2)^2 + x_2^2 \right] + \frac{1}{2} \exp \left[ -\frac{1}{2} (x_1+2)^2 + x_2^2 \right] \right\}
\]

\[
\times \left\{ \frac{1}{2} \exp \left[ -\frac{1}{2} (x_1+2)^2 + x_2^2 \right] + \frac{1}{2} \exp \left[ -\frac{1}{2} (x_1-2)^2 + x_2^2 \right] \right\} \tag{4.3}
\]
Obviously this density has four modes, namely at \((-2, -2), (-2, 2), (2, -2)\) and \((2, 2)\). Applying the integrated square error method (Section 2.4) and the weighted least squares method (Section 3.2) we found the following window width parameters

\[ \gamma_a^{(1)} = 31, \quad \gamma_a^{(2)} = 33. \]

In the figures A and B we show the true density \((4.3)\) and its SMINK estimate, respectively, and similarly in the figures C and D we show the true regression function \((4.2)\) and its SMINK estimate, where in all cases \(a_1\) and \(a_2\) go from \(-3.5\) to \(3.5\).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{true_density}
\caption{True density \(E(a_1, a_2)\)}
\end{figure}

Comparing the figures A and B 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 D is also flatter than the true regression function in figure C.
SMINK estimate of $H(x_1, x_2)$

Figure B

True regression function $G(x_1, x_2)$

Figure C
especially at the bottom and at the edges. These spots, however, have
relative low values of the density $\hat{h}$, as one can observe from figure A. Since

\begin{center}
SMINK estimate of $G(\hat{x}_1, \hat{x}_2)$
\end{center}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figureD.png}
\caption{Figure D}
\end{figure}

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 $\hat{h}(x)$.

In view of the small sample size of 100, the results are not unsatisfactory.
Of course, definite conclusions about the general performance of the SMINK
estimation approach cannot be drawn from these results, but that was not
our aim. The purpose of this numerical example is to show that our approach
has practical significance. In our opinion the above results provide sufficient
evidence for that.
Appendix

For proving the previous asymptotic results we need the following lemma's.

Lemma A : Let \( (y_n, x_1), \ldots, (y_n, x_n), \ldots \) be an i.i.d. stochastic process in \( \mathcal{R}^2 \). Assume that the distribution of the \( x_2 \) is absolutely continuous with continuous density \( h(x) \), say, and that \( g(x)h(x) \) is continuous and absolutely integrable on \( \mathcal{R}^2 \), where \( g(x) = \mathcal{E}(y_2 | x_2 = x) \). Furthermore we assume

\[
\mathcal{E}y_2^2 < \infty, \quad \ldots \quad (A1)
\]

\( \mathcal{E}x_2 = \mu \) and \( \mathcal{E}x_2' = W \) exist, \quad \ldots \quad (A2)

\( \text{plim} \bar{x} = \mu \) and \( \text{plim} V_n = V (= W - \mu \mu') \) exists, \quad \ldots \quad (A3)

where \( \bar{x} \) is the sample mean and \( V_n \) the sample variance of the \( x \) and \( V \) is positive definite. \quad \ldots \quad (A4)

Put

\[
\bar{a}_n(x | \gamma) = \frac{1}{n} \sum_{j=1}^{n} y_j \bar{h}_n(x | \gamma), \quad \ldots \quad (A5)
\]

where \( \bar{h}_n(x | \gamma) \) is defined by (2.24), and let

\[
a(x | \gamma) = \frac{\int \gamma(x, e^{-\frac{1}{2}y_2(1-y_2)(1-\gamma^2)\mu} - \sqrt{1-\gamma^2}y_2(1-\sqrt{1-\gamma^2})\mu |y_2| \sqrt{\frac{1}{2n}\gamma} |y_2| || |y_2| |y_2| \frac{h(x)}{dx} \bigg| V \bigg|}{(\sqrt{\frac{1}{2n}\gamma} |y_2| || |y_2| |y_2| \frac{h(x)}{dx} \bigg| V \bigg|)} \quad \ldots \quad (A6)
\]

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

\[
\text{lim} \xi_n \sqrt{n} = \infty \quad \text{and} \quad \text{lim} \xi_n = 0, \quad \ldots \quad (A7)
\]

Then

\[
\text{plim} \sup_{\gamma \in [\xi_n, 1]} \sup_{x \in \mathcal{R}^k} |\bar{a}_n(x | \gamma) - a(x | \gamma)| = 0, \quad \ldots \quad (A8)
\]

Moreover, \( a(x | \gamma) \) is a uniformly bounded and uniformly continuous function on \( \mathcal{R}^k \times [0, 1] \) satisfying

\[
\lim_{\gamma \downarrow 0} \sup_{x \in \mathcal{R}^k} |a(x | \gamma) - g(x)h(x)| = 0. \quad \ldots \quad (A9)
\]

Proof : From the inversion formula for characteristic functions it follows that \( \tilde{h}_n, g(x | \gamma) \) can be written as

\[
\tilde{h}_n, g(x | \gamma) = \left( \frac{1}{2\pi} \right)^k \int_{\mathcal{R}^k} e^{-xit - \bar{x}it} \sqrt{1-\gamma^2} x(1-\sqrt{1-\gamma^2}) e^{-\gamma y_2y_2'} d\gamma dt, \quad \ldots \quad (A10)
\]
hence
\[ a_n(x | \gamma) = \left( \frac{1}{2\pi} \right)^k \int_{\mathcal{K}^k} \left\{ \frac{1}{n} \sum_{j=1}^{n} y_{j} e^{i z_{j} \sqrt{1 - \gamma^2}} \right\} e^{-i u^2 x^2 - \gamma^2 (1 - \sqrt{1 - \gamma^2})} e^{-\frac{t^2}{1 - \gamma^2} V d t} dt. \]  

