Estimation of the solution of an Integral Equation of the Second Kind[1]

Marine Carrasco
University of Rochester

Jean-Pierre Florens
Université de Toulouse (GREMAQ and IDEI)

Eric Renault
Université de Montréal (CIRANO and CIREQ)

September 2003

[1]Carrasco gratefully aknowledges financial support from the National Science Foundation, grant # SES-0211418.

1 Introduction

The objective of this section is to study the properties of the solution of an integral equation of the second kind (also called Fredholm equation of the second type) defined by:

$$(I - K)\varphi = r \tag{1}$$

where φ is an element of an Hilbert space $\mathcal{H}$, K is a compact operator from $\mathcal{H}$ to $\mathcal{H}$ and r is an element of $\mathcal{H}$. As in the previous sections, K and r are known functions of a data generating process characterized by its c.d.f. F and the functional parameter of interest is the function φ.

In most cases, $\mathcal{H}$ is a functional space and K is an integral operator defined by its kernel k and Equation (1) becomes:

$$\varphi(t) - \int k(t,s)\varphi(s)\Pi(ds) = r(t) \tag{2}$$

The estimated operators are often degenerated, see Subsection 2.6.2. and, in that case, Equation (2) simplifies into:

$$\varphi(t) - \sum_{\ell=1}^{L} a_\ell(\varphi)\varepsilon_\ell(t) = r(t) \tag{3}$$

where the $a_\ell(\varphi)$ are linear forms on $\mathcal{H}$ and ε_ℓ belongs to $\mathcal{H}$ for any ℓ.

The essential difference between equations of the first kind and of the second kind is the compactness of the operator. In (1), K is compact but $I - K$ is not compact. Moreover, if $I - K$ is one-to-one, its inverse is bounded. In that case, the inverse problem is well-posed. Even if $I - K$ is not one-to-one the ill-poseness of equation (1) is less severe than in the first kind case because the solutions are stable in r.

In most cases, K is a self-adjoint operator (and hence $I - K$ is also self-adjoint) but we will not restrict our presentation to this case. On the other hand, Equation (1) could be extended by considering an equation $(S - K)\varphi = r$ where K is now a compact operator from $\mathcal{H}$ to $\mathcal{E}$ and S is a bounded operator from $\mathcal{H}$ to $\mathcal{E}$, one-to-one with a bounded inverse. This extension will not be considered in this paper.

This section will be organized in the following way. The next paragraph recalls the main mathematical properties of the equations of the second kind. The two following paragraphs present the statistical properties of the solution in the cases of well-posed and of ill-posed problems and the last paragraph applies these results to the two examples given in Section 1.

The implementation of the estimation procedures is not discussed here because this issue is similar to the implementation of the estimation of a regularized equation of the first kind (see Section 3). Actually, regularizations transform first kind equations into second kind equations and the numerical methods are then formally equivalent, even though statistical properties are fundamentally different.

2 Riesz theory and Fredholm alternative

We first briefly recall the main results about equations of the second kind as they were developed at the beginning of the 20th century by Fredholm and Riesz. The statements are given without proofs (see e.g. Kress, 1999, Chapters 3 and 4).

Let K be a compact operator from $\mathcal{H}$ to $\mathcal{H}$ and I be the identity on $\mathcal{H}$ (which is compact only if $\mathcal{H}$ is finite dimensional). Then the operator $I - K$ has a finite dimensional null space and its range is closed. Moreover $I - K$ is injective if and only if it is surjective. In that case $I - K$ is invertible and its inverse $(I - K)^{-1}$ is a bounded operator.

An element of the null space of $I-K$ verifies $K\varphi = \varphi$ and if $\varphi \neq 0$, it is an eigenfunction of K associated with the eigenvalue equal to 1. Equivalently the inverse problem (1) is well-posed if and only if 1 is not an eigenvalue of K. The Fredholm alternative follows from the previous results.

Theorem 1 (Fredholm alternative) *Let us consider the two equations of the second kind:*

$$(I - K)\varphi = r \tag{4}$$

and

$$(I - K^*)\psi = s \tag{5}$$

where K^* is the adjoint of K. Then:

i) Either the two homogeneous equations $(I - K)\varphi = 0$ and $(I - K^*)\psi = 0$ only have the trivial solutions $\varphi = 0$ and $\psi = 0$ and in that case (4) and (5) have a unique solution for any r and s in $\mathcal{H}$

ii) or the two homogeneous equations $(I - K)\varphi = 0$ and $(I - K^*)\psi = 0$ have the same finite number m of linearly independent solutions φ_j and ψ_j $(j = 1, ..., m)$ respectively and the solutions of (4) and (5) exist if and only if $\langle \psi_j, r\rangle = 0$ and $\langle \varphi_j, s\rangle = 0$ for any $j = 1, ..., m$.

3 Statistical properties of the solution of a well-posed equation of the second kind

In the case of a one to one equation of the second kind, the asymptotic properties are easily deduced from the properties of the estimation of the operator K and of the right-hand side r.

The starting point of this analysis is the relation:

$$\begin{aligned}
\hat{\varphi}_n - \varphi_0 &= \left(I - \hat{K}_n\right)^{-1} \hat{r}_n - (I-K)^{-1} r \\
&= \left(I - \hat{K}_n\right)^{-1} (\hat{r}_n - r) + \left[\left(I - \hat{K}_n\right)^{-1} - (I-K)^{-1}\right] r \\
&= \left(I - \hat{K}_n\right)^{-1} \left[\hat{r}_n - r + \left(\hat{K}_n - K\right)(I-K)^{-1} r\right] \\
&= \left(I - \hat{K}_n\right)^{-1} \left[\hat{r}_n - r + \left(\hat{K}_n - K\right)\varphi_0\right] \qquad (6)
\end{aligned}$$

where the third equality follows from $A^{-1} - B^{-1} = A^{-1}(B-A)B^{-1}$.

Theorem 2 *If*

i) $\left\|\hat{K}_n - K\right\| = o(1)$

ii) $\left\|\left(\hat{r}_n + \hat{K}_n\varphi_0\right) - (r + K\varphi_0)\right\| = O\left(\frac{1}{a_n}\right)$

Then $\|\hat{\varphi}_n - \varphi_0\| = O\left(\frac{1}{a_n}\right)$

Proof. As $I-K$ is invertible and admits a continuous inverse, i) implies that $\| \left(I - \hat{K}_n\right)^{-1} \|$ converges to $\left\|(I-K)^{-1}\right\|$ and the result follows from (6). ■

In some cases $\|r - \hat{r}_n\| = O(\frac{1}{b_n})$ and $\|\hat{K}_n - K\| = O(\frac{1}{d_n})$. Then $\frac{1}{a_n} = \frac{1}{b_n} + \frac{1}{d_n}$. In some particular examples, as it will be illustrated in the last subsection the asymptotic behavior of $\hat{r}_n - \hat{K}_n\varphi$ is directly considered.

Asymptotic normality can be obtained from different sets of assumptions. The following theorems illustrate two kinds of asymptotic normality.

Theorem 3 *If*

i) $\left\|\hat{K}_n - K\right\| = o(1)$

ii) $a_n\left(\left(\hat{r}_n + \hat{K}_n\varphi_0\right) - (r + K\varphi_0)\right) \Longrightarrow N(0, \Sigma)$ *(weak convergence in $\mathcal{H}$)*

Then

$$a_n(\hat{\varphi}_n - \varphi_0) \Longrightarrow \mathcal{N}\left(0, (I-K)^{-1}\Sigma(I-K^*)^{-1}\right)$$

Proof. The proof follows immediately from (6) and Theorem **??** in Section 2. ■

Theorem 4 *We consider the case where $\mathcal{H} = L^2(\mathbb{R}^p, \pi)$. If* *i)* $\|\hat{K}_n - K\| = o(1)$

ii) $\exists a_n$ *s.t* $a_n\left[\left(\hat{r}_n + \hat{K}_n\varphi_0\right) - (r + K\varphi_0)\right](x) \xrightarrow{d} N\left(0, \sigma^2(x)\right), \quad \forall x \in \mathbb{R}^p$

iii) $\exists b_n$ *s.t* $\frac{a_n}{b_n} = o(1)$ *and*

$$b_n\hat{K}\left[\left(\hat{r}_n + \hat{K}_n\varphi\right) - (r + K\varphi_0)\right] \Longrightarrow N(0, \Omega)$$

(weak convergence in $\mathcal{H}$)

Then

$$a_n(\hat{\varphi}_n - \varphi_0)(x) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2(x)\right) \qquad \forall x$$

Proof. Using

$$(I-K)^{-1} = I + (I-K)^{-1}K$$

we deduce from (6):

$$\begin{aligned} a_n(\hat{\varphi}_n - \varphi_0)(x) &= a_n\left\{(I-\hat{K}_n)^{-1}\left[\hat{r}_n + \hat{K}\varphi_0 - r - K\varphi_0\right]\right\} \\ &= a_n(\hat{r} + \hat{K}\varphi_0 - r - K\varphi_0)(x) \\ &+ \frac{a_n}{b_n}\left\{b_n(I-\hat{K})^{-1}\hat{K}(\hat{r} + \hat{K}\varphi_0 - r - K\varphi_0)\right\}(x) \end{aligned} \tag{7}$$

The last term into bracket converges (weakly in L^2). to a $N(0, (I-K)^{-1}\Omega(I-K)^{-1})$ and the value of this function at any point x also converges to a normal distribution (weak convergence implies finite dimensional convergences). Then the last term into brackets is bounded and the result is verified. ■

Note that condition (iii) is justified by circumstances when K is an integral operator which increases the rate of convergence of $\hat{r}_n + \hat{K}_n\varphi$.

