Computational Statistics (1997) 12:109-130

Computational Statistics

Sliced Inversed Regression (SIR): An Appraisal of Small Sample Alternatives to Slicing

Yves Aragon[1] and Jérôme Saracco[2]

[1] GREMAQ - Université Toulouse I
Place Anatole-France - 31042 Toulouse cedex - France
e-mail: aragon@cict.fr
[2] Laboratoire de Statistique et Probabiltés
Université Toulouse III - 31062 Toulouse cedex - France -
e-mail: saracco@cict.fr

Summary

We consider a semi-parametric regression model where the dependent variable depends upon the explanatory variables through a linear combination of these variables: the index. The dependent variable is linked to the index through an unknown function. Sliced Inverse Regression (SIR) (Li-1991, Duan and Li-1991) is concerned with the estimation of the direction of the index. SIR is computationally simple and fast. For small samples, one observes that the number and the position of slices have influence on the estimated direction. In this paper, we suggest some new methods, and obtain their asymptotic properties. We compare the small sample behaviour of the existing and new methods on simulated data.

Keywords: Inverse Regression, Semi-parametric Regression Model, Efficient Computing, Kernel Estimation.

0943-4062/97/1/109-130 $2.50

1 Introduction

In its simplest version, SIR is concerned with the estimation of the direction of the vector parameter β in the single index semiparametric regression model:

$$\left.\begin{array}{rcl} y & = & g(x'\beta, \epsilon), \\ \epsilon & \perp & x \end{array}\right\} \tag{1}$$

where: y is a univariate dependent variable, x a p-dimensional random regressor, and ϵ an error term independent of x. g is called the *link function*, β the *effective dimension reduction* (e.d.r.) direction and $x'\beta$, the *index*. g and β are unknown.

This model has been extensively studied by Li, see e.g. Li [8] for an early reference. Estimation of β relies on the fact that β is related to the principal eigenvector of a rank one matrix, $cov(E(x|T(y)))$, with respect to the inner product $(b_1, b_2) = b_1'\Sigma_x b_2)$ where T is a function of y and $\Sigma_x = cov(x)$. If y is a discrete variable, estimation of $E(x|y)$ is not difficult. This paper is mainly concerned with a continuous y.

Li's suggestion of using a discretization of y as $T()$ is computationally simple and it gives good results for samples not too small. This paper examines several other methods which may be of interest for small sample situations. For each method, asymptotic properties are recalled or established. Their small sample properties are compared on simulated data.

In Section 2 we review some basic properties about the semiparametric regression model, its connection with inverse regression and the structure of the ensuing estimation methods. Section 3 gives a description of the various estimation methods: they are based either on one or several slicings or on the direct estimation of $cov(E(x|y))$ through nonparametric methods. The small sample behavior of these methods is analyzed on simulated data. The results of these simulations are analyzed in Section 4. Technical tools and the proof of asymptotic properties are deferred to the Appendix.

2 Population Properties of Sliced Inverse Regression

2.1 Semiparametric model

The semiparametric model in the introduction can be generalized in two ways. The first generalization is to allow more than one indices:

$$\left.\begin{array}{rcl} y & = & g(x'\beta_1, \cdots, x'\beta_K, \epsilon), \\ \epsilon & \perp & x \end{array}\right\} \tag{2}$$

where $K < p$.

Let Σ_x denote the covariance matrix of x. The identifiability of the directions is guaranteed by the constraints: $\beta_k' \Sigma_x \beta_k = 1$ and $\beta_k' \Sigma_x \beta_l = 0, k \neq l$, which will be enforced hereafter.

Another generalization deals with a q-component dependent variable y, with a specific link function for each component and one set of indices common for all components. Li et al [9] examine specific estimation methods for this multivariate setting. These methods use slicing.

It is often simpler to regress y on the standardized version of x: $z = \Sigma_x^{-\frac{1}{2}}(x - E(x))$. Parameters are then: $\eta_k = \Sigma_x^{+\frac{1}{2}} \beta_k$, $k = 1, \cdots, K$. The β_k's are called the e.d.r. directions while the η_k are called the standardized e.d.r. directions.

2.2 Characterization of the e.d.r. direction

The design condition underlying the fundamental theory of Sliced Inverse Regression is the following:

Design Condition (D.C.) *The regressor x is sampled from a nondegenerate probability distribution such that*

$$E(x'b|x'\beta) = c_0 + c_1 x'\beta \text{ for all } b \in R^p$$

where c_0 and c_1 are scalars.

The condition may hold only for the true β. It holds for all β if and only if x is elliptically distributed. A study of the bias if the D.C. does not hold is given by Duan and Li [3].

A characterization of the e.d.r. direction is given in the following theorem (see [3] for the single index model or [8] for the multiple indices situation).

Theorem *For model (2), under D.C. the inverse curve $E(z \mid T(y))$ falls in the K-dimensional linear subspace of R^p spanned by the standardized e.d.r. directions $\eta_1, \cdots, \eta_K$; specifically, for model (1) we get:*

$$E(z \mid T(y)) = c_T(y)\eta \quad \text{where} \quad c_T(y) = E(z'\eta \mid T(y))$$

For model (1), a straightforward consequence of this theorem is that

$$\Gamma_T := cov(E(z|T(y)))$$

is degenerate in any direction orthogonal to η. It follows that the principal eigenvector (that is the eigenvector corresponding to the non-nul eigenvalue), v, of Γ_T gives a standardized direction. Transforming back to the original scale, $\Sigma_x^{-\frac{1}{2}} v$ is an e.d.r. direction.

The next section reviews a number of estimation methods.

3 Estimation methods

All the methods that we are to examine depend on the eigenvalue decomposition of an estimator of $cov(E(z|T(y)))$. Due to the ANOVA identity:

$$cov(z) = I_p = cov(E(z \mid T(y))) + E(cov(z \mid T(y))),$$

it is equivalent to estimating any of the two quantities on the right hand side.

Hsing and Carroll [5] have studied the asymptotic properties of estimators of $E(cov(z \mid y))$. All other methods considered here deal with estimators of the other right-hand side term. They split into two families: direct methods which nonparametrically estimate $cov(E(z \mid y))$ or slicing based methods which use one or several discretizations of y.
The asymptotic distributions are obtained here for the single index model ($K = 1$), while the consistency results are valid for the multiple indices model. We do not explicitly consider the multivariate y situation, but, with the exception of Hsing and Carroll's approach, the methods discussed here may be used in that setting.

3.1 Slicing and pooled slicing

This subsection is devoted to methods based on one or several slicings.

3.1.1 Basic Sliced Inverse Regression

Let $T(\ )$ be some partition of the range of y into H slices: $I_1, \cdots, I_H$. Then $\Gamma_T = \sum_{h=1}^{H} p_h m_h m_h'$ where $p_h = \Pr(y \in I_h)$ and $m_h = E(z \mid y \in I_h)$.

These quantities are estimated by their empirical version:
$\hat{p}_h = \frac{1}{n}\sum_{i=1}^{n} I\ (y_i \in I_h)$ and $\hat{m}_h = \frac{1}{n\hat{p}_h}\sum_{i=1}^{n} z_i I\ (y_i \in I_h)$, which converge respectively to p_h and m_h.

Then Γ_T is estimated by:

$$\hat{\Gamma}_T = \sum_{h=1}^{H} \hat{p}_h \hat{m}_h \hat{m}_h'.$$

The eigenvector, $\hat{v}$, corresponding to the largest eigenvalue of $\hat{\Gamma}_T$ is the estimated standardized e.d.r. direction.

Asymptotics. It follows that $\hat{\Gamma}_T = \Gamma_T + O_p(1/\sqrt{n})$. Consequently, the estimated standardized e.d.r. direction converges to the principal eigenvector of Γ_T, which is a standardized e.d.r. direction, at the root n rate.
Although the asymptotic distribution of slicing based estimators may be directly obtained (Duan and Li [3]), it will be useful, in particular for the asymptotic study of pooled slicing based estimators in the next section, to relate the problem to a linear least squares estimator. This will be done in the Appendix.