Put
\[ a_n(x | \gamma) = \left( \frac{1}{2\pi} \right)^k \int_{\mathcal{K}^k} \left\{ \frac{1}{n} \sum_{j=1}^{n} E_{y_{j}} e^{i z_{j} \sqrt{1 - \gamma^2}} \right\} e^{-i u^2 x^2 - \gamma^2 (1 - \sqrt{1 - \gamma^2})} e^{-\frac{t^2}{1 - \gamma^2} V d t} dt. \]  

We shall prove (A8) now in four steps. Assuming that \((\xi_n)\) is monotone and choosing \(\varepsilon \in (0, 1 - \xi_n)\) for all \(n\) it will be shown that
\[ \text{plim sup}_{\gamma \in [\min 1 - \varepsilon]} \sup_{x \in \mathcal{K}^k} |\tilde{a}_n(x | \gamma) - a_n^*(x | \gamma)| = 0; \]  
\[ \text{lim sup}_{\varepsilon \downarrow 0} \sup_{\gamma \in [1 - \varepsilon, 1]} \sup_{x \in \mathcal{K}^k} |\tilde{a}_n(x | \gamma) - a_n^*(x | \gamma)| = 0; \]  
\[ \text{plim sup}_{\gamma \downarrow 1 - \varepsilon} \sup_{x \in \mathcal{K}^k} |a_n^*(x | \gamma) - a(x | \gamma)| = 0; \]  
\[ \text{plim sup}_{\gamma \rightarrow 0} \sup_{x \in \mathcal{K}^k} |a_n^*(x | \gamma) - a(x | \gamma)| = 0. \]

and finally
\[ \text{plim sup}_{\gamma \rightarrow 0} \sup_{x \in \mathcal{K}^k} |a_n^*(x | \gamma) - a(x | \gamma)| = 0. \]

Obviously (A13) through (A16) imply (A8).

Step 1: Proof of (A13). From (A11) and (A12) we have
\[ \sup_{x \in \mathcal{K}^k} |\tilde{a}_n(x | \gamma) - a_n^*(x | \gamma)| \leq \left( \frac{1}{2\pi} \right)^k \int_{\mathcal{K}^k} \left| \frac{1}{n} \sum_{j=1}^{n} y_{j} e^{i z_{j} \sqrt{1 - \gamma^2}} - E_{y_{j}} e^{i z_{j} \sqrt{1 - \gamma^2}} \right| e^{-\frac{t^2}{1 - \gamma^2} V d t} dt \]
\[ = \left( \frac{1}{2\pi} \right)^k \left( \frac{1}{\sqrt{1 - \gamma^2}} \right)^k \int_{\mathcal{K}^k} \left| \frac{1}{n} \sum_{j=1}^{n} (y_{j} e^{i z_{j} \sqrt{1 - \gamma^2}} - E_{y_{j}} e^{i z_{j} \sqrt{1 - \gamma^2}}) \right| e^{-\frac{t^2}{1 - \gamma^2} V d t} dt =: b_n(\gamma) \]  

say. Now let \(S\) be the set of all positive definite \(k \times k\) matrices with minimum eigenvalue larger than \(\sigma_0 > 0\), where \(\sigma_0\) is such that \(V \in S\). If \(V \in S\) then \(b_n(\gamma)\) is bounded from above by
\[ \tilde{b}_n(\gamma) = \left( \frac{1}{2\pi} \right)^k \left( \frac{1}{\sqrt{1 - \gamma^2}} \right)^k \int_{\mathcal{K}^k} \left| \frac{1}{n} \sum_{j=1}^{n} (y_{j} e^{i z_{j} \sqrt{1 - \gamma^2}} - E_{y_{j}} e^{i z_{j} \sqrt{1 - \gamma^2}}) \right| e^{-\frac{t^2}{1 - \gamma^2} \sigma_0 V d t} dt \]  

... (A18)
because then \( t^2 V_t \geq \sigma_t^2 t \). But

\[
\begin{align*}
\sup_{t \in \mathcal{T}_n} E \left| \frac{1}{n} \sum_{j=1}^{n} (y_{j+} \tilde{w}_{j} - E y_{j+} \tilde{w}_{j}) \right|^2 & \\
& \leq \sup_{t \in \mathcal{T}_n} \left\{ E \left[ \frac{1}{n} \sum_{j=1}^{n} (y_j \cos(t' x_j) - E y \cos(t' x_j)) \right]^2 + E \left[ \frac{1}{n} \sum_{j=1}^{n} (y_j \sin(t' x_j) - E y \sin(t' x_j)) \right]^2 \right\} = O(1/\sqrt{n}) \to 0, \quad \ldots \ (A19)
\end{align*}
\]

hence

\[
\begin{align*}
E \sup_{\gamma \in [\theta, 1-\varepsilon]} \tilde{b}_n(\gamma) & \\
& \leq \left( \frac{1}{2\pi} \right)^k \left( \frac{1}{\sqrt{1 - (1-\varepsilon)^2}} \right)^k \int_{\mathcal{T}_n} \left| \frac{1}{n} \sum_{j=1}^{n} (y_{j+} \tilde{w}_{j} - E y_{j+} \tilde{w}_{j}) \right| e^{-\frac{1}{2n} \sigma^2 \gamma^2 t^2} dt \\
& = O \left( \frac{1}{\sqrt{n}} \right) \to 0, \quad \ldots \ (A20)
\end{align*}
\]

where the last conclusion follows from condition (A7). Consequently by Chebyshev's inequality,

\[
\text{plim}_{n \to \infty} \sup_{\gamma \in [\theta, 1-\varepsilon]} \tilde{b}_n(\gamma) = 0 \quad \ldots \ (A21)
\]

