Ann. Inst. Statist. Math.
36 (1984), Part A, 403-417

# KERNEL ESTIMATION AND INTERPOLATION FOR TIME SERIES CONTAINING MISSING OBSERVATIONS

P. M. ROBINSON

(Received Mar. 7, 1983; revised Mar. 12, 1984)

## Summary

Kernel estimators of conditional expectations are adapted for use in the analysis of stationary time series containing missing observations. Estimators of conditional expectations at fixed points are shown to have an asymptotic distribution with a relatively simple variance-covariance structure. The kernel method is also used to interpolate missing observations, and is shown to converge in probability to the least squares predictor. The results are established under the strong mixing condition and moment conditions, and the methods are applied to a real data set.

## 1. Introduction

Time series data sometimes contain missing observations. Recording of a discrete-time stochastic process $\{X_t;\ t=0, \pm 1, \cdots\}$ commences at time $t=1$ and ends at $t=T$, but this stretch of length $T$ includes gaps. Observations can be missed for a variety of reasons, such as clerical error, malfunction of recording equipment, deletion of apparently "bad" observations, and the inability to observe the process at certain times, for example at night-time or on weekends. Most methods of time series analysis are ideal only for Gaussian data, and even here the presence of missing observations causes complications. There has been little attempt to develop methods of analyzing non-Gaussian series with missing values by explicitly non-Gaussian methods, and this paper attempts to fill this gap.

We assume throughout that $X_t$ is strictly stationary and ergodic, so any necessary detrending has been carried out in advance. The method of trend estimation of Akaike and Ishiguro [2] seems particularly convenient when missing values occur. Following Parzen [10], introduce a sequence $b_t$, $t=1, 2, \cdots$, such that

$$b_t = \begin{cases} 1, & \text{if } X_t \text{ is observed}, \\ 0, & \text{if } X_t \text{ is missed}. \end{cases}$$

Throughout we assume $b_t$ is independent of $X_s$ for all $t$, $s$, so $X_t$ does not go missing on account of the value it takes. Define

$$b(j) = T^{-1} \sum_{t=j+1}^{T} b_t b_{t-j}, \qquad 0 \leqq j \leqq T-1; \; b = b(0) = T^{-1} \sum_{t=1}^{T} b_t .$$

The best predictor, in the least squares sense, of $X_t$ given $X_{t-j} = x$ (for $j$ positive or negative) is the conditional expectation

$$\nu_j(x) = \mathrm{E}(X_t \mid X_{t-j} = x) .$$

For non-Gaussian processes, such as non-linear AR ones, $\nu_j(x)$ need not be linear in $x$, and need not depend on only the second-order properties of the process. Therefore nonparametric estimation of $\nu_j(x)$ for various $j$, $x$ should provide detailed information on structure without imposition of implicit assumptions of linearity or Gaussianity. In addition a missing $X_{m+j}$ can be interpolated by a nonparametric "estimate" of $\nu_j(X_m)$, which hopefully will capture any nonlinearity in the process.

Let $K(x)$ be a real integrable function and

$$c(x) = (Th)^{-1} \sum_{t=1}^{T} b_t K((X_t - x)h^{-1}) ,$$

$$c_j(x) = (Th)^{-1} \sum_{t=j+1}^{T} b_t b_{t-j} X_t K((X_{t-j} - x)h^{-1}) , \qquad j > 0 ,$$

where $h$ is a positive "bandwidth" parameter regarded as decreasing as $T$ increases. We estimate $\nu_j(x)$ by

$$\hat{\nu}_j(x) = c_j(x) / \{b(j) \hat{f}(x)\} , \tag{1}$$

where

$$\hat{f}(x) = c(x)/b \tag{2}$$

estimates the probability density function (pdf) of $X_t$, $f(x)$, which we assume throughout exists but is unknown, and estimation of $f(x)$ is itself of interest particularly as $\nu_j(x)$ might be linear but $X_t$ non-Gaussian (as in linear AR processes with non-Gaussian innovations). Given that $X_m$ is observed, the formula (1) can be applied to interpolate a missing $X_{m+j}$ by $\hat{\nu}_j(X_m)$. With no loss of generality we assume throughout that

$$\int_R K(u) du = 1 .$$

For example, we have the "Gaussian" kernel

$$(3) \qquad K(x)=(2\pi)^{-1/2}\exp(-x^2/2), \qquad x\in R.$$

The methods described here are in principle readily extendable to estimation of higher-order conditional expectations such as

$$(4) \qquad \mathrm{E}(X_t \mid X_{t-j}, \cdots, X_{t-p-j}),$$

as well as joint pdfs of $X_t, \cdots, X_{t-p}$ and conditional pdfs.

Masry [9] has recently studied probability density estimation for a continuous process from random sampling. The references [1], [3], [11], [13]-[16], [20], [21] are among those which are concerned with nonparametric kernel and other estimators from equally-spaced time series data.

## 2. Asymptotic normality of conditional expectation estimator

In order to build up a comprehensive picture of the functions $\nu_j(x)$ they will be estimated, for $j=1, \cdots, q$, over a grid of distinct points $x=x_1, \cdots, x_r$. Of interest, therefore, is the multivariate CLT for the $qr$-dimensional vector $\{\hat{\nu}(x_1)^\tau, \cdots, \hat{\nu}(x_r)^\tau\}$, where

$$\hat{\nu}(x_i)=(\hat{\nu}_1(x_i), \cdots, \hat{\nu}_q(x_i))^\tau, \qquad i=1, \cdots, r.$$

For an open set $S\subset R$ we assume

$$f(x)>0, \qquad x\in S,$$

$$(5) \qquad f(x)\in \mathrm{Lip}(\lambda, S), \quad \nu_j(x)\in \mathrm{Lip}(\lambda, S), \qquad 0<\lambda<1, \ j=1, \cdots, q.$$