We illustrate these results by the following three examples even if the first one appears to be a little artificial.

Example. Consider $L^2(\mathbb{R}, \pi)$ and (Y, Z) is a random element of $\mathbb{R} \times L^2(\mathbb{R}, \pi)$. We study the integral equation of the second kind defined by

$$\varphi(x) + \int E^F\left(Z(x)Z(y)\right)\varphi(y)\pi(dy) = E^F(YZ(x)) \tag{8}$$

denoted by $\varphi + V\varphi = r$.

This equation defines a well posed inverse problem because the covariance operator is positive. We assume that an i.i.d. sample of (Y, Z) is available and the estimated equation

(8) defines the parameter of interest as the solution of an integral equation having the following form:

$$\varphi(x) + \frac{1}{n}\sum_{i=1}^{n} z_i(x) \int z_i(y)\varphi(y)\pi\,(dy) = \frac{1}{n}\sum_{i=1}^{n} y_i z_i(x) \tag{9}$$

Under regularity conditions one can check that $\|\hat{V}_n - V\| = O\left(\frac{1}{\sqrt{n}}\right)$ and that

$$\sqrt{n}\frac{1}{n}\sum_{i}\left\{z_i(\cdot)\left[y_i - \int z_i(y)\varphi(y)\pi(dy)\right] - E^F\left(YZ\left(\cdot\right)\right) - \int E^F(Z(.)Z(y))\varphi(y)\pi(dy)\right\}$$

$\Rightarrow \quad N(0,\Sigma)$ in $L^2(\mathbb{R},\pi)$.

If for instance $E^F(Y|Z) = \int Z(y)\varphi(y)\pi(y)dy$ and under a homoscedasticity hypothesis the operator Σ is a covariance operator with kernel $\sigma^2 E^F(Z(x)Z(y))$ where

$$\sigma^2 = Var\left(Y - \int Z(y)\varphi(y)\pi(dy)|Z\right).$$

Then, from Theorem 3,

$$\sqrt{n}\left(\hat{\varphi}_n - \varphi_0\right) \Rightarrow N\left(0, \sigma^2(I+V)^{-1}V(I+V)^{-1}\right) \tag{10}$$

Example. Rational expectations asset pricing models:

Following Lucas (1978), such models characterize the pricing functional as a function φ of the Markov state solution of an integral equation:

$$\varphi\left(x\right) - \int a(x,y)\varphi\left(y\right) f\left(y|x\right) dy = \int a(x,y)b(y)f\left(y|x\right) dy \tag{11}$$

While f is the transition density of the Markov state, the function a denotes the marginal rate of substitution and b the dividend function. For sake of expositional simplicity, we assume here that the functions a and b are both known while f is estimated nonparametrically by a kernel method. Note that if the marginal rate of substitution a involves some unknown preference parameters (subjective discount factor, risk aversion parameter), they will be estimated, for instance by GMM, with a parametric root n rate of convergence. Therefore, the nonparametric inference about φ (deduced by solution of (11) of a kernel estimation of f) is not contaminated by this parametric estimation; all the statistical asymptotic theory can be derived as if the preference parameters were known.

As far as kernel density estimation is concerned, it is well known that under mild conditions (see e.g. Bosq (1998)) it is possible to get with stationary strongly mixing stochastic processes the same convergence rates and the same asymptotic distribution as in the i.i.d. case. Then, we do not make explicit in this presentation the assumed dynamic properties of the observations y and x of present and lagged values of a Markov process.

Let us then consider a n-dimensional stationary stochastic process X_t and $\mathcal{H}$ the space of square integrable functions of one realization of this process. In this example, $\mathcal{H}$ is defined with respect to the true distribution. The operator K is defined by

$$K\varphi(x) = E^F\left(a\left(X_{t-1}, X_t\right)\varphi\left(X_t\right)|X_{t-1} = x\right) \tag{12}$$

and

$$r(x) = E^F\left(a\left(X_{t-1}, X_t\right)b(X_t)|X_{t-1} = x\right) \tag{13}$$

We will assume that K is compact though for example a Hilbert-Schmidt condition (see assumption A.1 of Section **??** for such a condition). A common assumption in rational expectation models is that K is a contraction mapping, due to discounting. Then, 1 is not an eigenvalue of K and (11) is a well-posed Fredholm integral equation.

Under these hypotheses, both numerical and statistical issues associated with the solution of (11) are well documented. See Rust, Traub and Wozniakowski (2002) and references therein for numerical issues. The statistical consistency of the estimator $\hat{\varphi}_n$ deduced from the kernel estimator $\hat{K}_n$ is deduced from Theorem 2 above. Assumption i) is satisfied because $\hat{K}_n - K$ has the same behavior as the conditional expectation operator and

$$\begin{aligned}
&\hat{r}_n + \hat{K}_n\varphi - r - K\varphi \\
&\quad = E^{F_n}\left(a\left(X_{t-1}, X_t\right)\left(b(X_t) + \varphi\left(X_t\right)\right)|X_{t-1}\right) \\
&\quad -E^F\left(a\left(X_{t-1}, X_t\right)\left(b(X_t) + \varphi\left(X_t\right)\right)|X_{t-1}\right)
\end{aligned}$$

converges at the speed $\frac{1}{a_n} = \left(\frac{1}{nc_n^m} + c_n^4\right)^{1/2}$ if c_n is the bandwidth of the (second order) kernel estimator and m is the dimension of X.

The weak convergence obtained through Theorem 4, Assumption ii) is the usual result on the normality of kernel estimation of conditional expectation. As K is an integral operator, the transformation by K increases the speed of convergence which implies iii).

Example: Partially Nonparametric forecasting model:

Nonparametric prediction of a stationary ergodic scalar random process X_t is often performed by looking for a predictor $\varphi\left(X_{t-1}, ..., X_{t-d}\right)$ able to minimize the mean square error of prediction:

$$E\left[X_t - \varphi\left(X_{t-1}, ..., X_{t-d}\right)\right]^2$$

In other words, if φ can be any squared integrable function, the optimal predictor is the conditional expectation

$$\varphi_0\left(X_{t-1,...,}X_{t-d}\right) = E\left[X_t|X_{t-1,...,}X_{t-d}\right]$$

and can be estimated by kernel smoothing or any other nonparametric way to estimate a regression function. The problems with this kind of approach are twofold. First, it is often necessary to include many lagged variables and the resulting nonparametric estimation surface suffers from the well-known "curse of dimensionality". Second, it is hard to describe and interpret the estimated regression surface when the dimension is more than two.

A solution to deal with these problems is to think about a kind of nonparametric generalization of ARMA processes. For this purpose, let us consider semiparametric predictors of the following form

$$E\left[X_t|I_{t-1}\right]=m_{\varphi}\left(\theta,I_{t-1}\right)=\sum_{j=1}^{\infty}a_j\left(\theta\right)\varphi\left(X_{t-j}\right) \tag{14}$$

where θ is an unknown finite dimensional vector of parameters, $a_j(.)$, $j\geq 1$, are known given scalar functions and $\varphi(.)$ is the unknown functional parameter of interest. The notation

$$E\left[X_t|I_{t-1}\right]=m_{\varphi}\left(\theta,I_{t-1}\right)$$

stresses the fact that the predictor depends on the true unknown value of the parameters θ and φ and of the information I_{t-1} available at time $(t-1)$ as well. This information is actually the σ-field generated by $X_{t-j}, j\geq 1$. A typical example is

$$a_j\left(\theta\right)=\theta^{j-1} \text{ for } j\geq 1 \text{ with } 0<\theta<1. \tag{15}$$

Then, the predictor (7.14) is actually characterized by

$$m_{\varphi}\left(\theta,I_{t-1}\right)=\theta m_{\varphi}\left(\theta,I_{t-2}\right)+\varphi\left(X_{t-1}\right) \tag{16}$$

In the context of volatility modelling, X_t would denote a squared asset return over period $[t-1,t]$ and $m_{\varphi}\left(\theta,I_{t-1}\right)$ the so-called squared volatility of this return as expected at the beginning of the period. Engle and Ng (1993) have studied such a partially nonparametric (PNP for short) model of volatility and called the function φ the "news impact function". They proposed an estimation strategy based on piecewise linear splines. Note that the PNP model includes several popular parametric volatility models as special cases. For instance, the GARCH (1,1) model of Bollerslev (1986) corresponds to $\varphi(x)=w+\alpha x$ while the Engle (1990) asymmetric model is obtained for $\varphi(x)=w+\alpha(x+\delta)^2$. See also Linton and Mammen (2003) and references therein.