Since (A3) and (A4) imply

\[
\lim_{n \to \infty} P \left( V_n \in S \right) = 1,
\]

and since \( b_n(\gamma) \leq \tilde{b}_n(\gamma) \) if \( V_n \in S \), (A21) implies (A13).

Step 2: Proof of (A14). Observe from (A18) that

\[
\begin{align*}
\tilde{b}_n(\gamma) & = \left( \frac{1}{2\pi} \right)^k \int_{\mathcal{T}_n} \left| \frac{1}{n} \sum_{j=1}^{n} (y_{j+} \tilde{w}_{j} \sqrt{1-\gamma^2} - E y_{j+} \tilde{w}_{j} \sqrt{1-\gamma^2}) \right| e^{-\frac{1}{2n} \sigma^2 \gamma^2 t^2} dt \\
& \leq \left( \frac{1}{2\pi} \right)^k \int_{\mathcal{T}_n} \left| \frac{1}{n} \sum_{j=1}^{n} (y_{j+} \tilde{w}_{j} \sqrt{1-\gamma^2} - 1 - E y_{j+} \tilde{w}_{j} \sqrt{1-\gamma^2} - 1) \right| e^{-\frac{1}{2n} \sigma^2 \gamma^2 t^2} dt \\
& + \left( \frac{1}{2\pi} \right)^k \left| \frac{1}{n} \sum_{j=1}^{n} (y_{j+} - E y_{j+}) \right| e^{-\frac{1}{2n} \sigma^2 \gamma^2 t^2} dt
\end{align*}
\]
\[\begin{align*}
&\leq \left(\frac{1}{2\pi}\right)^k \frac{1}{n} \sum_{j=1}^n |y_j| |x_j| \sqrt{1 - \gamma^2} \int_{\mathbb{R}^k} |t| e^{-\frac{1}{2} \sigma^2 t^2} dt \\
&+ \left(\frac{1}{2n}\right)^k \frac{1}{n} \sum_{j=1}^n E |y_j| |x_j| \sqrt{1 - \gamma^2} \int_{\mathbb{R}^k} |t| e^{-\frac{1}{2} \sigma^2 t^2} dt \\
&+ \left(\frac{1}{2n}\right)^k \left| \frac{1}{n} \sum_{j=1}^n (y_j - E y_j) \right| \int_{\mathbb{R}^k} e^{-\frac{1}{2} \sigma^2 t^2} dt, \quad \text{(A22)}
\end{align*}\]

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

\[E \sup_{\gamma \in \mathbb{R}^k} \hat{b}_n(\gamma) \leq \left(\frac{1}{2\pi}\right)^k 2E |y_j| |x_j| |y_j - E y_j| \sqrt{1 - (1 - \epsilon)^2} \int_{\mathbb{R}^k} |t| e^{-\frac{1}{2} \sigma^2 t^2} dt.\]

\[+ \left(\frac{1}{2n}\right)^k E \left| \frac{1}{n} \sum_{j=1}^n (y_j - E y_j) \right| \int_{\mathbb{R}^k} e^{-\frac{1}{2} \sigma^2 t^2} dt \quad \text{(A23)}\]

Since \(E |y_j| |x_j| < \infty\) the first term of the upperbound above converges to zero if \(\epsilon \downarrow 0\). Moreover,

\[E \left| \frac{1}{n} \sum_{j=1}^n (y_j - E y_j) \right| \leq \sqrt{E \left( \frac{1}{n} \sum_{j=1}^n (y_j - E y_j)^2 \right)} \to 0, \quad \text{(A24)}\]

hence the second term converges to zero for \(n \to \infty\). Thus

\[\lim_{\epsilon \downarrow 0} \lim_{n \to \infty} E \sup_{\gamma \in \mathbb{R}^k} \hat{b}_n(\gamma) = 0. \quad \text{(A25)}\]

Using this result and the trivial equality

\[\hat{b}_n(\gamma) = b_n(\gamma)I(V \in S) + b_n(\gamma)I(V \notin S),\]

where \(I(\cdot)\) is the well-known indicator function, (A14) can easily be proved.