The joint pdf of $X_t$, $X_{t+u}$ exists and is continuous and bounded on $S\times S$, the bound being uniform in $u$. For some $\gamma>\delta>0$,

$$(6) \qquad \mathrm{E}|X_t|^{2+\delta}\leqq\infty,$$

$$(7) \qquad \mathrm{E}(|X_t|^{2+\gamma}|X_{t-j}=x) \text{ is bounded on } S, \ j=1, \cdots, q.$$

Moreover,

$$\omega_{jk}(x)\triangleq\mathrm{E}(X_tX_{t-j}\mid X_{t-k}=x) \text{ is continuous on } S, \ 0\leqq j<k\leqq q.$$

We assume $X_t$ is strongly mixing (e.g. Ibragimov and Linnik [8], pp. 305–306) with mixing coefficient $\alpha_k$ satisfying

$$(8) \qquad \sum_{k=n}^{\infty}\alpha_k^{\delta/(2+\delta)}=O(n^{-1}), \qquad \text{as } n\to\infty,$$

with $\delta$ the same as in (6). For $0\leqq j\leqq k\leqq q$ we define

$$b(j, k)=T^{-1}\sum_{t=k+1}^{T} b_t b_{t-j} b_{t-k} ,$$

and impose an "asymptotic stationarity" condition on $b_t$ (cf. [10]), assuming the fixed probability limits

$$\text{(9)} \qquad \operatorname{p\,lim}_{T\to\infty} b(j, k) \triangleq \beta(j, k) , \qquad 0\leqq j\leqq k\leqq q ,$$

exist. Note that $b=b(0, 0)$, $b(j)=b(0, j)$, and assume

$$\text{(10)} \qquad \beta(j)\triangleq\beta(j, j)>0 , \qquad 1\leqq j\leqq q ,$$

(thus $\beta\triangleq\beta(0, 0)>0$). As far as $h=h_T$ and $K$ are concerned we require

$$\text{(11)} \qquad hT\to\infty , \quad h^{1+2\lambda}T\to 0 , \qquad \text{as } T\to\infty ,$$

$$\text{(12)} \qquad |K(x)|\leqq C(1+|x|)^{-1-\lambda} ,$$

where $\lambda$ is as in (5), and $C$ denotes a generic constant throughout the paper. Finally define

$$\nu(x)=\{\nu_1(x), \cdots, \nu_q(x)\}^{\tau} , \qquad \kappa=\int_R K(x)^2 dx .$$

THEOREM 1. *Let the above conditions hold. Then for any distinct* $x_i \in S$ *the vectors*

$$(Th)^{1/2}\{\hat{\nu}(x_i)-\nu(x_i)\} , \qquad i=1,\cdots, r ,$$

*converge as* $T\to\infty$ *to independent q-dimensional normal variables with zero means and covariance matrices with* $(j, k)$*th elements* $(j\leqq k)$

$$\text{(13)} \qquad \kappa f(x_i)^{-1}\{\beta(j)^{-1}\beta(k)^{-1}\beta(k-j, k)\omega_{k-j,k}(x_i)-\beta^{-1}\nu_j(x_i)\nu_k(x_i)\} .$$

*A consistent estimator of* (13) *is*

$$\text{(14)} \qquad \kappa \hat{f}(x_i)^{-1}[\{b(j)b(k)\hat{f}(x_i)\}^{-1}w_{k-j,k}(x_i)-b^{-1}\hat{\nu}_j(x_i)\hat{\nu}_k(x_i)] ,$$

*where*

$$w_{k-j,k}(x)=(Th)^{-1}\sum_t b_t b_{t+j-k} b_{t-k} X_t X_{t+j-k} K((X_{t-k}-x)h^{-1}) .$$

Thus estimators at distinct points are independent though estimators for different lags at the same point are correlated. The same is true when there are no missing data, $b_t\equiv 1$, when (13) reduces to $\kappa f(x_i)^{-1} \cdot \{\omega_{k-j,k}(x_i)-\nu_j(x_i)\nu_k(x_i)\}$, so matters are not greatly complicated by the missing of data. Of course this leads to imprecision, the extent of which depends on the pattern of missing values.

*Example* 1. The simplest stochastic generator of $b_t$ is Bernoulli sampling, the $b_t$ are independent and $\mathrm{P}(b_t=1)=\pi$. Then $\beta=\pi$, $\beta(j)=\pi^2$

$(j>0)$, $\beta(j, k)=\pi^3$ $(0<j<k)$ and (13) is

$$\begin{aligned} &\kappa f(x_i)^{-1}\{\pi^{-2}\omega_{0j}(x_i)-\pi^{-1}\nu_j^2(x_i)\}\ , \qquad j=k\ , \\ &\kappa f(x_i)^{-1}\pi^{-1}\{\omega_{k-j,k}(x_i)-\nu_j(x_i)\nu_k(x_i)\}\ , \qquad j<k\ . \end{aligned} \tag{15}$$

Variances, in particular, can therefore be seriously inflated.

*Example* 2. A simple deterministic pattern of missing data, considered in [10], [17], has repetitions of $N$ observed values followed by $M$ missing ones, so $\beta=N(M+N)^{-1}$. It clearly suffices to evaluate $\beta(j, k)$ for $1\leqq j\leqq k\leqq M+N$, and this is given by

$$\beta(j, k)=(M+N)^{-1}\{\max(N-k, 0)+\max(k-j-M, 0)+\max(j-M, 0)\}\ .$$

If $q\geqq N$ then $N>M$ is necessary for (10). For $j=k$ (13) is

$$\begin{aligned} &\kappa f(x_i)^{-1}(M+N)[\{\max(N-j, 0)+\max(j-M, 0)\}^{-1} \\ &\quad\cdot\ \omega_{0j}(x_i)-N^{-1}\nu_j^2(x_i)] \\ &\leqq \kappa f(x_i)^{-1}[(2\pi-1)^{-1}\omega_{0j}(x_i)-\pi^{-1}\nu_j^2(x_i)]\ , \qquad \text{for } N>M\ , \end{aligned} \tag{16}$$

where $\pi=(M+N)^{-1}N$ to make comparison with (15). The majorant (16), which is attained for $M\leqq j\leqq N$, is larger than (15).