The nonparametric identification and estimation of the news impact function can be derived for a given value of θ. After that, a profile criterion can be calculated to estimate θ. In any case, since θ will be estimated with a parametric rate of convergence, the asymptotic distribution theory of a nonparametric estimator of φ is the same as if θ were known. For sake of notational simplicity, the dependence on unknown finite dimensional parameters θ is no longer made explicit.

At least in the particular case (15)-(16), φ is easily characterized as the solution of a linear integral equation of the first kind

$$E\left[X_t-\theta X_{t-1}|I_{t-2}\right]=E\left[\varphi\left(X_{t-1}\right)|I_{t-2}\right] \tag{17}$$

Except for its dynamic features, this problem is completely similar to the nonparametric instrumental regression example described in Section **??**. However, as already mentioned, problems of the second kind are often preferable since they may be well-posed. As shown by Linton and Mammen (2003) in the particular case of a PNP volatility model, it is actually possible to identify and consistently estimate the function φ of interest in (**??**) from a well-posed linear inverse problem of the second kind. The main trick is to realize that φ is characterized by the first order conditions of the least squares problem

$$\min_{\varphi} E\left[X_t - \sum_{j=1}^{\infty} a_j \varphi\left(X_{t-j}\right)\right]^2 \tag{18}$$

Then, when φ is an element of the Hilbert space $L_F^2(X)$, its true unknown value is characterized by the first order conditions obtained by differentiating in the direction of any vector h

$$E\left[\left(X_t - \sum_{j=1}^{\infty} a_j \varphi\left(X_{t-j}\right)\right)\left(\sum_{l=1}^{\infty} a_l h\left(X_{t-l}\right)\right)\right] = 0$$

In other words, for any h in $L_F^2(X)$

$$\begin{aligned} &\sum_{j=1}^{\infty} a_j E^X\left[E\left[X_t | X_{t-j} = x\right] h(x)\right] \\ &-\sum_{j=1}^{\infty} a_j^2 E^X\left[\varphi(x) h(x)\right] \\ &-\sum_{j=1}^{\infty} \sum_{\substack{l=1 \\ l \neq j}}^{\infty} a_j a_l E^X\left[E\left[\varphi\left(X_{t-l}\right) | X_{t-j} = x\right] h(x)\right] = 0 \end{aligned} \tag{19}$$

where E^X denotes the expectation with respect to the stationary distribution of X_t. As the equality (19) is true for all h, it is in particular true for a complete sequence of functions of $L_F^2(X)$. It follows that

$$\begin{aligned} &\sum_{j=1}^{\infty} a_j E\left[X_t | X_{t-j}\right] - \left(\sum_{l=1}^{\infty} a_l^2\right) \varphi\left(X_{t-j}\right) \\ &-\sum_{j=1}^{\infty} \sum_{l \neq j}^{\infty} a_j a_l E\left[\varphi\left(X_{t-l}\right) | X_{t-j}\right] = 0 \end{aligned}$$

P^X- almost surely. Let us denote

$$r_j\left(X_t\right) = E\left[X_{t+j} | X_t\right] \text{ and } H_k(\varphi)\left(X_t\right) = E\left[\varphi\left(X_{t+k}\right) | X_t\right].$$

Then, we have proved that the unknown function φ of interest must be the solution of the linear inverse problem of the second kind

$$A(\varphi, F) = (I - K)\varphi - r = 0 \tag{20}$$

where

$$
\begin{aligned}
r &= \left(\sum_{j=1}^{\infty} a_j^2\right)^{-1} \sum_{j=1}^{\infty} a_j r_j \\
K &= -\left(\sum_{j=1}^{\infty} a_j^2\right)^{-1} \sum_{j=1}^{\infty} \sum_{l \neq j} a_j a_l H_{j-l}
\end{aligned},
$$

and, with a slight change of notation, F characterizes now the probability distribution of the stationary process (X_t).

To study the inverse problem (20), it is first worth noticing that K is a self adjoint integral operator. Indeed, while

$$
K = \left(\sum_{j=1}^{\infty} a_j^2\right)^{-1} \sum_{h=\pm 1}^{+\infty} H_k \left(\sum_{l=\max[1,1-k]}^{+\infty} a_l a_{l+k}\right)
$$

we immediately deduce from Subsection 2.5.1 that the conditional expectation operator H_k is such that

$$
H_k^* = H_{-k}
$$

and thus $K = K^*$, since

$$
\sum_{l=\max[1,1-k]}^{+\infty} a_l a_{l+k} = \sum_{l=\max[1,1+k]}^{+\infty} a_l a_{l-k}
$$

As noticed by Linton and Mammen (2003), this property greatly simplify the practical implementation of the solution of a sample counterpart of equation (7.19). But, even more importantly, the inverse problem (7.19) will be well-posed as soon as one maintains the following identification assumption about the news impact function φ

Assumption A: There exists no θ and $\varphi \in L_F^2(X)$ with $\varphi \neq 0$ such that $\sum_{j=1}^{\infty} a_j(\theta)\varphi(X_{t-j}) = 0$ almost certainly.

To see this, note that assumption A means that for any non-zero function φ

$$
0 < E\left[\sum_{j=1}^{\infty} a_j \varphi(X_{t-j})\right]^2
$$

that is

$$
\begin{aligned}
&0 < \textstyle\sum_{j=1}^{\infty} a_j^2 \langle \varphi, \varphi \rangle \\
&+ \textstyle\sum_{j=1}^{\infty} \sum_{\substack{l=1 \\ l \neq j}}^{\infty} a_l a_j \langle \varphi, H_{j-l}\varphi \rangle
\end{aligned}
$$

Therefore

$$
0 < \langle \varphi, \varphi \rangle - \langle \varphi, K\varphi \rangle \tag{21}
$$

for non zero φ. In other words, there is no non-zero φ such that

$$
K\varphi = \varphi
$$

and the operator $(I-K)$ is one-to-one. It is also worth noticing that the operator K is Hilbert-Schmidt and a fortiori compact under reasonable assumptions. As already mentioned in subsection 2.5.1, the Hilbert-Schmidt property for the conditional expectation operator H_k is tantamount to the integrability condition

$$\int\int\left[\frac{f_{X,X_{t-k}}(x,y)}{f_{X_{t(x)}}f_{X_t}(y)}\right]^2 f_{X_t}(x)\,f_{X_t}(y)\,dxdy<\infty$$

It amounts to say that there is not too much dependence between X_t and X_{t-k}. This should be tightly related to the ergodicity or mixing assumptions about the stationary process X_t. Then, if all the conditional expectation operator H_k, $k\geq 1$, are Hilbert-Schmidt, the operator K will also be Hilbert-Schmidt insofar as

$$\sum_{j=1}^{\infty}\sum_{l\neq j}a_j^2a_l^2<+\infty$$

Note that (21) implies that $(I-K)$ has eigenvalues bounded from below by a positive number.

Up to a straightforward generalization to stationary mixing processes of results only stated in the i.i.d. case, the general asymptotic theory of this subsection 7.3 can then be easily applied to nonparametric estimators of the new impact function φ based on the Fredholm equation of the second kind (7.19). An explicit formula for the asymptotic variance of $\hat{\varphi}_n$ as well as a practical implementation by solution of matricial equations similar to subsection 3.5 (without need of a Tikhonov regularization) is provided by Linton and Mammen (2003) in the particular case of volatility modelling.

However, an important difference with the i.i.d. case (see for instance assumption A.3 in section 5.4 about instrumental variables) is that the conditional homoskedasticity assumption cannot be maintained about conditional probability distribution of X_t given its own past. This should be particularly detrimental in the case of volatility modelling since, when X_t denotes a squared return, it will be in general even more conditionally heteroskedastic than returns themselves. Such a severe conditional heteroskedasticity will likely imply poor finite sample performance and large asymptotic variance of the estimator $\hat{\varphi}_n$ defined from the inverse problem (7.19), that is from the least squares problem (7.18). Indeed, $\hat{\varphi}_n$ is basically kind of OLS estimator in infinite dimension. In order to better take into account conditional heteroskedasticity of X_t in the context of volatility modelling, Linton and Mammen (2003) propose to replace the least squares problem (7.18) by a quasi-likelihood kind of approach where the criterion to optimize is defined from the density function of a normal conditional probability distribution of returns, with variance $m_\varphi(\theta,I_{t-1})$. Then, the difficulty is that the associated first order conditions now characterize the news impact function φ as solution of a nonlinear inverse problem. Linton and Mammen (2003) suggest to work with a version of this problem which is locally linearized around the previously described least squares estimator $\hat{\varphi}_n$ (and associated consistent estimator of θ).