3.1.2 Pooled slicing

When the sample size is small, the number or the position of slices may have substantial influence on the estimate of β. One remedy is to combine the results from a number of slicings.

One can then hope that pooling the covariances of $E(z \mid T(y))$ for a number of step functions $T(\ )$ may regularize the estimated eigenvectors. So let us consider the single index model and let D different ways of partition of y be given: $T_d(\ )$, $d = 1, \cdots, D$, then $\Gamma_d = cov(E(z \mid T_d(y))) = E[c_d(y)^2]\eta\ \eta'$. Let us choose a sequence of positive weights w_d with $\forall d,\ w_d \geq 0$, $\sum_{d=1}^{D} w_d = 1$, and let us define a pooled covariance matrix:

$$\Gamma^P = \sum_{d=1}^{D} w_d \Gamma_d.$$

Since $\Gamma^P = \sum_{d=1}^{D} w_d E[c_d(y)^2]\eta\ \eta'$, this matrix is degenerate in any direction orthogonal to η and the non-nul eigenvector is a standardized e.d.r. direction.

We shall denote as $\hat{v}^P$ the eigenvector corresponding to the largest eigenvalue of $\hat{\Gamma}^P = \sum_{d=1}^{D} \hat{w}_d \hat{\Gamma}_d$, the estimator derived from this pooling when each Γ_d is estimated according to the principle of Basic slicing for the slicing $T_d(\ )$. The choice of the slicings and of the weights is discussed below.

Practical choices

For each slicing T_d, the slices are built such that the numbers of cases in slices never differ by more than one. The number D of slicings is controlled through the minimum number of slices, $H_{\min}$, and the minimum number of cases in each slice, $n_{\min}$, which are given by the user. The parameter $H_{\min}$ must be chosen strictly larger than the number of e.d.r. directions in the model. Starting with a slicing based on $H_{\min}$ slices, the number of slices is incremented while the number of cases in each slice is greater than $n_{\min}$. The lower bound $n_{\min} = 2$ gives a slicing equivalent to the Hsing and Carroll nonoverlapping slices estimator discussed below. We will see in the Simulations section that this estimator behaves poorly. So we suggest to choose $n_{\min} \geq 3$.

For the weights $\hat{w}_d$, we can either choose equal weights or weights proportional to the largest eigenvalue of the $\hat{\Gamma}_d$ in the d-th slicing, $\hat{\lambda}_d$, which is a measure of the performance of that slicing. This suggests using the random weights: $\hat{w}_d = \hat{\lambda}_d \ / \ \sum_{d=1}^{D} \hat{\lambda}_d$.

Asymptotics

CONVERGENCE IN PROBABILITY OF $\hat{\Gamma}^P$. For each slicing,d, we have:

$$\hat{\Gamma}_d = \Gamma_d + O_p(n^{-1/2}) \ \text{ and } \ \hat{\lambda}_d = \lambda_d + O_p(n^{-1/2}),$$

so that:

$$\hat{\Gamma}^P = \Gamma^P + O_p(n^{-1/2}).$$

ASYMPTOTIC DISTRIBUTION OF THE PSIR ESTIMATOR. The asymptotic distribution, given in the Appendix, is derived in two steps. First, the SIR estimator is shown to be nearly the same as the linear least square estimator for the linear regression of a transformation of y, depending on the slicing, against x. Second, stacking the linear least squares problems of each slicing just as in the repeated measurement context of classical linear regression, allows us to obtain the asymptotic distribution of $\hat{v}^P$. The first step is taken from Carroll and Li's work [2] while the second step closely follows the asymptotic study of Pooled Marginal Slicing for Multivariate SIR given in [1].

3.2 Estimation of the expectation of the conditional covariances

The second term of the ANOVA identity is:

$$\Lambda = E(cov(z \mid y))$$

Each element in Λ, $cov(z_i\ ,\ z_j \mid y)$, can be viewed as the local covariance of errors in the regression of components of z on y. It does not need to be consistently estimated; it matters only that the expectation with respect to y gives a consistent estimator of $\Lambda_{i,j}$. In order to keep notations simple, we assume here that the observations have been sorted according to increasing values of y. The two estimators studied by Hsing and Carroll [5] are:

$$\hat{\Lambda} = \frac{1}{n}\sum_{i=1}^{[n/2]}(z_{2i} - z_{2i-1})(z_{2i} - z_{2i-1})' \qquad \text{(Non overlapping slices)}$$

$$\tilde{\Lambda} = \frac{1}{2n}\sum_{i=2}^{n}(z_i - z_{i-1})(z_i - z_{i-1})' \qquad \text{(Running slices)}$$

where $[n/2]$ is the integer part of $n/2$.

The running estimator also appears in the literature on heteroscedastic regression; see Mueller and Stadtmueller [10].

Asymptotics Under appropriate technical hypothesis, Hsing and Carroll show that: $n^{\frac{1}{2}}(\hat{\Lambda} - \Lambda) = O_p(1)$ and $n^{\frac{1}{2}}(\tilde{\Lambda} - \Lambda) = O_p(1)$.

3.3 Nonparametric estimation of the covariance of conditional expectations

Nonparametric techniques may be used to estimate $\Gamma := cov(E(z|y))$ if y is a continuous variable. The construction ot the estimators essentially follows the form of this covariance.

Let z^j be a component of z. Since $E(z) = 0$, each component of Γ is

$$\Gamma_{j,l} = cov\left(\nu^j(y), \nu^l(y)\right) = \int \nu^j(y)\nu^l(y)f(y)dy, \quad j,l = 1,\cdots,p,$$

where $\nu^j(y) = E(z^j|y)$ and f is the density function of y.

The conditional expectation, $\nu^j(y)$, and the density function, f, can be estimated by kernel estimators:

$$\forall y \in I\!R, \quad \hat{\nu}^j(y) = \frac{\sum_{i=1}^n z_i^j K\left(\frac{y-y_i}{h_n^j}\right)}{\sum_{i=1}^n K\left(\frac{y-y_i}{h_n^j}\right)} \quad \text{and} \quad \hat{f}(y) = \frac{1}{nh_n}\sum_{i=1}^n K\left(\frac{y-y_i}{h_n}\right)$$

where $K(.)$ is a kernel. The optimal bandwidths, h_n^j and h_n, may be obtained, for example, through cross-validation.

Substituting in the formula of $\Gamma_{j,l}$ the true functions ν^j, ν^l and f for their kernel estimator $\hat{\nu}^j$, $\hat{\nu}^l$ and $\hat{f}$, we obtain an "integral" estimator:

$$\hat{\Gamma}_{j,l} = \int_{I\!R} \hat{\nu}^j(y)\hat{\nu}^l(y)\hat{f}(y)dy.$$

Alternatively, replacing the integral by a sum, a "sum" estimator, $\widetilde{\Gamma}_{j,l}$, may be used:

$$\widetilde{\Gamma}_{j,l} = \frac{1}{n}\sum_{i=1}^n \hat{\nu}^j(y_i)\hat{\nu}^l(y_i)\, I\left[\hat{f}(y_i) > b_n\right],$$

where the trimming parameter b_n is positive and depends on the sample size n. Our treatment of the estimation of the off-diagonal terms of Γ is inspired by the work of Lavergne [7].

Whatever method is used (sum or integral), $p+1$ smoothing parameters have to be computed. For sum estimators we have to choose the trimming parameter, while for integral estimators, $p(p+1)/2$ integrals have to be numerically evaluated. Thus, these nonparametric methods may be time consuming unless a fast implementation is used (see the simulation section).