We briefly comment on the conditions of the theorem. Under appropriate smoothness conditions on the power spectrum, Gaussian processes, and certain functionals of Gaussian processes, satisfy (8), as do ARMA processes whose innovations admit a density, and certain Markov processes. The $\phi$-mixing condition on the other hand, excludes some of these cases. For further discussion see Deo [4]. If stronger smoothness conditions are imposed on $\nu_j$ and $f$, and additional conditions imposed on $K$, then larger values of $\lambda$ may be chosen in (11), see [13].

## 3. Proof of Theorem 1

A full proof is lengthly, so instead we emphasize certain aspects, in particular how we cope with the presence of missing data, and indicate how in other respects the theorem is similar to others in the literature. Full details may be obtained from the author. The asymptotic distribution of the $\hat{\nu}(x_i)$ follows from Rao ([12], pp. 387, 520) and Slutzky's theorem once we show that, for all constants $\zeta_{ij}$, $i=1,\cdots, r$, $j=0, 1,\cdots, q$,

$$\begin{aligned} &(Th)^{1/2}\sum_{i=1}^{r}\Big[\zeta_{i0}\{c(x_i)-bf(x_i)\} \\ &\qquad+\sum_{j=1}^{q}\zeta_{ij}\{c_j(x_i)-b(j)\nu_j(x_i)f(x_i)\}\Big]\xrightarrow{\mathcal{D}}\mathcal{N}(0, \sigma)\ , \end{aligned} \tag{17}$$

where

$$\sigma=\kappa\sum_{i=1}^{r}\left(\zeta_{i0}^{2}\beta+\sum_{j=1}^{q}\zeta_{ij}\left[\{2\zeta_{0j}\nu_j(x_i)+\zeta_{ij}\omega_{0j}(x_i)\}\beta(j)\right.\right.$$
$$\left.\left.+2\sum_{k>j}\zeta_{ik}\beta(k-j,k)\omega_{k-j,k}(x_i)\right]\right)f(x_i)\ .$$

It suffices to show that the statement holds true conditional on $\{b_t\}$ (which may be stochastic), because the $\mathcal{N}(0, \sigma)$ distribution is fixed and depends on $\{b_t\}$ only through the nonstochastic $\beta(j, k)$ (9), and is therefore the unconditional distribution also. We use the subscript $b$ to indicate operations conditional on $\{b_t\}$. Then

$$|\mathrm{E}_b\, c(x)-bf(x)|\leqq T^{-1}\sum_t b_t|\mathrm{E}\,\{h^{-1}K((X_t-x)h^{-1})\}-f(x)|$$
$$\leqq|\mathrm{E}\,\{h^{-1}K((X_t-x)h^{-1})\}-f(x)|\ ,$$

because $0\leqq b_t\leqq 1$, and likewise

$$|\mathrm{E}_b\, c_j(x)-b(j)\nu_j(x)f(x)|\leqq|\mathrm{E}\,\{h^{-1}\nu_j(X_t)K((X_t-x)h^{-1})\}-\nu_j(x)f(x)|\ ,$$

and both these bias terms are seen to be $O(h^{-1/2}T^{-1/2})$, much as in [13]. Thus the LHS of (17) is replaced by

$$(Th)^{1/2}\sum_{i=1}^{r}\left[\zeta_{i0}\{c(x_i)-\mathrm{E}_b\, c(x_i)\}+\sum_{j=1}^{q}\zeta_{ij}\{c_j(x_i)-\mathrm{E}_b\, c_j(x_i)\}\right]\ .$$

Now for some $D$, $0<D<\infty$, introduce

$$X_t'=X_tI(|X_t|\leqq D)\ ,\qquad X_t''=X_t-X_t'\ ,$$
$$Y_{tj}'(x)=X_t'K((X_{t-j}-x)h^{-1})\ ,\qquad Y_{tj}''(x)=X_t''K((X_{t-j}-x)h^{-1})\ ,$$
$$c_j'(x)=(Th)^{-1}\sum_{j+1}^{T}b_tb_{t-j}Y_{tj}'(x)\ ,\qquad c_j''(x)=c_j(x)-c_j'(x)\ ,$$
$$\nu_j'(x)=\mathrm{E}\,(X_t'\,|\,X_{t-j}=x)\ ,\qquad \omega_{jk}'(x)=\mathrm{E}\,(X_t'X_{t-j}'\,|\,X_{t-k}=x)\ .$$

Consider first, with argument $x$ omitted.

$$\mathrm{Var}_b\, c_j''\leqq(Th)^{-2}\sum_t b_tb_{t-j}\left\{\mathrm{Var}\,Y_{tj}''+\sum_{s\neq t}b_sb_{s-j}\,\mathrm{Cov}\,(Y_{tj}'', Y_{sj}'')\right\}$$
$$\leqq T^{-1}h^{-2}\left\{\mathrm{E}\,Y_{tj}''^{2}-(\mathrm{E}\,Y_{tj}'')^2+\sum_{s=1}^{T}|\mathrm{Cov}\,(Y_{tj}'', Y_{t+s,j}'')|\right\}$$
$$\leqq T^{-1}h^{-2}\left\{\varepsilon_h(2, 2)+\varepsilon_h(1, 1)^2+\sum_{s=1}^{T}|\mathrm{Cov}\,(Y_{tj}'', Y_{t+s,j}'')|\right\}\tag{18}$$

on defining

$$\varepsilon_h(\theta, \phi)=\mathrm{E}\,\{|X_t''|^{\theta}|K((X_{t-j}-x)h^{-1})|^{\phi}\}\ .$$