4 Regularized solution of an ill posed equation of the second kind and statistical implications

The objective of this section is to study equations $(I-K)\varphi = r$ where 1 is an eigenvalue of K, i.e. where $I-K$ is not injective (or one-to-one). For simplicity we restrict our analysis to the case where the order of multiplicity of the eigenvalue 1 is one and the operator K is self-adjoint. This implies that the dimension of the null spaces of $I-K$ is one and using the results of Section 2, the space $\mathcal{H}$ may be decomposed into

$$\mathcal{H} = \mathcal{N}(I-K) \oplus \mathcal{R}(I-K)$$

i.e. $\mathcal{H}$ is the direct sum between the null space and the range of $I-K$, both closed. We denote by $P_{\mathcal{N}}r$ the projection of r on $\mathcal{N}(I-K)$ and by $P_{\mathcal{R}}r$ the projection of r on the range $\mathcal{R}(I-K)$.

Using ii) of Theorem 1, a solution of $(I-K)\varphi = r$ exists in the non injective case only if r is orthogonal to $\mathcal{N}(I-K)$ or, equivalently, if r belongs to $\mathcal{R}(I-K)$. In other words, a solution exists if and only if $r = P_{\mathcal{R}}r$. However in this case, this solution is not unique and there exists a one dimensional linear manifold of solutions. Obviously, if φ is a solution, φ plus any element of $\mathcal{N}(I-K)$ is again a solution. This non uniqueness problem will be solved by a normalization rule which selects a unique element in the set of solutions. The normalization we adopt is

$$\langle \varphi, \phi_0 \rangle = 0 \tag{22}$$

where ϕ_0 is the eigenfunction of K corresponding to the eigenvalue equal to 1.

In most statistical applications of equations of the second kind, the r element corresponding to the true data generating process is assumed to be in the range of $I-K$ where K is also associated with the true DGP. However this property is no longer true if F is estimated and we need to extend the resolution of $(I-K)\varphi = r$ to cases where $I-K$ is not injective and r is not in the range of this operator. This extension must be done in such a way that the continuity properties of inversion are preserved.

For this purpose we consider the following generalized inverse of $(I-K)$. As K is a compact operator it has a discrete spectrum $\lambda_0 = 1, \lambda_1, ...$ where only 0 may be an accumulation point (in particular 1 cannot be an accumulation point). The associated eigenfunctions are $\phi_0, \phi_1, ...$ Then we define:

$$Lu = \sum_{j=1}^{\infty} \frac{1}{1-\lambda_j} \langle u, \phi_j \rangle \phi_j, \quad u \in \mathcal{H} \tag{23}$$

This operator computes the unique solution of $(I-K)\varphi = P_{\mathcal{R}}u$ satisfying the normalization rule (22). It can be easily verified that L satisfies:

$$\begin{aligned} LP_{\mathcal{R}} &= L = P_{\mathcal{R}}L \\ L(I-K) &= (I-K)L = P_{\mathcal{R}} \end{aligned} \tag{24}$$

It can easily be checked that L is the generalized inverse of $I - K$ as it was defined in Luenberger (1969).

We now consider estimation. For an observed sample, we obtain the estimator F_n of F (that may be built from a kernel estimator of the density) and then the estimators $\hat{r}_n$ and $\hat{K}_n$ of r and K respectively. Let $\hat{\phi}_0$, $\hat{\phi}_1, ...$ denote the eigenfunctions of $\hat{K}_n$ associated with $\hat{\lambda}_0$, $\hat{\lambda}_1, ...$ We restrict our attention to the cases where 1 is also an eigenvalue of multiplicity one of $\hat{K}_n$ (i.e. $\hat{\lambda}_0 = 1$). However $\hat{\phi}_0$ may be different from ϕ_0.

We have to make a distinction between two cases. First assume that the Hilbert space $\mathcal{H}$ of reference is known and in particular the inner product is given (for example $\mathcal{H} = L^2(\mathbb{R}^p, \pi)$ with π given), the normalization rule imposed to $\hat{\varphi}_n$ is

$$\langle \hat{\varphi}_n, \hat{\phi}_0 \rangle = 0$$

and $\hat{L}_n$ is the generalized inverse of $I - \hat{K}_n$ in $\mathcal{H}$ (which depends on the Hilbert space structure) where

$$\hat{L}_n u = \sum_{j=1}^{\infty} \frac{1}{1 - \hat{\lambda}_j} \langle u, \hat{\phi}_j \rangle \hat{\phi}_j, \quad u \in \mathcal{H}$$

Formula (24) applies immediately for F_n.

If however the Hilbert space $\mathcal{H}$ depends on F (e.g. $\mathcal{H} = L^2(\mathbb{R}^p, F)$), we need to assume that $L^2(\mathbb{R}, F_n) \subset L^2(\mathbb{R}^p, F)$. The orthogonality condition which defines the normalization rule (22) is related to $L^2(\mathbb{R}^p, F)$ but the estimator $\hat{\varphi}_n$ of φ will be normalized by

$$\langle \hat{\varphi}_n, \hat{\phi}_0 \rangle_n = 0$$

where $\langle \ , \ \rangle_n$ denotes the inner product relative to F_n. This orthogonality is different from an orthogonality relative to $\langle \ , \ \rangle$.

In the same way $\hat{L}_n$ is now defined as the generalized inverse of $I - \hat{K}_n$ with respect to the estimated Hilbert structure, i.e.

$$\hat{L}_n u = \sum_{j=1}^{\infty} \frac{1}{1 - \hat{\lambda}_j} \langle u, \hat{\phi}_j \rangle_n \hat{\phi}_j$$

and $\hat{L}_n$ is not the generalized inverse of $I - \hat{K}_n$ in the original space $\mathcal{H}$. The advantages of this definition is that $\hat{L}_n$ may be effectively computed and satisfies the formula (24) where F_n replaces F. In the sequel $P_{\mathcal{R}_n}$ denotes the projection for the inner product $<, >_n$on $\mathcal{R}_n = \mathcal{R}\left(I - \hat{K}_n\right)$.

>From (24) one can deduce that:

$$\begin{aligned} \hat{L}_n - L &= \hat{L}_n(\hat{K}_n - K)L \\ &+ \hat{L}_n(P_{\mathcal{R}_n} - P_{\mathcal{R}}) + (P_{\mathcal{R}_n} - P_{\mathcal{R}})L \end{aligned} \tag{25}$$

since $\hat{L}_n - L = \hat{L}_n P_{\mathcal{R}_n} - P_{\mathcal{R}} L = \hat{L}_n(P_{\mathcal{R}_n} - P_{\mathcal{R}}) + (P_{\mathcal{R}_n} - P_{\mathcal{R}})L - P_{\mathcal{R}_n} L + \hat{L}_n P_r$

and $\hat{L}_n (K_n - K) L = \hat{L}_n (K_n - I) L + \hat{L}_n (I - K) L = P_{\mathcal{R}_n} L + \hat{L}_n P_r$.

The convergence property is given by the following theorem:

Theorem 5 *Let us define* $\varphi_0 = Lr$ *and* $\hat{\varphi}_n = \hat{L}_n \hat{r}_n$. *If*

i) $\left\| \hat{K}_n - K \right\| = o(1)$

ii) $\| P_{\mathcal{R}_n} - P_{\mathcal{R}} \| = O\left(\frac{1}{b_n}\right)$

iii) $\left\| (\hat{r}_n + \hat{K}_n \varphi_0) - (r + K\varphi_0) \right\| = O\left(\frac{1}{a_n}\right)$

Then

$$\|\hat{\varphi}_n - \varphi_0\| = O\left(\frac{1}{a_n} + \frac{1}{b_n}\right)$$

Proof. The proof is based on:

$$\begin{aligned}
\hat{\varphi}_n - \varphi_0 &= \hat{L}_n \hat{r}_n - Lr \\
&= \hat{L}_n(\hat{r}_n - r) + (\hat{L}_n - L)r \\
&= \hat{L}_n(\hat{r}_n - r) + \hat{L}_n(\hat{K}_n - K)\varphi_0 \\
&+ \hat{L}_n \left(P_{\mathcal{R}_n} - P_{\mathcal{R}}\right) r + (P_{\mathcal{R}_n} - P_{\mathcal{R}})\varphi_0
\end{aligned} \tag{26}$$

deduced from (25). Then

$$\begin{aligned}
\|\hat{\varphi}_n - \varphi_0\| &\leq \|\hat{L}_n\| \|(\hat{r}_n + \hat{K}_n\varphi_0) - (r + K\varphi_0)\| \\
&+ (\|\hat{L}_n\|\|r\| + \|\varphi\|)\|P_{\mathcal{R}_n} - P_{\mathcal{R}}\|
\end{aligned} \tag{27}$$

Under i) and ii) $\|\hat{L}_n - L\| = o(1)$ from (25). This implies $\|\hat{L}_n\| \to \|L\|$ and the result follows. ■

If $\frac{a_n}{b_n} \sim O(1)$, the actual speed of convergence is bounded by $\frac{1}{a_n}$. This will be the case in the two examples of 5 where $\frac{a_n}{b_n} \to 0$.