**Step 3:** Proof of (A15). From (A12) and the inversion formula for characteristic functions it follows

\[a_n(x|\gamma) = \int_{\mathbb{R}^k} \frac{e^{-\frac{1}{2} \sigma^2 t^2} - (1 - \sqrt{1 - \gamma^2})\gamma^2 |V|^{-1} - (1 - \sqrt{1 - \gamma^2})\gamma^2}{(\sqrt{2\pi})^k \gamma^k |V|^{-1}} \cdot g(z) h(z) \, dz. \quad \text{(A26)}\]

Now put for convenience

\[f(x) = g(x) h(x). \quad \text{(A27)}\]
Then \( a_n^*(x | γ) \) can be written as

\[
a_n^*(x | γ) = \left( \frac{1}{\sqrt{1-γ^2}} \right)^k \int_{\mathbb{R}^k} f \left( \frac{x - (1 - \sqrt{1-γ^2})m - γV^t w}{\sqrt{1-γ^2}} \right) e^{-iw'w} \frac{1}{(\sqrt{2π})^k} dw
\]

and similarly we can write

\[
a(x | γ) = \left( \frac{1}{\sqrt{1-γ^2}} \right)^k \int_{\mathbb{R}^k} f \left( \frac{x - (1 - \sqrt{1-γ^2})μ - γV^t w}{\sqrt{1-γ^2}} \right) e^{-iw'w} \frac{1}{(\sqrt{2π})^k} dw.
\]

Since \( f(x) = g(x) \delta(x) \) is continuous and absolutely integrable on \( \mathbb{R}^k \) it is uniformly continuous on \( \mathbb{R}^k \), hence

\[
\sup_{γ \in [0,1]} \sup_{x \in \mathbb{R}^k} \left| f \left( \frac{x - (1 - \sqrt{1-γ^2})m - γV^t w}{\sqrt{1-γ^2}} \right) \right| \rightarrow 0
\]

in prob. pointwise in \( w \) if

\[
\sup_{γ \in [0,1]} \frac{1 - \sqrt{1-γ^2}}{\sqrt{1-γ^2}} |m - μ| + \sup_{γ \in [0,1]} \frac{γ}{\sqrt{1-γ^2}} |V^t w - V^t v| \rightarrow 0
\]

in prob. pointwise in \( w \). However, condition (A31) follows from condition (A3). Therefore (A15) follows from (A28), (A29) and (A30) by bounded convergence.

**Step 4**: **Proof of (A16).** From the inversion formula for characteristic functions we have

\[
a(x | γ) = \left( \frac{1}{2π} \right)^k \int_{\mathbb{R}^k} \left( \frac{e^{2π(x-a) \cdot \gamma V^t w}}{\sqrt{1-γ^2}} \right) e^{-iw'w} \frac{1}{(\sqrt(2π)^k} dw
dl
\]

hence from (A12) and (A13) it follows

\[
\sup_{x \in \mathbb{R}^k} |a_n^*(x | γ) - a(x | γ)| \leq \left( \frac{1}{2π} \right)^k \left( E |y| \right) \int_{\mathbb{R}^k} \left| e^{2π(x-a) \cdot \gamma V^t w} e^{-iw'w} \right| \frac{1}{(\sqrt(2π)^k} dw
\]

\[
- e^{2π(x-a) \cdot \gamma V^t w} e^{-iw'w} \leq \left( \frac{1}{2π} \right)^k \left( E |y| \right) \int_{\mathbb{R}^k} \left| e^{2π(x-a) \cdot \gamma V^t w} e^{-iw'w} \right| \frac{1}{(\sqrt(2π)^k} dw
\]
\[ + \left( \frac{1}{2\pi} \right)^k (E | y_i |) \int_{\mathbb{R}^k} | t | e^{-i\gamma \theta^T V t} - e^{-i\theta^T V t} dt \]

\[ \leq \left( \frac{1}{2\pi} \right)^k (E | y_i |) \int_{\mathbb{R}^k} | x | e^{-i\gamma \theta^T V t} - e^{-i\theta^T V t} dt \]

\[ + \left( \frac{1}{2\pi} \right)^k (E | y_i |) |1 - \sqrt{1 - \gamma^2}| x - \mu | \int_{\mathbb{R}^k} | t | e^{-i\gamma \theta^T V t} dt. \quad \ldots \quad (A33) \]

For \( \gamma \in [1-\varepsilon, 1] \) and \( V_x \in S \) the first term at the right side of (A33) is bounded by

\[ 2 \left( \frac{1}{2\pi} \right)^k (E | y_i |) \int_{\mathbb{R}^k} \sup_{\gamma \in [1-\varepsilon, 1]} | e^{-i\theta^T V t} - e^{-i\gamma \theta^T V t} | e^{-i\gamma \theta^T V t} dt, \quad \ldots \quad (A34) \]

which converges to zero in probability by condition (A3) and bounded convergence. The convergence in prob. to zero uniformly in \( \gamma \in [1-\varepsilon, 1] \) of the second term follows straightforwardly from condition (A3). Similar as in Step 1, this completes the proof of (A16) and hence of (A8).

We now complete the proof of the lemma under review. First, the uniform boundedness of \( a(x | \gamma) \) on \( \mathbb{K}^k \times [0, 1] \) follows straightforwardly from (A29) and (A32). Next we show that \( a(x | \gamma) \) is uniformly continuously on \( \mathbb{K}^k \times [0, 1] \). Observe from (A29) and the uniform continuity of \( f(x) = g(x)b(x) \) that

\[ \lim_{\varepsilon \to 0} \sup_{x \in \mathbb{K}^k \times [0, 1]} \| a(x_1 | \gamma) - a(x_2 | \gamma) \| = 0 \quad \ldots \quad (A35) \]

and from (A33) we see that

\[ \lim_{\varepsilon \to 0} \sup_{x \in \mathbb{K}^k \times [0, 1]} \sup_{\gamma, \gamma' \in [1-\varepsilon, 1]} | a(x_1 | \gamma) - a(x_2 | \gamma') | = 0, \quad \ldots \quad (A36) \]

hence

\[ \lim_{\varepsilon \to 0} \sup_{x \in \mathbb{K}^k \times [0, 1]} \sup_{\gamma, \gamma' \in [1-\varepsilon, 1]} | a(x_1 | \gamma) - a(x_2 | \gamma') | = 0. \quad \ldots \quad (A37) \]