Asymptotics Under appropriate technical hypothesis, we show in the appendix that $\hat{\Gamma}$ and $\widetilde{\Gamma}$ converge respectively in quadratic mean and in probability to Γ.

The asymptotic properties that we have reported cannot discriminate between methods. Usually, statistical methods behave well with large samples. So it is worth comparing these methods on small samples. This is the subject of the next section. The results of a number of simulation experiments are therein reported.

4 Simulations

We conducted a simulation study to compare the performance of different estimators. All works were carried out in S-PLUS. We first introduce an efficiency measure to be used. Then we describe the simulated models and comment on the results.

4.1 An efficiency measure

Some notations must be introduced.
Let $N = [\eta_1, ..., \eta_K]$ and $\hat{N} = [\hat{\eta}_1, \cdots, \hat{\eta}_K]$ be the p by K matrix of the true and estimated standardized e.d.r. directions. The standardized e.d.r. space is $E = \text{Span}(N)$, the linear space generated by the η_k's, while $\hat{E} = \text{Span}(\hat{N})$ is the estimated standardized e.d.r. space. Let P_E and $P_{\hat{E}}$ be the orthogonal projectors on E and $\hat{E}$. From the definition of the η_k's and $\hat{\eta}_k$'s: $N'N = \hat{N}'\hat{N} = I_K$. Therefore:

$$P_E = N\ N' \qquad \text{and} \qquad P_{\hat{E}} = \hat{N}\ \hat{N}'.$$

An efficiency measure for the estimates can be defined as:

$$m(\hat{E}, E) \ = \ \text{Tr}(P_{\hat{E}} P_E)/K.$$

This measure takes values in the interval $[0, 1]$. If $\hat{E} = E$, then $m(\hat{E}, E) = 1$. If $\hat{E}$ and E are orthogonal, then $m(\hat{E}, E) = 0$. The closer the measure is to one, the better is the estimation. For $K = 1$, this measure is just the squared cosine of the angle between $\hat{\eta}_1$ and η_1.

4.2 Models for simulation

We consider a single index regression model:

$$y = 1/\left[1 \ + \ (x'\beta_1)^4 \exp\{(x'\beta_1)^3 + (x'\beta_1)\varepsilon\}\right] \qquad \text{Model A}$$

and a two indices model:

$$y = 1/\left[0.5 \ + \ \exp(x'\beta_1 + \varepsilon) \ + \ \exp(x'\beta_2 - \varepsilon)\right] \qquad \text{Model B}$$

We have choosen these models in order to have a bounded y. This boundedness property allows us to monitor the lower and upper limits of bandwidth in the automatic selection step in nonparametric estimation (Haerdle's procedures [4] were used).

The dimension p is set at 5, and x and ε are independent. x follows an elliptically contoured distribution: a p-dimensionnal Pearson Type II distribution (described in Johnson [6]) with density

$$f(x) \ = \ \frac{\Gamma(p/2 + m + 1)}{\Gamma(m+1)\pi^{p/2}} |\Sigma_*|^{-1/2} \left[1 - (x-\mu)'\Sigma_*^{-1}(x-\mu)\right]^m$$

having support $(x-\mu)'\Sigma_*^{-1}(x-\mu) \leq -1$ and shape parameter $m > -1$. Futher, the parameters μ and Σ_* are interpreted as: $E[x] = \mu$ and $cov[x] = \frac{1}{2m+p+2}\Sigma_*$. Here, x is generated by means of the Cambanis representation: $x = RBU^{(p)} + \mu$, where R^2 has a beta distribution with parameters $p/2$ and $m+1$, B is a p by p matrix such that $BB' = \Sigma_*$, and $U^{(p)}$ is a random vector that is uniformly distributed on the unit hypersphere. We take $\mu = (0,0,0,0,0)'$, $\Sigma_* = (2m+p+2)I_5$ and $m = -0.9$. (So $\Sigma_x = I_5$ and $\eta_k = \beta_k$ for $k = 1, ..., K$). ε is standard normally distributed.

We take $\beta_1 = \eta_1 = \frac{1}{2}(1,1,-1,1,0)'$ and $\beta_2 = \eta_2 = \frac{1}{\sqrt{3}}(0,1,1,0,-1)'$.

1000 samples of $n = 25$ data points are generated from model A, and 1000 samples of $n = 50$ data points for model B.

4.3 Results

Acronyms for the competing methods are given Table 1.

SIR	Sliced Inverse Regression
PSIR	Pooled Slicing Inverse Regression
HC1	Hsing and Carroll method with nonoverlapping slices
HC2	Hsing and Carroll method with running slices
K-Sum	Kernel and Sum based estimator
K-Int	Kernel and Integral based estimator

TABLE 1 - *Notations of the different methods.*

For samples of $n = 25$ cases we take 5 slices, and for $n = 50$, 6 slices of 7 cases and one randomly chosen slice of 8 cases. For Pooled Slicing, the minimum number of slices is 3, and the minimum number of cases in each slice is 4.

For the one-component model A, Table 2 gives the mean and standard deviation of the efficiency measure for the different methods, over the 1000 simulated samples of size $n = 25$.

	SIR	PSIR	HC1	HC2	K-Sum	K-Int
Mean	0.693	0.761	0.491	0.680	0.817	0.819
St. Dev.	0.233	0.169	0.303	0.238	0.116	0.114

TABLE 2 - *Mean and standard deviation of the efficiency measure for the different methods in model A.*

As we see from this table and Figure 1, the kernel method based estimates are better than the others estimates. PSIR-method gives also good estimations, while HC1-method appears to be the worst one. The barplots of Figure 2 confirm these observations: the PSIR, K-Sum and K-Int methods rarely provide the worst measure among the six methods.