We consider asymptotic normality in this case. By (24), we have $\hat{L}_n = P_{\mathcal{R}_n} + \hat{L}_n \hat{K}_n$, hence:

$$\begin{aligned}
\hat{\varphi}_n - \varphi_0 &= P_{\mathcal{R}_n} \left[(\hat{r}_n + \hat{K}_n\varphi_0) - (r + K\varphi_0) \right] \\
&+ \hat{L}_n \hat{K}_n \left[(\hat{r}_n + \hat{K}_n\varphi_0) - (r + K\varphi_0) \right] \\
&+ \hat{L}_n(P_{\mathcal{R}_n} - P_{\mathcal{R}})r + (P_{\mathcal{R}_n} - P_{\mathcal{R}})\varphi_0
\end{aligned} \tag{28}$$

Let us assume that there exists a sequence a_n such that i) and ii) below are satisfied

i) $a_n P_{\mathcal{R}_n} \left[(\hat{r}_n + \hat{K}_n\varphi_0) - (r + K\varphi_0) \right](x)$ has an asymptotic normal distribution

ii) $a_n \left[\hat{L}_n \hat{K}_n (\hat{r}_n + \hat{K}_n \varphi_0 - r - K\varphi_0)\right](x) \rightarrow 0$, $a_n \left[\hat{L}_n (P_{\mathcal{R}_n} - P_{\mathcal{R}}) r\right](x) \rightarrow 0$ and $[(P_{\mathcal{R}_n} - P_{\mathcal{R}})\varphi_0](x) \rightarrow 0$

Then the asymptotic normality of $a_n(\hat{\varphi}_n - \varphi_0)$ is driven by the behavior of the first term of the decomposition (28). This situation occurs in the non parametric estimation as illustrated in the next section.

5 Two examples: backfitting estimation in additive models and panel model

5.0.1 Backfitting estimation in additive models

Let us recall that in an additive model defined by

$$\begin{array}{l} (Y, Z, W) \in \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^q \\ Y = \varphi(Z) + \psi(W) + U \\ E(U|Z, W) = 0, \end{array} \tag{29}$$

in which case (see 1.24), the function φ is solution of the equation:

$$\varphi - E\left[E(\varphi(Z)|W)|Z\right] = E(Y|Z) - E\left[E(Y|W)|Z\right]$$

and ψ is the solution of an equation of the same nature obtained by a permutation of W and Z. We focus our presentation on the estimation of φ. It appears as the resolution of a linear equation of the second kind. More precisely, we have in that case :

- $\mathcal{H}$ is the space of the square integrable functions of Z with respect to the true data generating process. This definition simplifies our presentation but an extension to different spaces is possible.
- The unknown function φ is an element of $\mathcal{H}$. Actually asymptotic considerations will restrict the class of functions φ by smoothness restrictions.
- The operator K is defined by $K\varphi = E\left[E(\varphi(Z)|W)|Z\right]$. This operator is self adjoint and we assume its compactness. This compactness may be obtained through the Hilbert Schmidt assumption A.1 of section 5.
- The function r is equal to $E(Y|Z) - E\left[E(Y|W)|Z\right]$.

The operator $I - K$ is not one-to-one because the constant functions belong to the null space of this operator. Indeed the additive model (29) does not identify φ and ψ. We introduce the following hypothesis which warrants that φ (and ψ) are exactly identified up to an additive constant or, equivalently, that the null space of $I - K$ only contains the constants.

Identification assumption. Z and W are measurably separated w.r.t. the distribution F i.e. a function of Z almost surely equal to a function of W is almost surely constant.

This assumption implies that if $\varphi_1, \varphi_2, \psi_1, \psi_2$ are such that $E(Y|Z, W) = \varphi_1(Z) + \psi_1(W) = \varphi_2(Z) + \psi_2(W)$ then $\varphi_1(Z) - \varphi_2(Z) = \psi_2(W) - \psi_1(W)$ which implies that $\varphi_1 - \varphi_2$ and $\psi_2 - \psi_1$ are a.s. constant. In terms of the null set of $I - K$ we have:

$$\begin{aligned} K\varphi &= \varphi \Longleftrightarrow E\left[E(\varphi\left(Z\right)|W)|Z\right] = \varphi\left(Z\right) \\ &\Longrightarrow E\left[\left(E\left[\varphi\left(Z\right)|W\right]\right)^2\right] \\ &= E\left[\varphi\left(Z\right) E\left(\varphi(Z)|W\right)\right] = E\left(\varphi^2\left(Z\right)\right). \end{aligned}$$

But, by Pythagore theorem:

$$\begin{aligned} \varphi(Z) &= E(\varphi\left(Z\right)|W) + \upsilon \\ E\left(\varphi^2\left(Z\right)\right) &= E\left(\left(E\left(\varphi\left(Z\right)|W\right)\right)^2\right) + E\upsilon^2 \end{aligned}$$

Then:

$$\begin{aligned} K\varphi &= \varphi \Longrightarrow \upsilon = 0 \\ &\Leftrightarrow \varphi(Z) = E\left[\varphi(Z)\,|W\right]. \end{aligned}$$

Then if φ is an element of the null set of $I - K$, φ is almost surely equal to a function of W and is therefore constant.

The eigenvalues of K are real positive and smaller than 1 except for the first one. We have $1 = \lambda_0 > \lambda_1 > \lambda_2 ... > .$[1] The eigenfunctions are such that $\phi_0 = 1$ and the condition $\langle \varphi, \phi_0 \rangle = 0$ means that φ has an expectation equal to zero. The range of $I - K$ is the set of functions with a mean equal to 0 and the projection of u, $P_{\mathcal{R}}u$, equals $u - E(u)$.

It should be noticed that under the hypothesis of additive model, r has zero mean and is then an element of $\mathcal{R}(I - K)$. Then a unique (up to the normalization condition) solution of the structural equation $(I - K)\varphi = r$ exists.

The estimation may be done by kernel smoothing. The joint density is estimated by

$$f_n(y, z, w) = \frac{1}{nc_n^{1+p+q}} \sum_{i=1}^{n} \omega\left(\frac{y - y_i}{c_n}\right) \omega\left(\frac{z - z_i}{c_n}\right) \omega\left(\frac{w - w_i}{c_n}\right) \tag{30}$$

and F_n is the c.d.f. associated to f_n. The estimated $\hat{K}_n$ operator verifies:

$$(\hat{K}_n\varphi)(z) = \int \varphi\left(u\right) \hat{a}_n\left(u, z\right) du \tag{31}$$

[1] Actually $K = T^*T$ when $T\varphi = E(\varphi|W)$ and $T^*\psi = E(\psi|Z)$ when ψ is a function of W. The eigenvalues of K correspond to the squared singular values of the T and T^* defined in Section 2.

where

$$\hat{a}_n(u,z) = \int \frac{\hat{f}_n(.,u,w)\,\hat{f}_n(.,z,w)}{\hat{f}_n(.,.,w)\,\hat{f}_n(.,z,.)}dw.$$

The operator $\hat{K}_n$ must be an operator from $\mathcal{H}$ to $\mathcal{H}$ (it is by construction an operator from $L^2_Z(F_n)$ into $L^2_Z(F_n)$). Therefore $\frac{\omega\left(\frac{z-z_\ell}{c_n}\right)}{\sum_\ell \omega\left(\frac{z-z_\ell}{c_n}\right)}$ must be square integrable w.r.t. F.

The estimation of r by $\hat{r}_n$ verifies

$$\hat{r}_n(z) = \frac{1}{\sum_{\ell=1}^n \omega\left(\frac{z-z_\ell}{c_n}\right)} \sum_{\ell=1}^n \left(y_\ell - \sum_{i=1}^n y_i \omega_{li}\right) \omega\left(\frac{z-z_\ell}{c_n}\right)$$

where $\omega_{li} = \dfrac{\omega\left(\dfrac{w_l - w_i}{c_n}\right)}{\sum_{i=1}^n \omega\left(\dfrac{w_l - w_i}{c_n}\right)}$.