Moreover, from (A33) we see that for \( \gamma, \gamma' \in [\varepsilon, 1] \)

\[ \sup_{x \in \mathbb{K}^k} | a(x_1 | \gamma) - a(x_1 | \gamma') | \leq \left( \frac{1}{2\pi} \right)^k (E | y_i |) \int_{\mathbb{K}^k} | e^{i\mu (1 - \sqrt{1 - \gamma^2})} - e^{i\mu (1 - \sqrt{1 - \gamma'^2})} | dt \]

\[ \leq \left( \frac{1}{2\pi} \right)^k (E | y_i |) \int_{\mathbb{K}^k} | e^{i\mu (1 - \sqrt{1 - \gamma^2})} - e^{i\mu (1 - \sqrt{1 - \gamma'^2})} | dt \]

\[ \leq \left( \frac{1}{2\pi} \right)^k (E | y_i |) \int_{\mathbb{K}^k} | e^{i\mu (1 - \sqrt{1 - \gamma^2})} - e^{i\mu (1 - \sqrt{1 - \gamma'^2})} | dt \]

\[ \leq \left( \frac{1}{2\pi} \right)^k (E | y_i |) \int_{\mathbb{K}^k} | e^{i\mu (1 - \sqrt{1 - \gamma^2})} - e^{i\mu (1 - \sqrt{1 - \gamma'^2})} | dt \]
\[
+ \left( \frac{1}{2\pi} \right)^k \mathbb{E} \left| y_1 \right| \int_{\mathbb{R}^k} \left| e^{-i\mathbf{y}^T \mathbf{V}_1} - e^{-i\mathbf{y}^T \mathbf{V}_2} \right| \, dt 
\leq \left( \frac{1}{2\pi} \right)^k \mathbb{E} \left| y_1 \right| \left\| \sqrt{1 - \gamma^2} - \sqrt{1 - \gamma^2} \right\|_2 \int \left| t \right| e^{-i\mathbf{y}^T \mathbf{V}_1} \, dt 
+ \left( \frac{1}{2\pi} \right)^k \mathbb{E} \left| y_1 \right| \left( \frac{1}{2} \left| \gamma^2 - \gamma^2 \right| \right) \int \mathbf{V}_1 e^{-i\mathbf{y}^T \mathbf{V}_1} \, dt.
\] ...
(A38)

hence

\[
\lim_{\mathbf{x} \to \mathbf{a}} \sup_{\gamma_1 \in \{0, 1\}, \gamma_2 \in \{0, 1\}} \sup_{\gamma_1 - \gamma_2 \neq 0} \left| a(x|\gamma_1) - a(x|\gamma_2) \right| = 0. \quad (A39)
\]

If (A9) is true, then it follows from (A9) and (A39) that

\[
\lim_{\mathbf{x} \to \mathbf{a}} \sup_{\gamma_1 \in \{0, 1\}, \gamma_2 \in \{0, 1\}, \gamma_1 - \gamma_2 \neq 0} \sup_{\gamma_2} \left| a(x|\gamma_1) - a(x|\gamma_2) \right| = 0. \quad (A40)
\]

Combining (A37) and (A40) we then see that \( a(x|\gamma) \) is uniformly continuous on \( \mathbb{R}^k \times [0, 1] \). So it remains to prove (A9). Observe from (A29) and (A27) that for every \( M > 0 \) and every \( \gamma \in (0, 1) \)

\[
\sup_{\mathbf{x} \in \mathbb{R}^k} \left| a(x|\gamma) - \left( \frac{1}{\sqrt{1 - \gamma^2}} \right)^k \mathbb{E} \left( \frac{x - (1 - \sqrt{1 - \gamma^2}) \mu}{\sqrt{1 - \gamma^2}} \right) \right|
\leq \left( \frac{1}{\sqrt{1 - \gamma^2}} \right)^k \sup_{\mathbf{z} \in \mathbb{R}^k} \sup_{\gamma \in \{0, 1\}} \left| f(x - z) - f(z) \right| \int_{\gamma \sqrt{1 - \gamma^2}}^{\gamma \sqrt{1 - \gamma^2}} \frac{e^{-i\mathbf{w}^T \mathbf{w}}}{(\sqrt{2\pi})^k} \, dw 
+ \left( \frac{1}{\sqrt{1 - \gamma^2}} \right)^k \sup_{\mathbf{x} \in \mathbb{R}^k} \left| f(x) \right| \int_{\gamma \sqrt{1 - \gamma^2}}^{\gamma \sqrt{1 - \gamma^2}} \frac{e^{-i\mathbf{w}^T \mathbf{w}}}{(\sqrt{2\pi})^k} \, dw
\leq \left( \frac{1}{\sqrt{1 - \gamma^2}} \right)^k \left\{ \sup_{\mathbf{z} \in \mathbb{R}^k} \sup_{\gamma \in \{0, 1\}} \left| f(x - z) - f(z) \right| 
+ 2 \sup_{\mathbf{x} \in \mathbb{R}^k} \left| f(x) \right| \int_{\gamma \sqrt{1 - \gamma^2}}^{\gamma \sqrt{1 - \gamma^2}} \frac{e^{-i\mathbf{w}^T \mathbf{w}}}{(\sqrt{2\pi})^k} \, dw \right\} 
\quad (A41)
\]