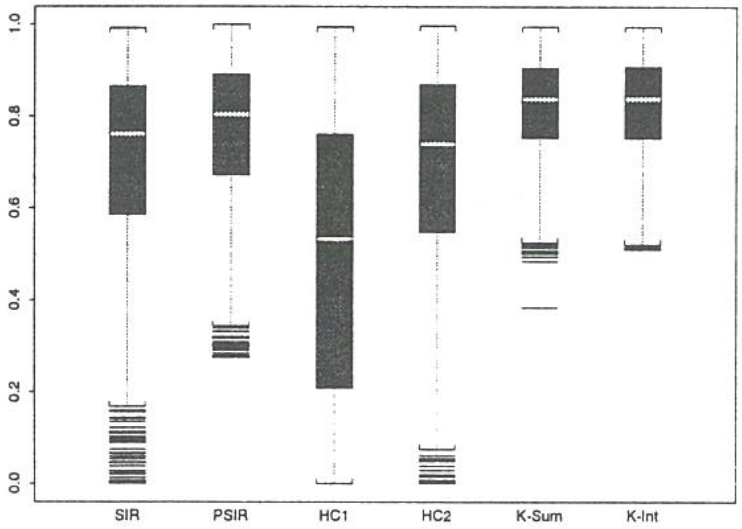

Figure 1: *Model A: Boxplots of the measures (by method).*

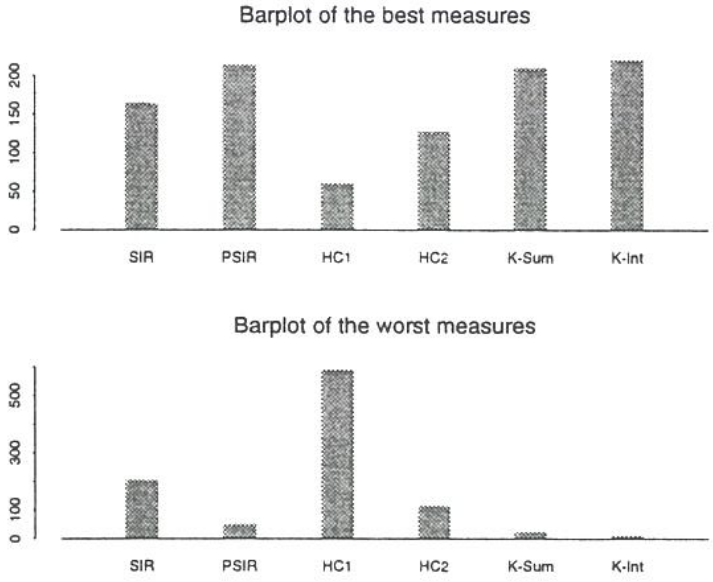

Figure 2: *Model A: Barplot of the best and worst measures (by method).*

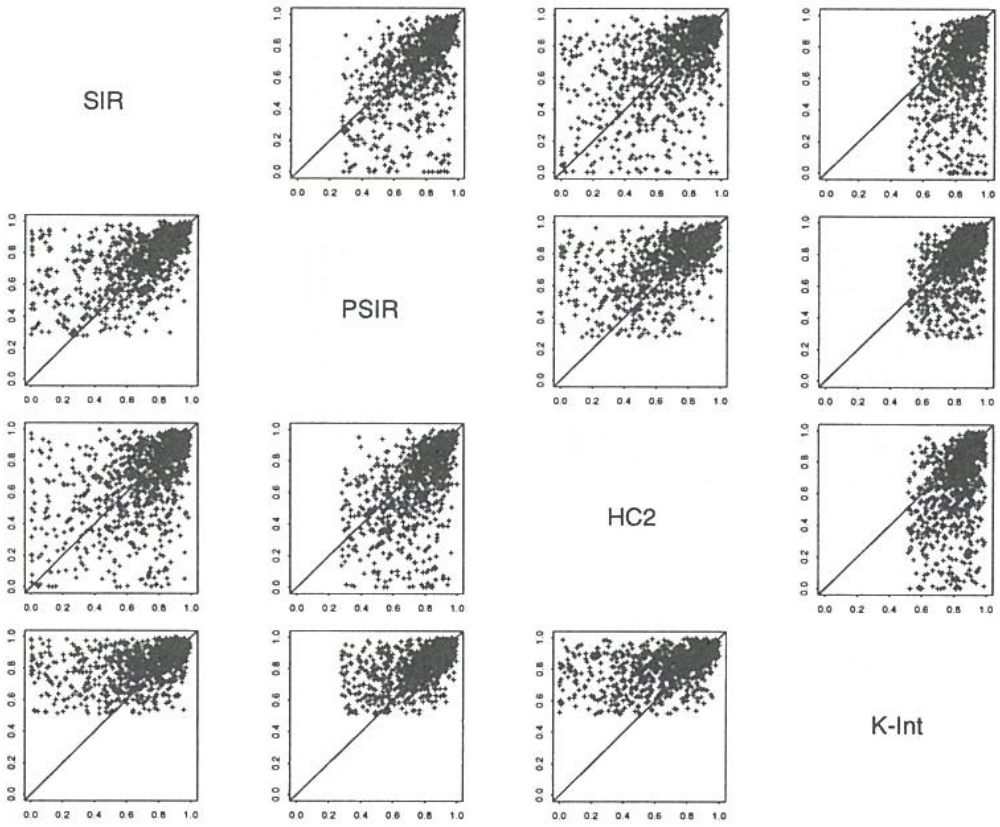

Figure 3: *Model A: Comparison of SIR, PSIR, HC2, K-Int.*

Figures 3 and 4 give scatterplots of simulations of model A for different methods. The coordinates of a point are the efficiency measure of the methods corresponding to the rows and columns. First, we observe on Figure 4 that the K-Sum and K-Int methods often give very similar results, and are uniformly better than the other methods. PSIR-method also seems to be uniformly better than SIR-method.

For the multicomponent model B, Table 3 gives the mean and standard deviation of the efficiency measure for the different methods on the 1000 samples of simulated data.

	SIR	PSIR	HC1	HC2	K-Sum	K-Int
Mean	0.715	0.768	0.691	0.726	0.733	0.735
St. Dev.	0.120	0.106	0.117	0.117	0.086	0.086

TABLE 3 - *Mean and standard deviation of the efficiency measure for the different methods in model B.*

Table 3 and Figures 5, and 6 show that no method accurately finds the true standardized e.d.r. space from samples of size $n = 50$ in model B. However, it is interesting to observe that PSIR-method performs better than the other methods.