The operator $\hat{K}_n$ has also 1 as the greatest eigenvalue corresponding to an eigenfunction equal to 1. Since F_n is a mixture of probabilities for which z and w are independent, the measurable separability between Z and W is fulfilled. Then the null set of $I - \hat{K}_n$ reduces to a.s. (w.r.t. F_n) constant functions. The generalized inverse of an operator depends on the inner product of the Hilbert space because it is defined as the function φ of minimal norm which minimizes the norm of $\hat{K}_n\varphi - \hat{r}_n$. The generalized inverse in the space $L^2_Z(F)$ cannot be used for the estimation because it depends on the actual unknown F. Then we construct $\hat{L}_n$ as the generalized inverse in $L^2_Z(F_n)$ of $I - \hat{K}_n$. The practical computation of $\hat{L}_n$ can be done by computing the n eigenvalues of $\hat{K}_n$, $\hat{\lambda}_0 = 1, ..., \hat{\lambda}_{n-1}$ and the n eigenfunctions $\hat{\phi}_0 = 1, \hat{\phi}_1, ..., \hat{\phi}_{n-1}$. Then

$$\hat{L}_n u = \sum_{j=1}^{n-1} \frac{1}{1-\hat{\lambda}_j} \left\{ \int u(z)\hat{\phi}_j(z)\hat{f}_n(z)dz \right\} \hat{\phi}_j \tag{32}$$

It can be easily checked that property (24) is verified where $P_{\mathcal{R}_n}$ is the projection (w.r.t. F_n) on the orthogonal of the constant function. This operator subtracts to any functions its empirical mean computed through the smoothed density :

$$P_{\mathcal{R}_n} u = u - \frac{1}{nc_n^p} \sum_i \int u(z)\omega\left(\frac{z-z_i}{c_n}\right) dz \tag{33}$$

The right hand side of the equation $(I - \hat{K}_n)\varphi = \hat{r}_n$ has a mean equal to 0 (w.r.t. F_n). Hence, this equation has a unique solution $\hat{\varphi}_n = \hat{L}_n\varphi_0$ which satisfies the normalization condition $\frac{1}{nc_n^p}\sum_i \int \hat{\varphi}_n(z)\omega\left(\frac{z-z_i}{c_n}\right) dz = 0$.

The general results of Section 4 apply.

1) Under very general assumptions, $\|\hat{K}_n - K\| \to 0$ in probability.

2) We have to check the properties of $P_{\mathcal{R}_n} - P_{\mathcal{R}}$

$$(P_{\mathcal{R}_n} - P_{\mathcal{R}})\varphi = \frac{1}{nc_n^p}\sum_i \int \varphi(z)\omega\left(\frac{z-z_i}{c_n}\right)dz - \int \varphi(z)f(z)dz$$

The asymptotic behavior of $\|(P_{\mathcal{R}_n}-P_{\mathcal{R}})\varphi\|^2 = \left|\frac{1}{nc_n^p}\sum_{i=1}^n \int \varphi(z)\omega\left(\frac{z-z_i}{c_n}\right)dz - E(\varphi)\right|^2$ is the same as the asymptotic behavior of the expectation of this positive random variable :

$$E\left(\frac{1}{nc_n^p}\sum_{i=1}^n \int \varphi(z)\omega\left(\frac{z-z_i}{c_n}\right)dz - E(\varphi)\right)^2$$

Standard computation on this expression shows that this mean square error is $O\left(\frac{1}{n} + c_n^{2min(d,d')}\right)\|\varphi\|^2$, where d is the smoothness degree of φ and d' the order of the kernel.

3) The last term we have to consider is actually not computable but its asymptotic behavior is easily characterized. We simplify the notation by denoting $E^{F_n}(.|.)$ the estimation of a conditional expectation. The term we have to consider is

$$\begin{array}{rcl}(\hat{r}_n + \hat{K}_n\varphi) - (r + K\varphi) & = & E^{F_n}(Y|Z) - E^{F_n}(E^{F_n}(Y|W)|Z) + E^{F_n}(E^{F_n}(\varphi(Z)|W)|Z) \\ & - & E^F(Y|Z) + E^F(E^F(Y|W)|Z) - E^F(E^F(\varphi(Z)|W)|Z) \\ & = & E^{F_n}\left(Y - E^F(Y|W) + E^F(\varphi(Z)|W)|Z\right) \\ & - & E^F\left(Y - E^F(Y|W) + E^F(\varphi(Z)|W)|Z\right) \\ & - & R\end{array}$$

where $R = E^F\left\{E^{F_n}(Y - \varphi(Z)|W) - E^F(Y - \varphi(Z)|W)\right\}$

1. Moreover

$$E^F(Y|W) = E^F(\varphi(Z)|W) + \psi|W)$$

Then

$$\begin{array}{rcl}(r_n + K_n\varphi) - (r + K\varphi) & = & E^{F_n}(Y - \psi(W)|Z) - E^F(Y - \psi(W)|Z) \\ & - & R\end{array}$$

The R term converges at a faster speed than the first part of the r.h.s. of this equation and can be neglected.

We have seen in the other parts of this chapter that

$$\|E^{F_n}(Y - \psi(W)|Z) - E^F(Y - \psi(W)|Z)\|^2 \sim 0\left(\frac{1}{nc_n^p} + c_n^{2\rho}\right)$$

where ρ depends on the regularity assumptions.

We can conclude that $\|\hat{\varphi}_n - \varphi_0\| \to 0$ in probability and that $\|\hat{\varphi}_n - \varphi_0\| \sim 0\left(\frac{1}{\sqrt{nc_n^p}} + c_n^\rho\right)$.

The pointwise asymptotic normality is now easy to verify. Consider $\sqrt{nc_n^\rho}(\hat{\varphi}_n(z) - \varphi_0(z))$. We adapt in this framework the formula (28) and Theorem 4.

1) Under a suitable condition on c_n (typically $nc_n^{\rho+2\min(d,r)} \to 0$), we have: $\sqrt{nc_n^p}\left\{\hat{L}_n(P_{\mathcal{R}_n} - P_{\mathcal{R}})r + (P_{\mathcal{R}_n} - P_{\mathcal{R}})\varphi\right\} \to 0$ in probability.

2) Using the same argument as in 4, a suitable choice of c_n implies that

$$\sqrt{nc_n^\rho}\hat{L}_n\hat{K}_n\left[(\hat{r}_n + \hat{K}_n\varphi_0) - (r + K\varphi_0)\right] \to 0$$

Actually, while $E^{F_n}(Y - \psi(W)|Z) - E^F(Y - \psi(W)|Z)$ only converges pointwise at a non parametric speed, the transformation by the operator $\hat{K}_n$ transforms this convergence into a functional convergence at a parametric speed. Then

$$\sqrt{nc_n^p}\left\|\hat{K}_n\left[E^{F_n}(Y - \psi(W)|Z) - E^F(Y - \psi(W)|Z\right]\right\| \to 0$$

Moreover $\hat{L}_n$ converge in norm to L which is a bounded operator and the result follows.

3) The convergence of $\sqrt{nc_n^p}(\varphi_{F_n}(z) - \varphi_F(z))$ is then identical to the convergence of

$$\sqrt{nc_n^p}P_{\mathcal{R}_n}\left[E^{F_n}(Y - \psi(W)|Z) - E^F(Y - \psi(W)|Z\right]$$

$$= \sqrt{nc_n^p}\Bigg[E^{F_n}(Y - \psi(W)|Z) - E^F(Y - \psi(W)|Z)$$

$$-\tfrac{1}{n}\sum_i(Y_i - \psi(W_i)) - \tfrac{1}{nc_n^p}\sum_i\int(Y - \psi(W))f(Y,W|Z)\omega\left(\frac{z - z_i}{c_n}\right)dz\Bigg]$$

Then also it can be easily checked that the difference between the two sample means converge to zero at a higher speed than $\sqrt{nc_n^p}$ and these two last terms can be cancelled. Then using standard results on nonparametric estimation, we obtain:

$$\sqrt{nc_n^p}(\varphi_{F_n}(z) - \varphi_F(z)) \xrightarrow{d} \mathcal{N}\left(0, \frac{\int\omega^2}{f_Z(z)}Var(Y - \psi(W)|Z = z)\right)$$

where the 0 mean of the asymptotic distribution is obtained thanks to a suitable choice of the bandwidth, which needs to converge to 0 faster than the optimal speed.

5.0.2 Estimation of the bias function in a measurement error equation

We have introduced in Example 1.3.6, Section 1, the measurement error model:

$$\begin{cases} Y_1 = \eta + \varphi(Z_1) + U_1 & Y_1, Y_2 \in \mathbb{R} \\ Y_2 = \eta + \varphi(Z_2) + U_2 & Z_1, Z_2 \in \mathbb{R}^p \end{cases}$$

when η, U_i are random unknown elements and Y_1 and Y_2 are two measurements of η contaminated by a bias term depending on observable elements Z_1 and Z_2. The unobservable component η is eliminated by difference and we get the model under consideration :

$$Y = \varphi(Z_2) - \varphi(Z_1) + U \tag{34}$$

when $Y = Y_2 - Y_1$ and $E(Y|Z_1, Z_2) = \varphi(Z_2) - \varphi(Z_1)$. We assume that i.i.d. observations of (Y, Z_1, Z_2) are available. Moreover the order of measurements is arbitrary or equivalently (Y_1, Y_2, Z_1, Z_2) is distributed identically to (Y_2, Y_1, Z_2, Z_1). This implies that (Y, Z_1, Z_2) and $(-Y, Z_2, Z_1)$ have the same distribution. In particular, Z_1 and Z_2 are identically distributed.