Using Holder's inequality for $s\neq t$,

$$(19)\qquad |\mathrm{Cov}\,(Y''_{tj}, Y''_{sj})| \leqq \varepsilon_h(1,1)^2 + \varepsilon_h(2+\delta, 2)^{2/(2+\delta)}\{\mathrm{E}\,|K((X_t-x)h^{-1}) \cdot K((X_s-s)h^{-1})|\}^{\delta/(2+\delta)},$$

but also, using inequality (2.1) in [4],

$$(20)\qquad |\mathrm{Cov}\,(Y''_{tj}, Y''_{sj})| \leqq C\varepsilon_h(2+\delta, 2+\delta)^{2/(2+\delta)}\alpha^{\delta/(2+\delta)}_{\max(|t-s|-j,0)}\,.$$

It is easily seen [13] that for $\theta \leqq 2+\delta$, $\phi \geqq 1$ and $\eta > 0$, $(x-\eta, x+\eta) \in S$,

$$\begin{aligned}\varepsilon_h(\theta,\phi) &= h \max_{|u|<\eta}\{\mathrm{E}\,(|X''_t|^\theta | X_{t-j}=x+u) f(x+u)\} \int |K(u)|^\phi du \\ &\quad + (\sup_{|u|\geqq \eta/h} |K(u)|^\phi)\, \mathrm{E}\,|X''_t|^\theta \\ &\leqq Ch\{D^{\theta-\gamma} \max_{|u|<\eta} \mathrm{E}\,(|X_t|^\gamma | X_{t-j}=x+u) + \eta^{-1}\,\mathrm{E}\,|X''_t|^\theta\}\,,\end{aligned}$$

and the expression in curly brackets$\to 0$ as $D\to\infty$. Moreover the expression in curly brackets in (19) is $O(h^2)$, like in [15], [20], [13]. Combining (18), (19) and (20) we have

$$Th\,\mathrm{Var}_b\, c''_j \leqq \varepsilon\left(1 + h + nh^{\delta/(2+\delta)} + h^{-\delta/(2+\delta)} \sum_n^\infty \alpha_k^{\delta/(2+\delta)}\right)$$

where $\varepsilon \to 0$ as $D\to\infty$. On choosing $n \sim h^{-\delta/(2+\delta)}$ we have $\lim_{D\to\infty} Th\,\mathrm{Var}_b\, c''_j = 0$ uniformly in $T$. Thus, as in [8], Theorem 18.5.3, (17) will follow if, for fixed $D$

$$(21)\qquad (Th)^{1/2} \sum_i [\zeta_{i0}\{c(x_i) - \mathrm{E}_b\, c(x_i)\} + \sum_j \zeta_{ij}\{c'_j(x_i) - \mathrm{E}_b\, c'_j(x_i)\}] \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma')\,,$$

and $\sigma' \to \sigma$ as $D\to\infty$. Let us check that the LHS of (21) has $\mathrm{Var}_b$ that $\to\sigma$ as $D\to\infty$. For $j\leqq k$ and with $\Delta(\cdot,\cdot)$ the Kronecker delta

$$\begin{aligned}&|Th\,\mathrm{Cov}_b\,\{c'_j(x), c'_k(y)\} - b(k-j,k)\kappa\omega'_{k-j,k}(x) f(x)\Delta(x,y)| \\ &\quad \leqq (Th)^{-1} \sum_t b_t b_{t-k}[b_{t+j-k}\{|\mathrm{E}\,\{Y'_{tk}(x) Y'_{t+j-k,j}(y)\} \\ &\qquad - h\kappa\omega'_{k-j,k}(x) f(x)\Delta(x,y)| + |\mathrm{E}\,Y'_{tk}(x)||\mathrm{E}\,Y'_{t+j-k,j}(y)|\} \\ &\qquad + \sum_{s\neq t} b_{s+j-k} b_{s-k} |\mathrm{Cov}\,\{Y'_{tk}(x), Y'_{s+j-k,j}(y)\}|]\end{aligned}$$

$$(22)\qquad \leqq |h^{-1}\,\mathrm{E}\,\{Y'_{tk}(x) Y'_{t+j-k,j}(y)\} - \kappa\omega'_{k-j,k}(x) f(x)\Delta(x,y)|$$

$$(23)\qquad \left\{\begin{aligned}&+h^{-1}|\mathrm{E}\,Y'_{tk}(x)||\mathrm{E}\,Y'_{t+j-k,k}(y)| \\ &+(Th)^{-1} \sum_{s\neq t}\sum |\mathrm{Cov}\,\{Y'_{tk}(x), Y'_{s+j-k,j}(y)\}|\,.\end{aligned}\right.$$

By the boundedness of $Y'$, inequality (17.2.2) in [8], and arguments rather similar to those used above, (23) is of order

$$D^2\left(h + nh + h^{-1}\sum_n^\infty \alpha_k\right) \leqq D^2\left\{h + nh + (\alpha_n^{2/(2+\delta)}/h) \sum_n^\infty \alpha_k^{\delta/(2+\delta)}\right\}$$

which $\to 0$ as $T\to\infty$ on taking $\alpha_n^\phi/n \sim h$ for $0<\phi<2/(2+\delta)$. For $x\neq y$, (22)$\to$0 much as in [14], [13]. As noted in [16], p. 1388, continuity of $\omega_{jk}(x)$ implies continuity of $\omega'_{jk}(x)$ and thus (22)$\to$0 for $x=y$ by Bochner's theorem [15]. Thus with (9) and $\omega'_{jk}(x)\to\omega_{jk}(x)$ we have established

$$\lim_{D\to\infty} \mathrm{p}\lim_{T\to\infty} Th \operatorname{Cov}_b \{c'_j(x), c'_k(y)\} = \beta(k-j, k)\kappa\omega_{k-j,k}(x)f(x)\Delta(x, y) ,$$

for $j\leqq k$. In the same way we have

$$\lim_{D\to\infty} \mathrm{p}\lim_{T\to\infty} Th \operatorname{Cov}_b \{c'_j(x), c(y)\} = \beta(j)\kappa\nu_j(x)f(x)\Delta(x, y) ,$$

$$\mathrm{p}\lim_{T\to\infty} Th \operatorname{Cov}_b \{c(x), c(y)\} = \beta\kappa f(x)\Delta(x, y) ,$$

so the asymptotic variance structure has been verified. The asymptotic normality part of the proof in (21) is omitted because it is lengthy and almost identical to that of Theorem 5.3 in [13] (which uses the method of Theorems 18.5.4 and 18.5.5 in [8] for the asymptotic normality of partial sums under a milder condition than (8)) except that all expectations involved are conditional on $\{b_t\}$, as above.