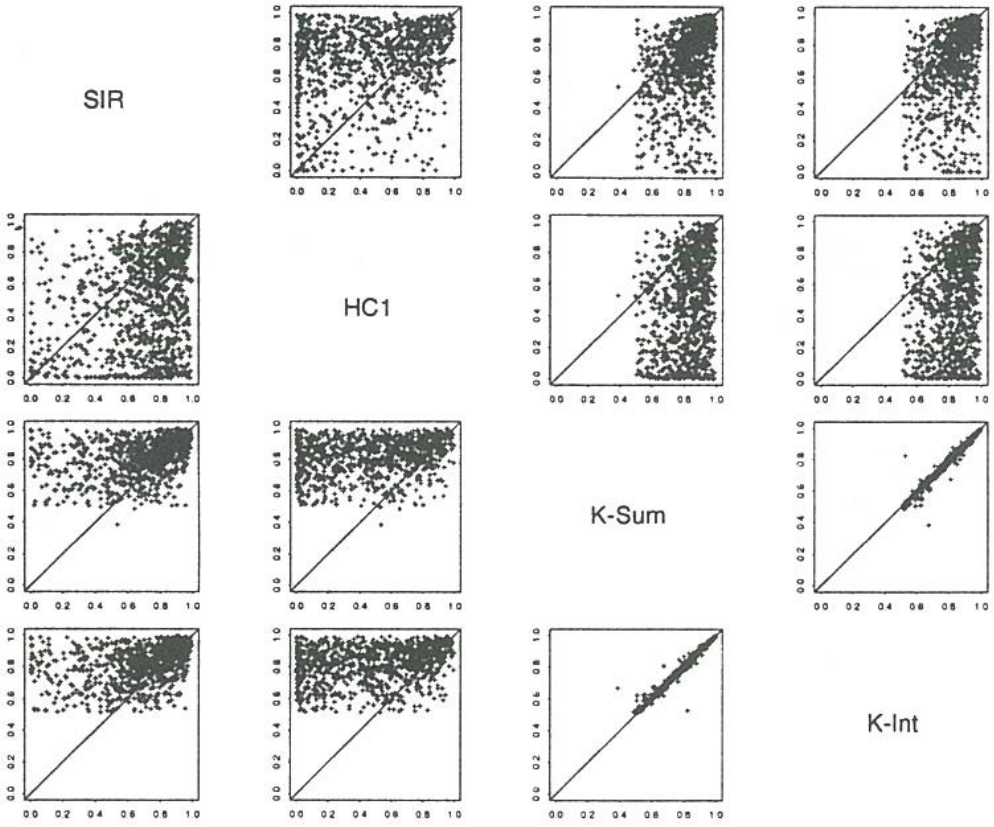

Figure 4: *Model A: Comparison of SIR, HC1, K-Sum, K-Int.*

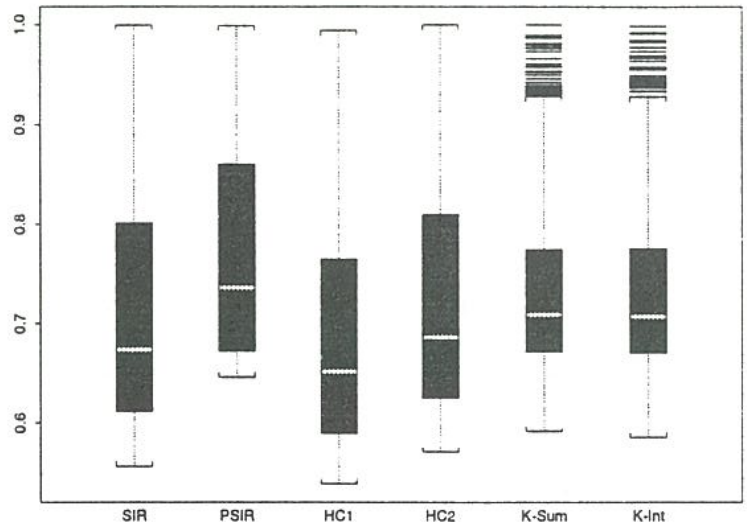

Figure 5: *Model B: Boxplots of the measures (by method).*

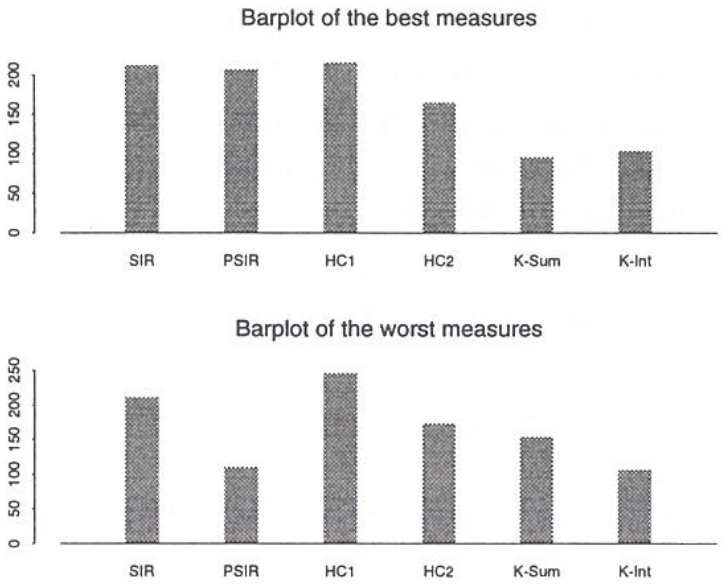

Figure 6: *Model B: Barplot of the best and worst measures (by method).*

5 Conclusion

The new estimation methods that we have introduced appear to perform better than the existing ones. But no one is uniformly better in our simulation study. For a multivariate y, kernel based estimators may suffer the same drawbacks as multivariate nonparametric regression. Pooled slicing is simple. It easily extends to multivariate y and seems to have good numerical properties. The Hsing-Carroll estimator based on running means is simple too and behaves quite well with small samples. But it cannot be extended to multivariate y.

6 Appendix

6.1 Defining a Least Squares Estimator Asymptotically equivalent to the SIR Estimator

6.1.1 Linear regression

The developpement of this subsection is essentially taken from Carroll and Li [2].
The population version of the linear least squares regression of y over z is:

$$\min_{a \in I\!R,\ v \in I\!R^p} E[y - a - v'z]^2.$$

The solution to this problem is, with respect to v:

$$v_{LS} = E(yz), \text{ since } E(z) = 0 \text{ and } cov(z) = I_p.$$

From the study of inverse regression we know that $E(z \mid y) = c(y)\ \eta$. Thus $v_{LS} = E(y\ c(y))\ \eta$, and v_{LS} is proportional to η.
Given a sample of (x, y) or (z, y), the moment estimator of v_{LS} is $\hat{v}_{LS} = 1/n \sum_{i=1}^{n} y_i z_i$. Let us define theoretical residuals: $e = y - Ey - v'_{LS} z$. Carrying this expression into $\hat{v}_{LS}$ we get:

$$\hat{v}_{LS} = v_{LS} + 1/n \sum_{i=1}^{n} e_i z_i + O_p(n^{-1})$$

then, by the Central Limit Theorem:

$$n^{1/2}(\hat{v}_{LS} - v_{LS}) \longrightarrow^{d} N_p(0, cov(ez)).$$

6.1.2 Asymptotic distribution of the SIR estimator

Some matrix notations must be introduced to closely examine the SIR estimator:

$$\begin{aligned}
&\hat{M} = [\hat{m}_1, ..., \hat{m}_H],\\
&M = [m_1, ..., m_H],\\
&\hat{D} = diag(\hat{p}_1, ..., \hat{p}_H),\\
&D = diag(p_1, ..., p_H),\\
&k = (k_1, ..., k_H)' \text{ where } k_h = E[c(y)|y \in I_h].
\end{aligned}$$

It is easy to check that $M = \eta\ k'$ and $\hat{M} = M + O_p(n^{-1/2})$.
After elementary manipulations we get:

$$\hat{\Gamma}_T = \hat{M} \hat{D} \hat{M}' = \tilde{\Gamma} + O_p(n^{-1})$$

with

$$\tilde{\Gamma} = (k' \hat{D} k)^{-1} [\eta + (k' \hat{D} k)^{-1} \hat{\Delta} \hat{D} k][\eta + (k' \hat{D} k)^{-1} \hat{\Delta} \hat{D} k]'$$

where $\hat{\Delta} = \hat{M} - M$.
Let $\hat{v}$ be the non-nul eigenvector of $\hat{\Gamma}_T$ and $\tilde{v}$ the non-nul eigenvector of the rank one matrix $\tilde{\Gamma}$. It is proportional to $[\eta + (k' \hat{D} k)^{-1} \hat{\Delta} \hat{D} k]$ or $[(k' \hat{D} k)\ \eta + \hat{\Delta} \hat{D} k]$. But this vector may be rewritten as: $\sum_{h=1}^{H} \hat{p}_h \hat{m}_h k_h = 1/n \sum_{i=1}^{n} z_i \tilde{y}_i$ with $\tilde{y}_i = \sum_{h=1}^{H} I\ [y_i \in I_h] k_h$.
So, let us write $\tilde{v}_{LS} = \frac{1}{n} \sum_{i=1}^{n} z_i \tilde{y}_i$. This is the empirical version of the linear regression of $\tilde{y}$ over z. The asymptotic distribution of $\tilde{v}_{LS}$ is then derived, using the same steps as in the linear regression of y over z. The theoretical residuals are now: $\tilde{e} = \tilde{y} - E\tilde{y} - v^{*\prime}_{LS} z$ where v^{*}_{LS} is the slope of the linear regression of $\tilde{y}$ over z and is equal to $k' D k\ \eta$.
Ultimately we have to normalize $\tilde{v}_{LS}$. The δ-method gives the asymptotic distribution of the normalized estimator:

$$n^{1/2}\left(\frac{\tilde{v}_{LS}}{||\tilde{v}_{LS}||} - \eta\right) \longrightarrow^{d} N_p(0, (k' D k)^{-2} P cov(\tilde{e} z) P)$$

with $P = I_p - \eta\eta'$.
But, $\tilde{v} = \tilde{v}_{LS}/||\tilde{v}_{LS}||$ is the normalized estimator. Since $\hat{v} = \tilde{v} + O_p(1/n)$, $\hat{v}$ is asymptotically distributed as $\tilde{v}$:

$$n^{1/2}(\hat{v} - \eta) \longrightarrow^d N_p(0, (k'Dk)^{-2} P cov(\tilde{e}z) P).$$

We are now ready to obtain the asymptotic distribution of the Pooled estimator.