Choosing \( M = \sqrt{\gamma} \) and letting \( \gamma \to 0 \) we now see that

\[
\lim_{\gamma \to 0} \sup_{\mathbf{x} \in \mathbb{R}^k} \left| a(x|\gamma) - \left( \frac{1}{\sqrt{1 - \gamma^2}} \right)^k \mathbb{E} \left( \frac{x - (1 - \sqrt{1 - \gamma^2}) \mu}{\sqrt{1 - \gamma^2}} \right) \right| = 0.
\quad (A42)
\]
From the uniform continuity and boundedness of \( f(x) \) it easily follows

\[
\lim_{\gamma \to 0} \sup_{x \in \mathbb{R}^k} \left[ \frac{1}{\sqrt{1-\gamma^2}} f\left( \frac{x - (1 - \sqrt{1-\gamma^2})\mu}{\sqrt{1-\gamma^2}} \right) - f(x) \right] = 0. \quad \cdots \quad (A43)
\]

Combining (A42) and (A43) (A9) follows. This completes the proof of Lemma A.

Q.E.D.

Lemma B: Let the conditions of Lemma A be satisfied. Put

\[
\bar{G}_\gamma(x | \gamma) = \frac{1}{n} \sum_{j=1}^{n} (x - \sqrt{1-\gamma^2}x_j - (1 - \sqrt{1-\gamma^2})\mu). \quad \cdots \quad (A44)
\]

and

\[
C(x | \gamma) = E \left\{ \left( x - \sqrt{1-\gamma^2}x_j - (1 - \sqrt{1-\gamma^2})\mu \right)^2 \right\} \times \frac{e^{-\frac{1}{2}x^2 - (1 - \sqrt{1-\gamma^2})x_j} V^{-\frac{1}{2}}_x S(x - \sqrt{1-\gamma^2}x_j - (1 - \sqrt{1-\gamma^2})\mu)^2}{(\sqrt{2\pi})^{k^2} | V_x |^{\frac{1}{2}}} \right\} \quad \cdots \quad (A45)
\]

Then

\[
\text{plim} \sup_{x \in \mathbb{R}^k} \sup_{\gamma \in [0, 1]} \bar{G}_\gamma(x | \gamma) - C(x | \gamma) = 0. \quad \cdots \quad (A46)
\]

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

\[
\lim_{\gamma \to 0} \sup_{x \in \mathbb{R}^k} | C(x | \gamma) | = 0. \quad \cdots \quad (A47)
\]

Proof: If we differentiate (A10) to \( x \) we see from (2.24) that

\[
(\partial / \partial x^i) \bar{G}_\gamma(x | \gamma) = \left( \frac{1}{2\pi} \right)^{k/2} \int_{\mathbb{R}^k} (-i) e^{-\frac{1}{2}x^2 + it^2x_j + (1 - \sqrt{1-\gamma^2})x_j} e^{-\frac{1}{2}t_x^2} dt_x^i \]

\[
= \frac{1}{\gamma^2} V^{-\frac{1}{2}}_x S(x - \sqrt{1-\gamma^2}x_j - (1 - \sqrt{1-\gamma^2})\mu). \quad \cdots \quad (A48)
\]