- The reference space $\mathcal{H}$ is the space of random variables defined on R^p that are square integrable with respect to the true marginal distribution on Z_1 (or Z_2). We are in a case where the Hilbert space structure depends on the unknown distribution
- The function φ is an element of $\mathcal{H}$ but this set has to be reduced by smoothness condition in order to obtain asymptotic properties of the estimation procedure.
- The operator K is the conditional expectation operator

  $$\begin{aligned} (K\varphi)(z) &= E^F(\varphi(Z_2)|Z_1 = z) \\ &= E^F(\varphi(Z_1)|Z_2 = z) \end{aligned}$$

  from $\mathcal{H}$ to $\mathcal{H}$. The two conditional expectations are equal because (Z_1, Z_2) and (Z_2, Z_1) are identically distributed (by the exchangeability property). This operator is self-adjoint and we suppose that K is compact. This property may be deduced as in previous cases from an Hilbert Schmidt argument.

Equation (34) introduces an overidentification property because it constrains the conditional expectation of Y given Z_1 and Z_2. In order to define φ for any F (and in particular for the estimated one), the parameter φ is now defined as the solution of the minimization problem:

$$\varphi = \arg\min_{\varphi} E\left(Y - \varphi(Z_2) + \varphi(Z_1)\right)^2$$

or, equivalently as the solution of the first-order conditions:

$$E^F\left[\varphi(Z_2)|Z_1 = z\right] - \varphi(z) = E(Y|Z_1 = z)$$

because $(Y, Z_1, Z_2) \sim (-Y, Z_2, Z_1)$.

Then the integral equation which defines the functions of interest φ may be denoted by

$$(I - K)\,\varphi = r$$

where $r = E(Y|Z_2 = z) = -E(Y|Z_1 = z)$. As in the additive models, this inverse problem is ill-posed because $I - K$ is not one-to-one. Indeed, 1 is the greatest eigenvalue of K and the eigenfunctions associated with 1 are the constant functions. We need an extra assumption to warranty that the order of multiplicity is one, or, in more statistical terms, that φ is identified up to a constant. This property is obtained if Z_1 and Z_2 are measurably separated i.e. if the functions of Z_1 almost surely equal to some functions of Z_2 are almost surely constant.

Then the normalization rule is

$$\langle \varphi, \phi_0 \rangle = 0$$

where ϕ_0 is constant. This normalization is then equivalent to

$$E^F(\varphi) = 0.$$

If F is estimated using standard kernel procedure, the estimated F_n does not satisfied, in general, the exchangeability assumption ((Y, Z_1, Z_2) and $(-Y, Z_2, Z_1)$ are identically distributed). A simple way to incorporate this constraint is to estimate F using a sample of size $2n$ by adding to the original sample $(y_i, z_{1i}, z_{2i})_{i=1,\ldots,n}$ a new sample $(-y_i, z_{2i}, z_{1i})_{i=1,\ldots,n}$. For simplicity we do not follow this method here and we consider an estimation of F which does not verify the exchangeability. In that case $\hat{r}_n$ is not, in general, an element of $\mathcal{R}\left(I - \hat{K}_n\right)$ and the estimator $\hat{\varphi}_n$ is defined as the unique solution of

$$\left(I - \hat{K}_n\right)\varphi = P_{\mathcal{R}_n}\hat{r}_n,$$

which satisfies the normalization rule

$$E^{F_n}(\varphi) = 0.$$

Equivalently we have seen that the functional equation $\left(I - \hat{K}_n\right)\varphi = \hat{r}_n$ reduces to a n dimensional linear system, which is solved by a generalized inversion. The asymptotic properties of this procedure follows immediately from the theorems of Section 4 and are obtained identically to the case of additive models.

References

[1] Ai, C. and X. Chen (1999) "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions", mimeo, New York University.

[2] Aronszajn, N. (1950) "Theory of Reproducing Kernels", *Transactions of the American Mathematical Society*, Vol. 68, 3, 337-404.

[3] Basmann, R.L. (1957), *A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equations*, Econometrica, 25, 77-83.

[4] Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics* 31, 307-327.

[5] Bosq, D. (1998) "Nonparametric Statistics for Stochastic Processes. Estimation and Prediction", Springer, Berlin.

[6] Carrasco, M., M. Chernov, J.-P. Florens, and E. Ghysels (2001) "Efficient estimation of jump diffusions and general dynamic models with a continuum of moment conditions", mimeo, University of Rochester.

[7] Carrasco, M. and J. P. Florens (2000) "Generalization of GMM to a continuum of moment conditions", *Econometric Theory*, 16, 797-834.

[8] Carrasco, M. and J. P. Florens (2001) "Efficient GMM Estimation Using the Empirical Characteristic Function", mimeo, University of Rochester.

[9] Carrasco, M. and J.-P. Florens (2002a) "Spectral method for deconvolving a density", mimeo, University of Rochester.

[10] Carrasco, M. and J. P. Florens (2002b) "On the Asymptotic Efficiency of GMM", mimeo, University of Rochester.

[11] Carroll, R. and P. Hall (1988) "Optimal Rates of Convergence for Deconvolving a Density", *Journal of American Statistical Association*, 83, No.404, 1184-1186.

[12] Carroll, R., A. Van Rooij, and F. Ruymgaart (1991) "Theoretical Aspects of Ill-posed Problems in Statistics", *Acta Applicandae Mathematicae*, 24, 113-140.

[13] Chacko, G. and L. Viceira (2000) "Spectral GMM Estimation of Continuous-Time Processes", forthcoming in *Journal of Econometrics.*

[14] Chen, X., L.P. Hansen and J. Scheinkman (1998) "Shape-preserving Estimation of Diffusions", mimeo, University of Chicago.

[15] Chen, X., and H. White (1992), *Central Limit and Functional Central Limit Theorems for Hilbert Space-Valued Dependent Processes*, Working Paper, University of San Diego.

[16] Chen, X. and H. White (1996) “Law of Large Numbers for Hilbert Space-Valued mixingales with Applications”, *Econometric Theory,* 12, 284-304.

[17] Chen, X. and H. White (1998) “Central Limit and Functional Central Limit Theorems for Hilbert Space-Valued Dependent Processes”, *Econometric Theory,* 14, 260-284.

[18] Darolles, S., J.P. Florens, and C. Gourieroux (2000) “Kernel Based Nonlinear Canonical Analysis and Time Reversibility”, forthcoming in *Journal of Econometrics.*

[19] Darolles, S., J.P. Florens, and E. Renault (1998), “Nonlinear Principal Components and Inference on a Conditional Expectation Operator with Applications to Markov Processes”, presented at Garchy, France, September 1998.

[20] Darolles, S., J.P. Florens, and E. Renault (2002), “Nonparametric Instrumental Regression”, Working paper 05-2002, CRDE.

[21] Dautray, R. and J.-L. Lions (1988) *Analyse mathématique et calcul numérique pour les sciences et les techniques.* Vol. 5. Spectre des opérateurs, Masson, Paris.

[22] Davidson, J. (1994) *Stochastic Limit Theory*, Oxford University Press, Oxford.

[23] Debnath, L. and P. Mikusinski (1999) *Introduction to Hilbert Spaces with Applications*, Academic Press. San Diego.

[24] Dunford, N. and J. Schwartz (1988) *Linear Operators, Part II: Spectral Theory*, Wiley, New York.

[25] Engle R.F., (1990), Discussion: Stock Market Volatility and the Crash of ’87, *Review of Financial Studies* 3, 103-106.

[26] Engle, R.F., D.F. Hendry and J.F. Richard, (1983), “Exogeneity”, *Econometrica*, 51 (2) 277-304.

[27] Engle, R.F., and V.K. Ng (1993), “Measuring and Testing the Impact of News on Volatility”, *The Journal of Finance* XLVIII, 1749-1778.

[28] Fan, J. (1993) “Adaptively local one-dimentional subproblems with application to a deconvolution problem”, *The Annals of Statistics*, 21, 600-610.

[29] Feuerverger, A. and P. McDunnough (1981), “On the Efficiency of Empirical Characteristic Function Procedures”, *J. R. Statist. Soc.* B, 43, 20-27.

[30] Florens, J.P., J. Heckman, C. Meghir and E. Vytlacil (2002), “ Instrumental Variables, Local Instrumental Variables and Control Functions”, Manuscript, University of Toulouse.

[31] Florens, J.P. and Malavolti (2002) "Instrumental Regression with Discrete Variables", mimeo University of Toulouse, presented at ESEM 2002, Venice.

[32] Florens, J.P. and M. Mouchart (1985), "Conditioning in Dynamic Models", *Journal of Time Series Analysis*, 53 (1), 15-35.

[33] Florens, J.P., M. Mouchart, and J.F. Richard (1974), *Bayesian Inference in Error-in-variables Models*, Journal of Multivariate Analysis, 4, 419-432.

[34] Florens, J.P., M. Mouchart, and J.F. Richard (1987), *Dynamic Error-in-variables Models and Limited Information Analysis*, Annales d'Economie et Statistiques, 6/7, 289-310.