3.1.3 Asymptotic distribution of the Pooled estimator

Denoting by an upper index, d, the d-th slicing, and using notations very similar to that introduced in the linear regression development, we get:

$$\hat{\Gamma}^P = \sum_{d=1}^{D} w_d \sum_{h=1}^{H_d} \hat{p}_h^d (\hat{\Delta}_h^d + k_h^d \eta)(\hat{\Delta}_h^d + k_h^d \eta)'$$

where $\hat{\Delta}_h^d = \hat{m}_h^d - m_h^d$.
Let $\hat{v}^P$ be the eigenvector corresponding to the largest eigenvalue of $\hat{\Gamma}^P$.
Let us define the scalar: $\hat{c}^P = \sum_{d=1}^{D} w_d \sum_{h=1}^{H_d} \hat{p}_h^d (k_h^d)^2$
and the vector: $\hat{u} = \sum_{d=1}^{D} w_d \sum_{h=1}^{H_d} \hat{p}_h^d k_h^d \hat{\Delta}_h^d$.
This yields, after some manipulations:

$$\hat{\Gamma}^P = \tilde{\Gamma}^P + O_p(1/n)$$

with

$$\tilde{\Gamma}^P = \hat{c}^P [\eta + (\hat{c}^P)^{-1} \hat{u}][\eta + (\hat{c}^P)^{-1} \hat{u}]'.$$

The principal eigenvector, denoted $\tilde{v}^P$, of this rank one matrix may be viewed as a linear regression estimator, indeed: $\tilde{v}_{LS}^P = [\eta + (\hat{c}^P)^{-1} \hat{u}] = 1/n \sum_{i=1}^{n} z_i \tilde{y}_i^P$ with $\tilde{y}_i^P = \sum_{d=1}^{D} w_d \tilde{y}_i^d$ and $\tilde{y}_i^d = \sum_{h=1}^{H_d} k_h^d I\,[y_i \in I_h^d]$.
The asymptotic distribution of $\tilde{v}_{LS}^P$ is derived in the same way as before. Theoretical residuals are defined as:

$$\tilde{e}^P = \tilde{y}^P - E\tilde{y}^P - v_{LS}^{P\prime} z$$

where

$$v_{LS}^P = \sum_{d=1}^{D} w_d \sum_{h=1}^{H_d} p_h^d (k_h^d)^2 \eta$$

is the slope of the linear regression of $\tilde{y}^P$ over z. With the same normalization as before we ultimately get the asymptotic distribution:

$$n^{1/2}(\hat{v}^P - \eta) \longrightarrow^d N_p(0, (1/n)\ \gamma\ P cov(\tilde{e}z) P)$$

where $\gamma = 1/(\sum_{d=1}^{D} w_d \sum_{h=1}^{H_d} p_h^d (k_h^d)^2)^2$.

6.2 Consistency of the Nonparametric estimators

6.2.1 Some useful notations

In the following proofs, we will denote: for $j = 1, ..., p$,

- $\nu^j(y) = E[z^j|Y = y] = g^j(y)/f(y)$, where $f(.)$ is the density function of Y;
- $\nu_n^j(y) = g_n^j(y)/f_n^j(y)$
  where $g_n^j(y) = \frac{1}{n}h_n^j \sum_{i=1}^n z_i^j K\left(\frac{y_i - y}{h_n^j}\right)$
  and $f_n^j(y) = \frac{1}{nh_n^j} \sum_{i=1}^n K\left(\frac{y_i - y}{h_n^j}\right)$;
- f_n the Nadaraya-Watson estimator of the density function f of Y.

Usually, the functions f_n, $f_n^1, ..., f_n^p$ are not identical. Indeed, smoothing parameters may be different. For instance, with the Cross-validation method, these parameters will be, for a positive bounded function w defined on a compact support:

- $\forall j = 1, ..., p$, $h_n^j = \text{argmin}_h\, CV^j(h)$
  where $CV^j(h) = \frac{1}{n} \sum_{i=1}^n \left[z_i^j - \nu_{n,-i}^j(y_i)\right]^2 w(y_i)$
  and $\nu_{n,-i}^j(y) = \frac{\sum_{l \neq i} z_l^j K\left(\frac{y_l - y}{h}\right)}{\sum_{l \neq i} K\left(\frac{y_l - y}{h}\right)}$;
- $h_n = \text{argmin}_h\, CV(h)$
  where $CV(h) = \int_{\mathbb{R}} [f_n(y)]^2 dy - \frac{2}{n} \sum_{i=1}^n f_{n,-i}(y_i)$,
  and $f_{n,-i}(y) = \frac{1}{nh} \sum_{j \neq i} K\left(\frac{y_j - y}{h}\right)$.

6.2.2 Convergence in probability of $\tilde{\Gamma}_{j,l} = \frac{1}{n} \sum_{i=1}^n \nu_n^j(y_i)\nu_n^l(y_i) \mathbb{1}_{[f_n(y_i) > b_n]}$ to $\Gamma_{j,l} = E\{E[z^j|y]E[z^l|y]\}$

Assumptions:

H_1 : (z_i, y_i) $i = 1, ..., n$ are independent and identically distributed

H_2 : $\lim_{n \to \infty} b_n = 0$ (trimming parameter)

H_3 : $1/b_n \sup_y \left|f_n^j(y) - f(y)\right| \longrightarrow 0$ in probability for $j = 1, ..., p$

H_4 : $1/b_n \sup_y \left|g_n^j(y) - g^j(y)\right| \longrightarrow 0$ in probability for $j = 1, ..., p$

H_5 : $1/b_n \sup_y |f_n(y) - f(y)| \longrightarrow 0$ in probability

H_6 : $E\left[\, |z^j z^l| \,\right] < \infty$ for $j, l = 1, ..., p$

Proof:

Let us define:

$$\bar{\theta} = \frac{1}{n}\sum_{i=1}^{n} \nu_n^j(y_i)\nu_n^l(y_i)I_i \qquad \theta_0 = \frac{1}{n}\sum_{i=1}^{n} \nu^j(y_i)\nu^l(y_i)I_i$$

where $I_i = I_{[f(y_i)>b_n]}$.
The proof splits into three parts.

STEP 1: Let us show that $\bar{\theta} - \theta_0 = o_p(1)$.

$$\bar{\theta} - \theta_0 = \frac{1}{n}\sum_{i=1}^{n} \left(\nu_n^j(y_i)\nu_n^l(y_i) - \nu^j(y_i)\nu^l(y_i)\right) I_i.$$

Using the following decomposition:

$$\begin{aligned}
&\nu_n^j(y_i)\nu_n^l(y_i) - \nu^j(y_i)\nu^l(y_i) \\
&= \left(\frac{g_n^j(y_i)-g^j(y_i)}{f_n^j(y_i)}\right)\left(\frac{g_n^l(y_i)-g^l(y_i)}{f_n^l(y_i)}\right) \\
&+ \left(\frac{g_n^j(y_i)-g^j(y_i)}{f_n^j(y_i)}\right)\left(\frac{f(y_i)-f_n^l(y_i)}{f_n^l(y_i)}+1\right)\nu^l(y_i) \\
&+ \left(\frac{g_n^l(y_i)-g^l(y_i)}{f_n^l(y_i)}\right)\left(\frac{f(y_i)-f_n^j(y_i)}{f_n^j(y_i)}+1\right)\nu^j(y_i) \\
&+ \left(\frac{f(y_i)-f_n^j(y_i)}{f_n^j(y_i)}\right)\left(\frac{f_n^l(y_i)-f(y_i)}{f_n^l(y_i)}+2\right)\nu^j(y_i)\nu^l(y_i) \\
&+ \left(\frac{f(y_i)-f_n^l(y_i)}{f_n^l(y_i)}\right)\left(\frac{f(y_i)-f_n^j(y_i)}{f_n^j(y_i)}+1\right)\nu^j(y_i)\nu^l(y_i) \\
&+ \left(\frac{f_n^j(y_i)-f(y_i)}{f_n^j(y_i)}\right)\left(\frac{f(y_i)-f_n^l(y_i)}{f_n^l(y_i)}+1\right)\nu^j(y_i)\nu^l(y_i)
\end{aligned}$$

and using the fact that under H_3, with a probability converging to 1:

$$\forall \epsilon \in (0,1),\ \forall y \in I\!R,\quad f(y) > b_n \Longrightarrow \frac{1}{f_n^j(y)} < \frac{1}{b_n(1-\epsilon)}$$

we thus obtain the following upper bound for $\left|\bar{\theta} - \theta_0\right|$, with a probability converging to 1:

$$\begin{aligned}
&1/(b_n(1-\epsilon))\,[\,(1/(b_n(1-\epsilon))\sup_y \left|g_n^j(y) - g^j(y)\right| \sup_y \left|g_n^l(y) - g^l(y)\right| \tfrac{1}{n}\textstyle\sum_{i=1}^n I_i \\
&+ \sup_y \left|g_n^j(y) - g^j(y)\right| \left(\tfrac{1}{b_n(1-\epsilon)}\sup_y \left|f(y) - f_n^l(y)\right| + 1\right) \tfrac{1}{n}\textstyle\sum_{i=1}^n |\nu^l(y_i)| I_i \\
&+ \sup_y \left|g_n^l(y) - g^l(y)\right| \left(\tfrac{1}{b_n(1-\epsilon)}\sup_y \left|f(y) - f_n^j(y)\right| + 1\right) \tfrac{1}{n}\textstyle\sum_{i=1}^n |\nu^j(y_i)| I_i \\
&+ \sup_y \left|f(y) - f_n^j(y)\right| \left(\tfrac{1}{b_n(1-\epsilon)}\sup_y \left|f_n^l(y) - f(y)\right| + 2\right) \tfrac{1}{n}\textstyle\sum_{i=1}^n |\nu^j(y_i)\nu^l(y_i)| I_i \\
&+ \sup_y \left|f(y) - f_n^l(y)\right| \left(\tfrac{1}{b_n(1-\epsilon)}\sup_y \left|f_n^j(y) - f(y)\right| + 1\right) \tfrac{1}{n}\textstyle\sum_{i=1}^n |\nu^j(y_i)\nu^l(y_i)| I_i \\
&+ \sup_y \left|f(y) - f_n^j(y)\right| \left(\tfrac{1}{b_n(1-\epsilon)}\sup_y \left|f_n^l(y) - f(y)\right| + 1\right) \tfrac{1}{n}\textstyle\sum_{i=1}^n |\nu^j(y_i)\nu^l(y_i)| I_i\,]
\end{aligned}$$

It is easy to show, using H_2, H_6 and Markov inequality, that:

$$1/n\sum_{i=1}^{n} I_i = O_p(1)\ , \quad 1/n\sum_{i=1}^{n} |\nu^l(y_i)| I_i = O_p(1),$$

$$1/n \sum_{i=1}^{n} |\nu^j(y_i)| I_i = O_p(1) \ , \quad \text{and} \quad 1/n \sum_{i=1}^{n} |\nu^j(y_i)\nu^l(y_i)| I_i = O_p(1).$$

Thus, eventually we have: $\left|\bar{\theta} - \theta_0\right| = o_p(1)$.

STEP 2: By H_1 and the weak law of large numbers, $\theta_0 \longrightarrow \Gamma_{j,l}$ in probability.

STEP 3: Let us show that $\tilde{\Gamma}_{j,l} - \bar{\theta} = o_p(1)$.
Some notations will be useful:

$$J_i = I_{[f_n(y_i) > b_n \text{ and } f(y_i) \leq b_n]} \quad \text{and} \quad M_i = I_{[f_n(y_i) \leq b_n \text{ and } f(y_i) > b_n]}.$$

Thus:

$$\tilde{\Gamma}_{j,l} - \bar{\theta} = \frac{1}{n} \sum_{i=1}^{n} \nu_n^j(y_i)\nu_n^l(y_i) \left(J_i - M_i\right).$$

Using H_5, it is straighforward to see that $P(M_i)$ is asymptotically zero. Thus, the term corresponding to M_i converges in probability to 0.
We now need to show that $\frac{1}{n}\sum_{i=1}^{n} \nu_n^j(y_i)\nu_n^l(y_i)J_i = o_p(1)$.
Using the same demonstration as in step 1, we can obtain an upper bound for the difference

$$\left| \frac{1}{n} \sum_{i=1}^{n} [\nu_n^j(y_i)\nu_n^l(y_i) - \nu^j(y_i)\nu^l(y_i)] J_i \right|.$$

By the Lebesgue dominated convergence theorem and Markov inequality, it can be shown that:

$$\frac{1}{n} \sum_{i=1}^{n} J_i = o_p(1) \ , \quad \frac{1}{n} \sum_{i=1}^{n} |\nu_n^l(y_i)| J_i = o_p(1),$$

$$\frac{1}{n} \sum_{i=1}^{n} |\nu_n^j(y_i)| J_i = o_p(1) \ , \quad \text{and} \quad \frac{1}{n} \sum_{i=1}^{n} |\nu_n^l(y_i)\nu_n^j(y_i)| J_i = o_p(1)$$

Thus as in step 1:

$$\left| \frac{1}{n} \sum_{i=1}^{n} [\nu_n^j(y_i)\nu_n^l(y_i) - \nu^j(y_i)\nu^l(y_i)] J_i \right| = o_p(1),$$

and therefore, $1/n \sum_{i=1}^{n} \nu_n^j(y_i)\nu_n^l(y_i)J_i = o_p(1)$.
Finally, we have shown that:

$$\tilde{\Gamma}_{j,l} - \bar{\theta} = o_p(1)$$

CONCLUSION: Using the results of steps 1, 2 and 3, we obtain:

$$\tilde{\Gamma}_{j,l} \longrightarrow \Gamma_{j,l} \text{ in probability.}$$

6.2.3 Convergence in quadratic mean of $\hat{\Gamma}_{j,l} = \int_{\mathbb{R}} \nu_n^j(y)\nu_n^l(y)f_n(y)dy$ to $\Gamma_{j,l} = E\{E[z^j|y]E[z^l|y]\}$

Assumptions:

H_0 : (z_i', y_i) $i = 1, ..., n$ are i.i.d.

H_1 : K is a density.

H_2 : Kernel estimators f_n, f_n^j, g_n^j, and $d_n^{j,l}$ based on K are uniformly convergent on $\mathbb{R}$, where $d_n^{j,l} = 1/nh_n \sum_{i=1}^n z_i^j z_i^l K((y_i - y)/(h_n))$, for $j, l = 1, ..., p$.

H_3 : $\forall y \in \mathbb{R}$, $f_n(y)$, $\nu_n^j(y)$, converge almost surely to $f(y)$, $\nu^j(y)$ for $j = 1, ..., p$.

H_4 : $E\left[(z^j))^4\right] < \infty$ for $j = 1, ..., p$.

Proof:

Let us denote $V_n = \frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}} (z_i^j - \nu_n^j(y))(z_i^l - \nu_n^l(y))K_n(y - y_i)dy$.
Under H_1, the following equality is easily derived:

$$V_n = \frac{1}{n}\sum_{i=1}^n z_i^j z_i^l - \int_{\mathbb{R}} \nu_n^j(y)\nu_n^l(y)f_n^l(y)dy - \int_{\mathbb{R}} \nu_n^j(y)\nu_n^l(y)f_n^j(y)dy$$

$$+ \int_{\mathbb{R}} \nu_n^j(y)\nu_n^l(y)f_n(y)dy$$

Without loss of generality, to keep notations simple, we assume here a common smoothing parameter h_n^* for the estimators ν_n^j, ν_n^l et f_n. Thus, we have:

$$V_n = \frac{1}{n}\sum_{i=1}^n z_i^j z_i^l - \int_{\mathbb{R}} \nu_n^j(y)\nu_n^l(y)f_n(y)dy$$

The proof splits into three steps.