\[
\times \frac{e^{-\frac{1}{2}x^2 - (1 - \sqrt{1-\gamma^2})x_j} V^{-\frac{1}{2}}_x S(x - \sqrt{1-\gamma^2}x_j - (1 - \sqrt{1-\gamma^2})\mu)^2}{(\sqrt{2\pi})^{k^2} | V_x |^{\frac{1}{2}}} \right\} \quad \cdots \quad (A45)
\]
hence
\[
\overline{G}_n(x | \gamma) = \gamma^r V_x \sum_{j=1}^{n} \overline{G}_n(x | \gamma), \quad \ldots \quad (A49)
\]
where
\[
\overline{G}_n(x | \gamma) = i \left( \frac{1}{2\pi} \right)^k \int_{\mathcal{C}^k} \frac{e^{-i x \cdot \zeta}}{\sqrt{1-\gamma^2}} \, d\zeta.
\]
Now put
\[
\overline{\Omega}_n(x | \gamma) = \frac{1}{n} \sum_{j=1}^{n} \overline{G}_n \left( x | \gamma \right), \quad \ldots \quad (A50)
\]
and
\[
\overline{\Omega}_n^*(x | \gamma) = i \left( \frac{1}{2\pi} \right)^k \int_{\mathcal{C}^k} \left[ \frac{e^{i x \cdot \zeta}}{\sqrt{1-\gamma^2}} \right] \frac{e^{-i x \cdot \zeta}}{\sqrt{1-\gamma^2}} \, d\zeta.
\]
Similar to (A13) and (A14) in the proof of Lemma A it can be shown that
\[
\lim_{n \to \infty} \sup_{\gamma \in [\zeta, 1]} \sup_{x \in \mathcal{C}^k} |\overline{\Omega}_n(x | \gamma) - \Omega_n^*(x | \gamma)| = 0, \quad \ldots \quad (A53)
\]
and
\[
\lim_{n \to \infty} \sup_{\gamma \in [\zeta, 1]} \sup_{x \in \mathcal{C}^k} |\overline{\Omega}_n^*(x | \gamma) - \Omega_n^*(x | \gamma)| = 0, \quad \ldots \quad (A54)
\]
hence
\[
\lim_{n \to \infty} \sup_{\gamma \in [\zeta, 1]} \sup_{x \in \mathcal{C}^k} |\overline{\Omega}_n(x | \gamma) - \Omega_n^*(x | \gamma)| = 0, \quad \ldots \quad (A55)
\]
and consequently
\[
\lim_{n \to \infty} \sup_{\gamma \in [\zeta, 1]} \sup_{x \in \mathcal{C}^k} |\overline{G}_n(x | \gamma) - \Omega_n^*(x | \gamma)| = 0, \quad \ldots \quad (A56)
\]
where
\[
\Omega_n^*(x | \gamma) = \gamma^r V_x \overline{\Omega}_n^*(x | \gamma). \quad \ldots \quad (A57)
\]
Next observe from (A48) and (A52) that similar to (A28),
\[
\Omega_n^*(x | \gamma) = \int_{\mathcal{C}^k} \frac{1}{\gamma^2} V_x \left( x - \sqrt{1-\gamma^2}z - (1-\sqrt{1-\gamma^2})E \right) \times \left( \sqrt{2\pi} \right)^k \gamma^2 \left| V_x \right| \frac{e^{-\frac{1}{2\gamma^2}}} {V_x} \, h(x) \, dz.
\]
\[
= \gamma^r V_x^{-1} \left( \frac{1}{\sqrt{1-\gamma^2}} \right) \int_{\mathcal{C}^k} h \left( \frac{x - \frac{1}{2}(1-\sqrt{1-\gamma^2})E - \gamma V_x z}{\sqrt{1-\gamma^2}} \right) \frac{we^{-iwv}dw}{\sqrt{2\pi}}.
\]
Thus if we put
\[ \Omega(x | \gamma) = \gamma^{-1} V^{-1} \left[ \frac{1}{\sqrt{1 - \gamma^2}} \right]^k \int_{\mathbb{R}^k} h \left[ \frac{2 - (1 - \sqrt{1 - \gamma^2}) \mu - \gamma V^1 \omega}{\sqrt{1 - \gamma^2}} \right] \frac{w e^{-\mid w \mid^2}}{(\sqrt{2\pi})^k} dw \]
then it follows similar to the proof of (A15) that
\[ \text{plim } \sup_{n \to \infty} \sup_{\omega \in \mathbb{R}^k} \left| \gamma V^1 \Omega_n(x | \gamma) - \gamma V^1 \Omega(x | \gamma) \right| = 0. \quad \ldots (A59) \]
Realizing that (A59) can also be written as
\[ \Omega(x | \gamma) = \frac{i}{(2\pi)^k} \int_{\mathbb{R}^k} \left[ Re^{i\omega \cdot x} \sqrt{1 - \gamma^2} \right] e^{-iw \cdot x} \omega \, \left[ 1 - \sqrt{1 - \gamma^2} \right] \frac{1}{i} e^{-i\omega \cdot x} \, d\omega \]
(compare (A52)), we see from (A52), (A61) and the proof of (A61) that
\[ \sup_{\gamma \in [0, 1]} \sup_{x \in \mathbb{R}^k} \left| \Omega_n(x | \gamma) - \Omega(x | \gamma) \right| = 0. \quad \ldots (A62) \]
From (A60) and (A56) we obtain
\[ \text{plim } \sup_{n \to \infty} \sup_{\omega \in \mathbb{R}^k} \left| \Omega_n(x | \gamma) - \gamma^2 V \Omega(x | \gamma) \right| \]
\[ \leq \text{plim } \sup_{n \to \infty} \sup_{\omega \in \mathbb{R}^k} \left| \gamma V^1(x | \gamma) \Omega_n(x | \gamma) - \gamma V^1 \Omega(x | \gamma) \right| \]
\[ + \text{plim } \sup_{n \to \infty} \sup_{\omega \in \mathbb{R}^k} \left| \gamma^2 (V^1 - V^1) \Omega(x | \gamma) \right| = 0 \quad \ldots (A63) \]
because \text{plim } V_n = V and \left| \Omega(x | \gamma) \right| is bounded on \mathbb{R}^k \times [0, 1]. Similarly we conclude from (A62) that
\[ \text{plim } \sup_{n \to \infty} \sup_{\omega \in \mathbb{R}^k} \left| \Omega_n(x | \gamma) - \gamma^2 V \Omega(x | \gamma) \right| = 0 \quad \ldots (A64) \]
and consequently, combining (A63) and (A64),
\[ \text{plim } \sup_{n \to \infty} \sup_{\omega \in \mathbb{R}^k} \left| \Omega_n(x | \gamma) - \gamma^2 V \Omega(x | \gamma) \right| = 0 \quad \ldots (A65) \]
Since \( C(x | \gamma) = \gamma^2 V \Omega(x | \gamma), (A64) \) follows from (A55) and (A65). The rest of the lemma can be proved in the same way as in the proof of Lemma A. Q.E.D.