We describe only the most difficult part of the proof that (14) is consistent for (13), namely that for $j\leqq k$

$$\mathrm{p}\lim_{T\to\infty} w_{k-j,k}(x) = \beta(k-j, k)\omega_{k-j,k}(x)f(x) .$$

Now

$$T^{-1}\sum_t b_t b_{t-j-k} b_{t+j-k} b_{t-k} |h^{-1} \mathrm{E}\{X_t X_{t+j-k} K((X_{t-k}-x)h^{-1})\} - \omega_{k-j,k}(x)f(x)| \to 0$$

by Bochner's theorem and in view of (9) it is only necessary to show that $\mathrm{p}\lim_{T\to\infty} \Delta=0$, where $\Delta = w_{k-j,k}(x) - \mathrm{E}_b\, w_{k-j,k}(x)$. Write $Y'_t = X_t X_{t+j-k} \cdot I(|X_t X_{t+j-k}|\leqq D)$, $Y''_t = Y_t - Y'_t$ and correspondingly $\Delta = \Delta' + \Delta''$. Then by arguments similar to those used previously we have

$$\mathrm{E}_b\,(\Delta')^2 \leqq D^2\left\{(Th)^{-1} + nT^{-1} + (Th^2)^{-1}\sum_n^\infty \alpha_k\right\} \to 0$$

as $T\to\infty$ with $n\sim h^{-1}$. On the other hand

$$\mathrm{E}_b\,|\Delta''| \leqq 2h^{-1}D^{-\delta}\, \mathrm{E}\,\{|X_t X_{t+j-k}|^{1+\delta/2} |K((X_{t-k}-x)h^{-1})|\} \leqq CD^{-\delta}$$

because the expectation is $O(h)$, and then let $D\to\infty$.

## 4. Numerical example (1)

Dunsmuir and Robinson [6] analyzed, by ARMA models, a time series of daily average carbon monoxide (CO) measurements. The series contains 663 observations (in parts per million) and 135 missing values, so $T=798$, and the configuration of missing values (which is very irregular) is apparently "asymptotically stationary". The observations themselves are not stationary but were made approximately so in [6] by linear detrending and extraction of a day-of-the week effect, and for further details see [6].

The Gaussian kernel (3) was used throughout in the formulae for $\hat{f}(x)$ and $\hat{\nu}_j(x)$, which were computed for $j=1,\cdots,q=28$ and an equally-spaced grid of $r=30$ $x$-values which covered the range of the data. In addition estimated asymptotic $28\times 28$ covariance matrices of the $\hat{\nu}(x)$ were computed by formula (14), and the diagonal elements used to estimate 95% confidence intervals for the $\nu_j(x)$, using the asymptotic normality. To estimate 95% confidence intervals for $f(x)$ the consistent estimator $\kappa\hat{f}(x)/b$ of the asymptotic variance $\kappa f(x)/\beta$ was used. A range of bandwidth values was tried, $h=0.001$, 0.01, 0.1, 0.2, 0.4, 0.6 and 1.0, and separate graphs of $\hat{\nu}_j(x)$ against $x$ produced for each $(j, h)$ combination, and of $\hat{f}(x)$ for each $h$. Smooth curves were plotted by GINOGRAF and the CALCOMP plotter on the University of Surrey's PRIME Network.

The value $h=0.001$ is much too small for sensible estimates to be achieved, and the instability of the estimators for other small $h$ values frequently produced very steep gradients which caused plotting difficulties. On the other hand the curves for $h=1.0$ were apparently too smooth. Graphs of $\hat{f}(x)$ and the $\hat{\nu}_j(x)$, $j=1,\cdots,7$ only are presented for $h=0.4$ only, in Figure I, the broken line representing 95% confidence limits. The graph in the top left hand corner of the figure is of $\hat{f}(x)$ while the legend "lag $j$" on the remaining graphs refers to $\hat{\nu}_j(x)$.

The broad message of Figure I is suggestive of a unimodal, positively skewed pdf, and $\nu_j(x)$ which are quite linear (offering some justification for the modelling in [6]) and which tend to approach the horizontal as $j$ increases, although not monotonically. Noticeable departures from linearity do appear in the $\hat{\nu}_j(x)$, particularly for extreme values of $x$. For large positive $x$ they are well-behaved for some $j$ but otherwise sharply rising or falling. These phenomena may in large part be spurious, in view of the small probability density for $x>4$, as expressed in the large confidence intervals for such $x$. For $j=7$ (and also $j=8$, 9, not shown) there is a rather marked downturn from $x=3.8$ on; in [6] it was found that the removal of day-of-the-week means did not altogether dispose of a weekly periodic effect, and perhaps the graphs are indicative of a weekly nonlinear effect. As far as the covariance estimates of the $\hat{\nu}_j(x)$ for different $x$ are concerned, these generally gave fairly small correlations, and the correlation between $\hat{\nu}_j(x)$ and $\hat{\nu}_{j+k}(x)$ falls off quite rapidly as $k$ increases.