[35] Florens, J.P., C. Protopopescu, and J.F. Richard, (1997), "Identification and Estimation of a Class of Game Theoretic Models", GREMAQ-University of Toulouse.

[36] Forni, M. and L. Reichlin (1998) "Let's Get Real: A Factor Analytical Approach to Disaggregated Business Cycle Dynamics", *Review of Economic Studies*, 65, 453-473.

[37] Gallant, A. R. and J. R. Long (1997) "Estimating Stochastic Differential Equations Efficiently by Minimum Chi-squared", *Biometrika*, 84, 125-141.

[38] Gaspar, P. and J.P. Florens, (1998), "Estimation of the Sea State Biais in Radar Altimeter Measurements of Sea Level: Results from a Nonparametric Method", *Journal of Geophysical Research*, 103 (15), 803-814.

[39] Guerre, E., I. Perrigne, and Q. Vuong, (2000), "Optimal Nonparametric Estimation of First-Price Auctions", *Econometrica*, 68 (3), 525-574.

[40] Hansen, L., (1982), "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica*, 50, 1029-1054.

[41] Hansen, L. (1985) "A method for calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators", *Journal of Econometrics*, 30, 203-238.

[42] Hardle, W. and O. Linton (1994) "Applied Nonparametric Methods", *Handbook of Econometrics*, Vol. IV, edited by R.F. Engle and D.L. McFadden, North Holland, Amsterdam.

[43] Hastie, T.J. and R.J. Tibshirani (1990), *Generalized Additive Models*, Chapman and Hall, London.

[44] Hausman, J., (1981), "Exact Consumer's Surplus and Deadweight Loss" *American Economic Review*, 71, 662-676.

[45] Hausman, J. (1985), "The Econometrics of Nonlinear Budget sets" *Econometrica*, 53, 1255-1282.

[46] Hausman, J. and W.K. Newey, (1995) "Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss", *Econometrica*, 63, 1445-1476.

[47] Heckman, J., H. Ichimura, J. Smith, and P. Todd (1998), *Characterizing Selection Bias Using Experimental Data*, Econometrica, 66, 1017-1098.

[48] Heckman, J., and V. Vytlacil (2000), *Local Instrumental Variables*, in Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya, ed. by C. Hsiao, K. Morimune, and J. Powells. Cambridge: Cambridge University Press, 1-46.

[49] Hoerl, A. E. and R. W. Kennard (1970) "Ridge Regression: Biased Estimation of Nonorthogonal Problems", *Technometrics*, 12, 55-67.

[50] Imbens, G., and J. Angrist (1994), *Identification and Estimation of Local Average Treatment Effects*, Econometrica, 62, 467-476.

[51] Jiang, G. and J. Knight (2000) "Estimation of Continuous Time Processes Via the Empirical Characteristic Function", forthcoming in *Journal of Business & Economic Statistics.*

[52] Judge, G., W. Griffiths, R. C. Hill, H. Lutkepohl, and T-C. Lee (1980) *The Theory and Practice of Econometrics*, John Wiley and Sons, New York.

[53] Kitamura, Y. and M. Stutzer (1997), "An Information Theoretic Alternative to Generalized Method of Moments Estimation", *Econometrica*, 65, 4, 861-874.

[54] Knight, J. L. and J. Yu (2002) "Empirical Characteristic Function in Time Series Estimation", *Econometric Theory*, 18, 691-721.

[55] Kress, R. (1999), *Linear Integral Equations,* Springer.

[56] Kutoyants, Yu. (1984), *Parameter estimation for stochastic processes,* Heldermann Verlag, Berlin.

[57] Lancaster, H. (1968), *The Structure of Bivariate Distributions,* Ann. Math. Statist., 29, 719-736.

[58] Linton, O. and E. Mammen (2003), "Estimating Semiparametric ARCH(∞) models by kernel smoothing methods", Discussion Paper, LSE. Invited Paper ESEM 2003, Stockholm.

[59] Loubes, J.M. and A. Vanhems (2001), "Differential Equation and Endogeneity", Discussion Paper, GREMAQ, University of Toulouse, presented at ESEM 2002, Venice.

[60] Loubes, J.M. and A. Vanhems (2003), "Saturation Spaces for Regularization Methods in Inverse Problems", Discussion Paper, GREMAQ, University of Toulouse, presented at ESEM 2003, Stockholm.

[61] Lucas, R. (1978) "Asset Prices in an Exchange Economy", *Econometrica*, 46, 1429-1446.

[62] Luenberger, D. G. (1969) *Optimization by Vector Space Methods*, Wiley, New York.

[63] Malinvaud, E. (1970), *Methodes Statistiques de l'Econométrie*, Dunod, Paris.

[64] Nashed, N. Z. and G. Wahba (1974) "Generalized inverses in reproducing kernel spaces: An approach to regularization of linear operator equations", *SIAM J. Math. Anal.* 5, 974-987.

[65] Newey, W., and J. Powell (2000), *Instrumental Variables for Nonparametric Models,* MIT Discussion Paper.

[66] Newey, W., Powell, J. and F. Vella (1999), *Nonparametric Estimation of Triangular Simultaneous Equations Models*, Econometrica, **67**, 565-604.

[67] Owen, A. (2001) *Empirical likelihood,* Monographs on Statistics and Applied Probability, vol. 92. Chapman and Hall, London.

[68] Pagan, A. and A. Ullah (1999), *Nonparametric Econometrics*, Cambridge University Press.

[69] Parzen, E. (1959) "Statistical Inference on time series by Hilbert Space Methods,I.", Technical Report No.23, Applied Mathematics and Statistics Laboratory, Stanford. Reprinted in (1967) *Time series analysis papers*, Holden-Day, San Francisco.

[70] Parzen, E. (1970) "Statistical Inference on time series by RKHS methods", Proc. 12th Biennal Canadian Mathematical Seminar, R. Pyke, ed. American Mathematical Society, Providence.

[71] Politis, D. and J. Romano (1994) "Limit theorems for weakly dependent Hilbert space valued random variables with application to the stationary bootstrap", Statistica Sinica, 4, 451-476.

[72] Qin, J. and J. Lawless, (1994), "Empirical Likelihood and General Estimating Equations", *The Annals of Statistics*, 22, 1, 300-325.

[73] Reiersol, O. (1941), *Confluence Analysis of Lag Moments and other Methods of Confluence Analysis*, Econometrica, 9, 1-24.

[74] Reiersol, O. (1945), *Confluence Analysis by Means of Instrumental Sets of Variables*, Arkiv for Mathematik, Astronomie och Fysik, 32.

[75] Rust, J., J. F. Traub, and H. Wozniakowski (2002) "Is There a Curse of Dimensionality for Contraction Fixed Points in the Worst Case?", *Econometrica*, 70, 285-330.

[76] Ruymgaart, F. (2001) "A short introduction to inverse statistical inference", lecture given at the conference "L'Odyssée de la Statistique", Institut Henri Poincaré, Paris.

[77] Saitoh, S. (1997) *Integral transforms, reproducing kernels and their applications*, Longman.

[78] Sansone, G. *Orthogonal Functions*, Dover Publications, New York.

[79] Sargan, J.D. (1958), *The Estimation of Economic Relationship using Instrumental Variables,* Econometrica, 26, 393-415.

[80] Singleton, K. (2001) "Estimation of Affine Pricing Models Using the Empirical Characteristic Function", *Journal of Econometrics*, 102, 111-141.

[81] Stefanski, L. and R. Carroll (1990) "Deconvoluting Kernel Density Estimators", *Statistics*, 2, 169-184.

[82] Stock, J. and M. Watson (1998) "Diffusion Indexes", NBER working paper 6702.

[83] Tauchen, G. (1997) "New Minimum Chi-Square Methods in Empirical Finance", in *Advances in Econometrics*, *Seventh World Congress*, eds. D. Kreps and K. Wallis, Cambridge University Press, Cambridge.

[84] Tauchen, G. and R. Hussey (1991) "Quadrature-Based Methods for Obtaining Approximate Solutions to Nonlinear Asset Pricing Models", *Econometrica*, 59, 371-396.

[85] Theil, H.(1953), *Repeated Least Squares Applied to complete Equations System*, The Hague: Central Planning Bureau (mimeo).

[86] Vanhems, A. (2000), "Nonparametric Solutions to Random Ordinary Differential Equations of First Orders", GREMAQ-University of Toulouse.

[87] Van Rooij and F. Ruymgaart (1991) "Regularized Deconvolution on the Circle and the Sphere", in *Nonparametric Functional Estimation and Related Topics*, edited by G. Roussas, 679-690, Kluwer Academic Publishers, the Netherlands.

[88] Van Rooij, A., F. Ruymgaart, and W. Van Zwet (2000) "Asymptotic Efficiency of Inverse Estimators", *Theory Probab. Appl.*, 44, 4, 722-738.

[89] Vapnik A.C.M. (1998), *Statistical Learning Theory*, Wiley, New York.