STEP 1: We show the convergence in probability of V_n to $E[cov(z^j, z^l|y)]$.
Let $\alpha \in \mathbb{R}^+ - \{0\}$. Note $D_\alpha = \{y|f(y) \geq \alpha\}$, $E_\alpha = \mathbb{R} - D_\alpha$, and $c(y) = cov(z^j, z^l|Y = y)$. To simplify notation, we will write d_n instead of $d_n^{j,l}$. Thus, we have:

$$\begin{aligned} & V_n - E[cov(z^j z^l|y)] \\ = & \int_{D_\alpha} \{d_n(y) - \nu_n^j(y)\nu_n^l(y)f_n(y) - c(y)f(y)\}dy \\ + & \int_{E_\alpha} \{d_n(y) - \nu_n^j(y)\nu_n^l(y)f_n(y)\}dy - \int_{D_\alpha} c(y)f(y)dy \end{aligned}$$

Let $\epsilon > 0$. It is straightforward to see that:

$$\Pr\left[\left|V_n - E[cov(z^j z^l|y)]\right| > 3\epsilon\right] \leq a + b + c$$

where: $a = \Pr\left[\left|\int_{D_\alpha} \{d_n(y) - \nu_n^j(y)\nu_n^l(y)f_n(y) - c(y)f(y)\}dy\right| > \epsilon\right]$,

$b = \Pr\left[\left|\int_{E_\alpha} \{d_n(y) - \nu_n^j(y)\nu_n^l(y)f_n(y)\}dy\right| > \epsilon\right]$ and $c = \Pr\left[\left|\int_{E_\alpha} c(y)f(y)dy\right| > \epsilon\right]$.

We can show that the probabilities a, b and c converge to zero as $n \to \infty$.

• For the probability c, $\forall \epsilon > 0$, by choosing α sufficiently small, $\int_{E_\alpha} c(y)f(y)dy$ will be lower than ϵ.

• For the probability b, note $B = \int_{E_\alpha} \{d_n(y) - \nu_n^j(y)\nu_n^l(y)f_n(y)\}dy$. We use the following inequality:

$$0 \leq |B| \leq \int_{E_\alpha} |d_n(y)|dy + \int_{E_\alpha} |\nu_n^j(y)\nu_n^l(y)f_n(y)|dy.$$

Under H_3, it is easily shown that $\forall \epsilon > 0$, by choosing α sufficiently small, the two right terms of this inequality will be lower than ϵ.

• For the probability a, we study the term

$$A = \sup_{y \in D_\alpha} |d_n(y) - \nu_n^j(y)\nu_n^l(y)f_n(y) - c(y)f(y)|.$$

By noting that $c(y) = s(y) - \nu^j(y)\nu^l(y)$ and $d(y) = s(y)f(y)$,

$$A \leq \sup_{y \in D_\alpha} |d_n(y) - d(y)| + \sup_{y \in D_\alpha} |\nu_n^j(y)\nu_n^l(y)f_n(y) - \nu^j(y)\nu^l(y)f(y)|.$$

Under H_2, we can show the two left terms of this majoration are $o_p(1)$. Thus, a converges to 0 $\forall \alpha \in \mathbb{R}^+ - \{0\}$.

Finally, we have the result:

$$\Pr[|V_n - E[cov(z^j, z^l|y)]| > 3\epsilon] \longrightarrow 0 \text{ for } n \to \infty,$$

$$\text{i.e} \quad V_n \longrightarrow E[cov(z^j, z^l|y)] \text{ in probability for } n \to \infty.$$

STEP 2: From step 1 and the weak law of large numbers, we get:

$$\hat{\Gamma}_{j,l} \longrightarrow \Gamma_{j,l} \text{ in probability.}$$

STEP 3: $\hat{\Gamma}_{j,l}$ converges in quadratic mean (q.m.) to $\Gamma_{j,l}$.

From the Cauchy-Schwarz inequality, we get:

$$\left|\int_{\mathbb{R}} \nu_n^j(y)\nu_n^l(y)f_n(y)dy\right| \leq \left\{\int_{\mathbb{R}} (\nu_n^j(y))^2 f_n(y)dy\right\}^{1/2} \left\{\int_{\mathbb{R}} (\nu_n^l(y))^2 f_n(y)dy\right\}^{1/2}$$

Let $V_n' = \frac{1}{n}\sum_{i=1}^n (z_i^j)^2 - \int_{\mathbb{R}} (\nu_n^j(y))^2 f_n(y)dy$ and $T_n = \frac{1}{n}\sum_{i=1}^n (z_i^j)^2$. Thus, $\forall n$, $0 \leq V_n' \leq T_n$. By the weak law of large numbers, $T_n \longrightarrow E[(z^j)^2]$ in probability. Under H_4, it can be shown that for $n \to \infty$, $E[T_n^2] < \infty$.

Thus, by lemma B of Serfling ([11], p. 15), $\{ T_n^2 \}$ is uniformly integrable, therefore $\{ V_n'^{\,2} \}$ is uniformly integrable.
From this and step 2, we can apply lemma C of Serfling ([11], p. 15) and conclude that

$$V_n' \longrightarrow E[var(z^j|y)] \text{ in q.m.}$$

A direct consequence of this result is:

$$\int_{I\!R} (\nu_n^j(y))^2 f_n(y)dy \longrightarrow E\left[\ E^2[z^j|y]\ \right] \quad \text{in q.m.}$$

From $E\left|T_n^2\right|^2 < \infty$ and $E\left|V_n'^{\,2}\right|^2 < \infty$, we have $E\left|\int_{I\!R} (\nu_n^j(y))^2 f_n(y)dy\right|^2 < \infty$. Thus, by the same lemma C, $\{(\int_{I\!R} (\nu_n^j(y))^2 f_n(y)dy)^2\}$ is uniformly integrable, and therefore $\{\int_{I\!R} (\nu_n^j(y))^2 f_n(y)dy\}$ is uniformly integrable.
From the Cauchy-Schwarz inequality, $\{(\int_{I\!R} \nu_n^j(y)\nu_n^l(y) f_n(y)dy)^2\}$ is uniformly integrable.
An inspection of $E\left|\int_{I\!R} \nu_n^j(y)\nu_n^l(y) f_n(y)dy\right|^2$ shows that it is bounded. From step 2 and lemma C, we ultimately obtain:

$$\int_{I\!R} \nu_n^j(y)\nu_n^l(y) f_n(y)dy \longrightarrow E\left[E\{z^j|y\}E\{z^l|y\}\right] \text{ in q.m.}$$

References

[1] Aragon, Y., Li, K.C. and Thomas–Agnan, C. (1994), Modelling income distributions using multivariate sliced inverse regression, *unpublished.*

[2] Carroll, R.J. and Li, K.C. (1992), Measurement error regression with unknown link: dimension reduction and data visualization, *J. Amer. Stat. Assoc.*, **87**, 1040-1050.

[3] Duan, N. and Li, K.C. (1991), Slicing regression: a link-free regression method, *The Annals of Statistics*, **19**, 505-530.

[4] Haerdle, W. (1990), *Smoothing techniques with implementation in S*, Springer.

[5] Hsing, T. and Carroll, R.J. (1992), An asympotic theory for Sliced Inverse regression, *The Annals of Statistics*, **20**, 1040-1061.

[6] Johnson, M.E. (1987), *Multivariate Statistical Simulation*, Wiley, New-York.

[7] Lavergne, P. (1993), Sélection non paramétrique de régresseurs, *Thèse pour le doctorat en Sciences Économiques - Université des Sciences Sociales de Toulouse.*

[8] LI, K.C. (1991), Sliced inverse regression for dimension reduction, with discussions, *Journal of the American Statistical Association*, **86**, 316-342.

[9] LI, K.C., ARAGON, Y. and THOMAS–AGNAN, C. (1994), Sliced inverse regression for multivariate outcome data: a nonlinear theory of Hotelling's most predictive variates, *preprint.*

[10] MUELLER, H.G. and STADTMUELLER, U. (1987), Estimation of heteroscedasticity in regression analysis, *The Annals of Statistics*, **15**, 610-625.

[11] SERFLING, R.J. (1980), *Approximation theorems of Mathematical Statistics*, Wiley, New-York.