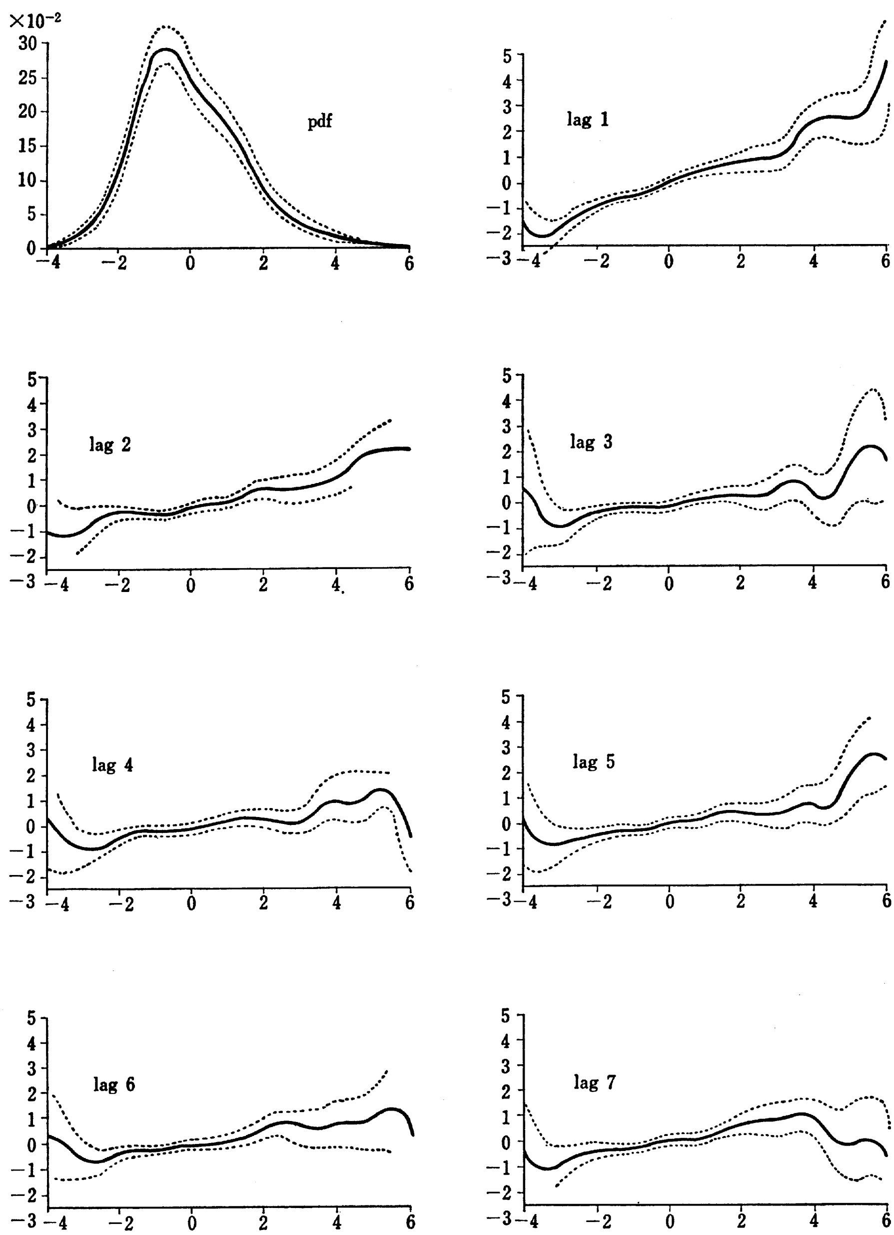


Figure I

## 5. Convergence of interpolator of missing values

Theorem 1 implies that $\hat{\nu}_j(x)=\nu_j(x)+O_p(h^{-1/2}T^{-1/2})$ for fixed $x$-values. A variety of other consistency results can be established, including uniform convergence over a bounded subset of $x$-values. For the problem of interpolating a missing $X_{m+j}$ on the basis of an observed $X_m$, it is of more relevance to study the convergence of $\hat{\nu}_j(X_m)$ to the random variable $\nu_j(X_m)$, which is the least squares interpolator.

We continue to assume $X_t$ has a pdf and is strongly mixing but now, for some $\delta>0$,

$$\mathrm{E}\,|X_t|^{1+\delta}<\infty \tag{24}$$

$$\sum_{k=1}^{\infty} \alpha_k^{\gamma/(2+\gamma)}<\infty\ , \qquad \gamma>0,\ \gamma+1\geqq\delta\ . \tag{25}$$

We assume $\beta(j)=\operatorname{p}\lim_{T\to\infty} b(j)$ exists and is positive. With regard to $K$ we require (12) and also that $K(x)$ has representation

$$K(x)=\int_R \tilde{K}(u)e^{iux}du\ , \tag{26}$$

where the integral converges absolutely. Finally

$$Th^{2(1+\delta)(1+\gamma)/\delta(2+\gamma)}\to\infty\ , \qquad h\to 0 \text{ as } T\to\infty\ . \tag{27}$$

THEOREM 2. *Under the above conditions*

$$\operatorname{p}\lim_{T\to\infty} \hat{\nu}_j(X_m)=\nu_j(X_m)$$

*for fixed* $j$, $m$.

The requirements on $K$ now exclude, for example, the uniform kernel $K(x)=(1/2)I(|x|\leqq 1)$. The mildest version of the first part of (27) is $Th^2\to\infty$, attained when $\delta=\gamma+1$. However, the moment and mixing conditions (24), (25) are weaker than in Section 2, including cases where $X_t$ has infinite variance. We now make no continuity assumptions on $f$ or $\nu_j$ (cf. [3], [5]), and this is achieved by application of a theorem of Stein [19], which employs the denseness in $L_p$ of all continuous functions with bounded support. A similar result for prediction is available.