**Proof of Theorem 2:** Theorem 2 follows from Lemma A by substituting \( y = 1 \).

**Proof of Theorem 4:** Theorem 4 follows straightforwardly from Lemma 1 in Section 2.4 if
\[ \text{plim } \sup_{\omega \in \mathbb{R}^k} \left| \frac{1}{n} \sum_{j=1}^{n} h(x_j | \gamma) - \int h(x | \gamma) h(x) dx \right| = 0, \quad \ldots (A66) \]
where \( h(x|\gamma) \) is defined by (2.36). But \( h(x|\gamma) \) is uniformly continuous and uniformly bounded on \( \mathbb{R}^k \times [0, 1] \), as follows from Lemma A. So (A66) follows from Jennrich (1969, Lemma 2) and hence Theorem 4 follows from Lemma 1.

Proof of Lemma 1: Lemma 1 follows straightforwardly from (3.29) and Lemma A by substituting \( y^* = 1 \). Q.E.D.

Proof of Theorem 6: From the conditions of Theorem 6 it follows that the vector \( \hat{\beta}_n \) of OLS parameter estimators and the OLS estimator \( \hat{\alpha}_n \) of the constant term of a linear regression model converge in probability:

\[
\text{plim } \hat{\alpha}_n = \alpha, \text{ plim } \hat{\beta}_n = \beta,
\]
say. Now observe from (3.5), (A5) and (A44) that

\[
\hat{\gamma}_n(x|\gamma, \gamma_n(x|\gamma)) = \int \frac{f(x|\gamma)}{y} f_n(x|\gamma) dy
\]

\[
= \sqrt{1-\gamma^2} \hat{a}_n(x|\gamma) + (1-\sqrt{1-\gamma^2}) \hat{b}_n(x|\gamma) + \beta \hat{C}_n(x|\gamma)
\]

\[
= \hat{a}_n(x|\gamma), \quad \cdots \quad (A67)
\]
say. Put

\[
d(x|\gamma) = \sqrt{1-\gamma^2} a(x|\gamma) + (1-\sqrt{1-\gamma^2})(x + \beta \mu) b(x|\gamma) + \beta \mu C(x|\gamma), \quad \cdots \quad (A68)
\]

where \( a(x|\gamma), b(x|\gamma) \) and \( C(x|\gamma) \) are defined by (A6), (2.38) and (A45), respectively and \( \mu = \text{plim } \hat{\mu} \). Realizing that \( \text{plim } \hat{y} = \alpha + \beta \mu \) and that \( d(x|\gamma) \) is uniformly bounded on \( \mathbb{R}^k \times [0, 1] \), we conclude from the Lemma's A, 1 and B that

\[
\text{plim } \sup \sup \left| \hat{a}_n(x|\gamma) - d(x|\gamma) \right| = 0. \quad \cdots \quad (A69)
\]

Using this result and the fact that by Theorem 4,

\[
\text{plim } \sup \left| \hat{a}_n(x|\gamma_n(x|\gamma)) - h(x) \right| = 0, \quad \cdots \quad (A70)
\]

it is easy to show

\[
\text{plim } \sup_{x \in \mathbb{R}^k} \left| \frac{1}{n} \sum_{j=1}^{n} \left( y_j \hat{h}_n(x_j|\gamma_n(x_j|\gamma)) - \hat{a}_n(x_j|\gamma) \right)^2 \right| = 0,
\]

\[
-\frac{1}{n} \sum_{j=1}^{n} (y_j h(x_j) - d(x_j|\gamma))^2 = 0. \quad \cdots \quad (A71)
\]
Moreover, from the conditions of the theorem under review and the fact that 
\( d(x|\gamma) \) is uniformly bounded and continuous on \( \mathcal{R}^k \times [0,1] \) it follows that 
Jennrich (1969, Lemma 2) is applicable, hence

\[
\operatorname{plim} \sup_{\gamma \in [0,1]} \left| \frac{1}{n} \sum_{j=1}^{n} (g_jh_n(x_j) - d(x_j|\gamma))^2 - Eh^2_n(x_j) \right| = 0 \quad \text{... (A72)}
\]

where \( u_j = y_j - g(x_j) \). Putting

\[
Q(\gamma) = Eh^2_n(x_j) + \int_{\mathcal{R}^k} \{g(x)h(x) - d(x|\gamma)\}^2h(x)dx, \quad \text{... (A73)}
\]

we have proved by now

\[
\operatorname{plim} \sup_{\gamma \in [0,1]} |Q_n(\gamma) - Q(\gamma)| = 0. \quad \text{... (A74)}
\]

Consequently we have

\[
\operatorname{plim} \int_{\mathcal{R}^k} \{g(x)h(x) - d(x|\gamma^{(n)})\}^2h(x)dx = 0. \quad \text{... (A75)}
\]

Similar as to the proof of Theorem 5 it can now be shown that (A75) implies

\[
\operatorname{plim} \sup_{x \in \mathcal{R}^k: h(x) \neq 0} \left| \tilde{d}_n(x|\gamma^{(n)}) - g(x)h(x) \right| = 0 \quad \text{... (A76)}
\]

for \( \epsilon \leq (0, \sup_{x \in \mathcal{R}^k} h(x)) \). The SMINK regression case considered in Theorem 6

now follows straightforwardly from (A76).

The modified SMINK regression case follows from

\[
\operatorname{plim} \sup_{\gamma \in [0,1]} \sup_{x \in \mathcal{R}^k} \left| \tilde{d}_n(x|\gamma, \gamma_2)h_n(x|\gamma_2) - \tilde{d}_n(x|\gamma, \gamma_2)h_n(x|\gamma_2) \right| = 0. \quad \text{... (A77)}
\]

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

**References**


*Paper received: July, 1982.*