## 6. Proof of Theorem 2

Define

$$\tilde{f}(y)=h^{-1}\int K((x-y)h^{-1})f(x)dx\ , \qquad \tilde{g}(y)=h^{-1}\int K((x-y)h^{-1})\nu_j(x)f(x)dx\ ,$$

and express $\hat{\nu}_j(X_m)-\nu_j(X_m)$ as

$$\hat{\nu}_j(X_m)-\nu_j(X_m)=\hat{f}(X_m)^{-1}\{b(j)^{-1}a_1+a_2-\nu_j(X_m)(b^{-1}a_3+a_4)\}\ ,$$

where

$$a_1=(Th)^{-1}\sum_{j+1}^{T} b_t b_{t-j} X_t k((X_{t-j}-X_m)h^{-1})-b(j)\tilde{g}(X_m)\ ,$$

$$a_2=\tilde{g}(X_m)-\nu_j(X_m)f(X_m)\ ,\qquad a_3=b\{\hat{f}(X_m)\}-\tilde{f}(X_m)\}\ ,$$

$$a_4=\tilde{f}(X_m)-f(X_m)\ .$$

We shall show that $\operatorname{p}\lim_{T\to\infty} a_i=0,\ i=1,\cdots,4$; this implies $\operatorname{p}\lim_{T\to\infty}\hat{f}(X_m)=f(X_m)$ which is positive a.s. because $f$ is the pdf of $X_m$. The convergence i.p. of $b$ and $b(j)$ to nonzero limits completes the proof.

On using (26) we may write $a_1$ as

$$a_1=T^{-1}\int_R \tilde{K}(uh)\exp(-iuX_m)\sum_t b_t b_{t-j}\{H(X_t)-\mathrm{E}\,H(X_t)\}du\ , \tag{28}$$

where $H(X_t)=X_t\exp(iuX_{t-j})$, reference to $X_{t-j}$ and $u$ being suppressed. Introduce

$$X_t'=X_t I(|X_t|\leqq Dh^{-1/\delta})\ ,\qquad X_t''=X_t-X_t'\ ,$$

for some $D$, $1<D<\infty$. By ([4], equn. (2.1))

$$|\mathrm{E}\,\{H(X_t')-\mathrm{E}\,H(X_t')\}\{\bar{H}(X_s')-\mathrm{E}\,\bar{H}(X_s')\}|$$
$$\leqq C(\mathrm{E}\,|X_t'|^{2+\gamma})^{2/(2+\gamma)}\alpha^{\gamma/(2+\gamma)}_{\max(|t-s|-j,0)}$$

where

$$\mathrm{E}\,|X_t'|^{2+\gamma}\leqq(Dh^{-1/\delta})^{1+\gamma-\delta}\,\mathrm{E}\,|X_t|^{1+\delta}\ .$$

On the other hand

$$\mathrm{E}\,|H(X_t'')-\mathrm{E}\,H(X_t'')|\leqq 2\,\mathrm{E}\,|H(X_t'')|\leqq 2D^{-\delta}h\,\mathrm{E}\,|X_t|^{1+\delta}\ .$$

Thus from (28)

$$\mathrm{E}\,|a_1|\leqq T^{-1}\int_R \tilde{K}(uh)|\,\mathrm{E}\,([\mathrm{E}_b\,|\sum_t b_t b_{t-j}\{H(X_t')-\mathrm{E}\,H(X_t')\}|^2]^{1/2}$$
$$+\mathrm{E}_b\,|\sum_t b_t b_{t-j}\{H(X_t'')-\mathrm{E}\,H(X_t'')\}|)du$$
$$\leqq CT^{-1}\int_R |\tilde{K}(uh)|\left\{\left(TD'h^{-2(1+\gamma-\delta)/\delta(2+\gamma)}\sum_0^T \alpha_k^{\gamma/(2+\gamma)}\right)^{1/2}+TD^{-\delta}h\right\}du$$
$$\leqq C(D'^{1/2}T^{-1/2}h^{-(1+\delta)(1+\gamma)/\delta(2+\gamma)}+D^{-\delta})\ ,$$

where $D'=D^{2(1+\gamma-\delta)/(2+\gamma)}$. On letting $T\to\infty$ and then $D\to\infty$ we establish $a_1\xrightarrow{L_1}0$. By a simpler proof, $a_3\xrightarrow{L_1}0$ also.

To deal with $a_2$ and $a_4$ we may apply Theorem 2, part (b) of [19], pp. 62, 63. Notice that $\nu_j(x)$ and $f(x)$ are integrable, $K$ integrates to 1, and

$$\sup_{|x|\geqq|y|}|K(x)|\leqq C(1+|y|)^{-1-\lambda}\ ,$$

which is integrable. Thus $\tilde f(y)\to f(y)$, $\tilde g(y)\to g(y)$ a.e. $y$, and so $a_2\to 0$, $a_4\to 0$ a.s., completing the proof.

## 7. Numerical example (2)

The sequence of 5 consecutive missing values at $t=367,\cdots,381$ in the CO data set was interpolated. As well as the basic interpolator discussed in § 5, a weighted average based on several of these was employed, along with backward interpolators and arithmetic means of backward and forward interpolators, all of which share the same consistency properties but may be more efficient than the one of § 5. A still more efficient, though computationally more onerous, method is possible, which estimates (4). We report only results for the "forward" interpolator of $X_{366+j}$

$$\hat X^{(n)}_{366+j}=n^{-1}\sum_{k=j}^{j+n-1}\hat\nu_k(X_{366-k+j})$$

for $j=1,\cdots,5$ and $n=1, 8$. Thus $\hat X^{(1)}_{366+j}=\hat\nu_j(X_{366})$ whereas $\hat X^{(8)}_{366+j}$ is the arithmetic mean of interpolators based on $X_{359},\cdots,X_{366}$, all of which were observed, the case $n=8$ selected to incorporate a possible weekly effect (see § 4 and [6]). Again the kernel (3) was used.

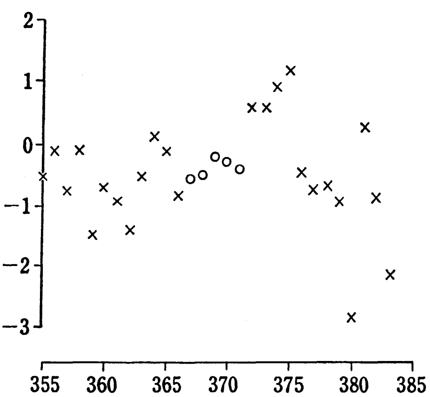

Figure II. $n=1$, $h=0.2$.

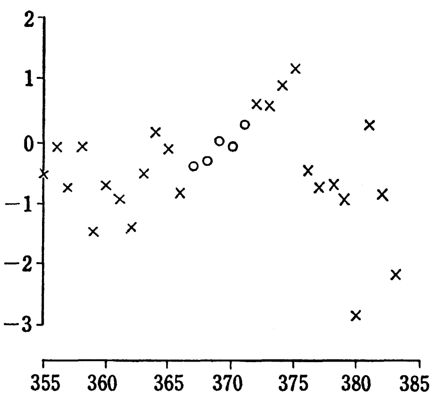

Figure III. $n=3$, $h=0.01$.

Only results for a single bandwidth are presented in each case, $h=0.2$ in Fig. II, when $n=1$, and $h=0.01$ in Fig. III, $n=8$. The asterisks represent observed values for $t=355,\cdots,366, 372,\cdots,383$, while circles are interpolations. Interpolation methods would not normally bridge the considerable discrepancy in level at $t=366$ and $t=372$, though $\hat X$ does so quite impressively in Fig. III. Increasing $n$ or decreasing $h$ moves interpolations towards the series mean.

## 8. Choice of bandwidth

In the applied work we reported above we tried several choices of $h$, and in practice this seems a wise, if possibly expensive precaution, to avoid over- or under-smoothing. Other information is available to guide the choice of $h$. The asymptotic theory provides upper and (in the central limit theorem) lower bounds for the decay of $h$ with $T$. In addition it may be desirable to use an $h$ in $\hat{\nu}_j(x)$ which varies over $x$, inversely with respect to supposed probability mass. There is also evidence that the stronger the serial dependence the smaller $h$ should be chosen, in order to avoid imprecision (though at risk of increased bias). In simulations we have examined the performance of bandwidth selection procedures which incorporate these features. More automatic procedures are possible. Silverman [18] suggests one, for density estimation from independent observations. The method of cross-validation has been considered by various authors, though again its performance in the presence of serial dependence has not been investigated. For example, Hall [7] shows that a cross-validated choice can minimize, asymptotically, mean integrated squared error (MISE) for nonparametric estimators using independent observations. Note, however, that bandwidths which aim for a small MISE will not necessarily produce a good approximation to our central limit result, or good interpolators.

This research was supported by the Social Science Research Council. I am grateful to Martin Dransfield for carrying out the computations described in the paper and preparing the figures.

LONDON SCHOOL OF ECONOMICS

### REFERENCES

[1] Ahmad, I. A. (1979). Strong consistency of density estimation by orthogonal series methods for dependent variables with applications, *Ann. Inst. Statist. Math.*, **31**, 279-288.

[2] Akaike, H. and Ishiguro, M. (1980). Trend estimation with missing observations, *Ann. Inst. Statist. Math.*, **32**, 481-488.

[3] Collomb, G. (1982). Proprietes de convergence presque complete du predicteur a noyau (preprint).

[4] Deo, C. M. (1973). A note on empirical processes of strong mixing sequences, *Ann. Prob.*, **5**, 870-875.

[5] Devroye, L. P. and Wagner, T. J. (1980). On the $L_1$ convergence of kernel estimators of regression with applications in discrimination, *Zeit. Wahrscheinlichkeitsth.*, **51**, 15-25.

[6] Dunsmuir, W. and Robinson, P. M. (1982). Estimation of time series models in the presence of missing data, *J. Amer. Statist. Ass.*, **76**, 560-568.

[7] Hall, P. (1983). Large sample optimality of least squares cross-validation in density

estimation, *Ann. Statist.*, **11**, 1156-1174.

[8] Ibragimov, I. A. and Linnik, Yu. V. (1971). *Independent and Stationary Sequences of Random Variables*, Wolters-Noordhoff, Groningen.

[9] Masry, E. (1983). Probability density estimation from sampled data, *IEEE Trans. Inf. Theory*, **29**, 696-709.

[10] Parzen, E. M. (1963). On spectral analysis with missing observations and amplitude modulation, *Sankhyā*, A, **25**, 383-392.

[11] Pham, T. D. (1981). Nonparametric estimation of the drift coefficient in the diffusion equation, *Math. Oper.*, **12**, 61-73.

[12] Rao, C. R. (1973). *Linear Statistical Inference and its Applications*, 2nd ed., Wiley, New York.

[13] Robinson, P. M. (1983). Nonparametric estimators for time series, *J. Time Series Anal.*, **4**, 185-207.

[14] Rosenblatt, M. (1970). Density estimators and Markov sequences, in *Nonparametric Techniques in Statistical Inference* (ed. Puri, M. C.), Cambridge University Press, Cambridge, 199-210.

[15] Roussas, G. (1967). Nonparametric estimation in Markov processes, *Ann. Inst. Statist. Math.*, **21**, 73-87.

[16] Roussas, G. (1969). Nonparametric estimation of the transition distribution function of a Markov process, *Ann. Math. Statist.*, **40**, 1386-1400.

[17] Sakai, H. (1980). Fitting autoregressions with regularly missed observations, *Ann. Inst. Statist. Math.*, **32**, 393-400.

[18] Silverman, B. (1978). Choosing the window width when estimating a density, *Biometrika*, **65**, 1-11.

[19] Stein, E. M. (1970). *Singular Integrals and Differentiability Properties of Functions*, Princeton University Press, Princeton, N.J.

[20] Takahata, H. (1977). Limiting behaviour of density estimates for stationary asymptotically uncorrelated processes, *Bull. Tokyo Gakugei Univ.*, IV, **29**, 1-9.

[21] Takahata, H. (1980). Almost sure convergence of density estimators for weakly dependent stationary processes, *Bull. Tokyo Gakugei Univ.* IV, **32**, 11-32.