ISSN 1440-771X
ISBN 0 7326 1056 7

MONASH UNIVERSITY

ANCORA · IMPARO

AUSTRALIA

Nonparameteric Estimation and Symmetry Tests for Conditional Density Functions

Rob J. Hyndman & Qiwei Yao

Working Paper 17/98
December 1998

DEPARTMENT OF ECONOMETRICS AND BUSINESS STATISTICS

Nonparametric estimation and symmetry tests for conditional density functions

Rob J Hyndman[1] and Qiwei Yao[2]

11 December 1998

Abstract: We suggest two new methods for conditional density estimation. The first is based on locally fitting a log-linear model, and is in the spirit of recent work on locally parametric techniques in density estimation. The second method is a constrained local polynomial estimator. Both methods always produce non-negative estimators. We propose an algorithm suitable for selecting the two bandwidths for either estimator. We also develop a new bootstrap test for the symmetry of conditional density functions. The proposed methods are illustrated by both simulation and application to a real data set.

Keywords: bandwidth selection; bootstrap; conditioning; density estimation; kernel smoothing; symmetry tests.

1 Introduction

The goal of this paper is two-fold. First, we propose two new methods for estimating the conditional density function of Y_t given X_t based on observations from a strictly stationary process $\{(X_t, Y_t)\}$. Second, we propose a new bootstrap method for testing the symmetry of conditional density functions.

Our first estimation method is locally parametric; it produces estimators of arbitrarily high order and is always non-negative. In spirit, this approach is related to recently-introduced local parametric methods for density estimation; see for example Copas (1995), Simonoff (1996, Section 3.4), Hjort and Jones (1997), Loader (1996) and Hall, Wolff and Yao (1999). Our second method is a constrained version of local polynomial estimation for conditional density functions, which was studied by Fan, Yao and Tong (1996). The simple constraint makes the estimator always non-negative while retaining the nice asymptotic properties of the local polynomial estimators. All the work mentioned above is based on the double kernel smoothing approach, which has also been adapted by Yu and Jones (1998) to estimate conditional quantiles.

We consider the mean square error properties of our estimators and show that the asymptotic optimal bandwidth in the x-direction is greater than that in ordinary kernel regression estimation in order to compensate for the data sparseness due to the smoothing in y-direction. Similarly, the optimal bandwidth in the y-direction is greater than that for unconditional density estimation to compensate for the smoothing in the x-direction. Based on the mean-square error properties, we propose a bandwidth selection algorithm for these estimators.

[1]Department of Econometrics and Business Statistics, Monash University, Clayton VIC 3168, Australia.
[2]Institute of Mathematics and Statistics, University of Kent at Canterbury, Canterbury, Kent CT2 7NF, U.K.

As far as we know, the symmetry of *conditional* density functions has never been addressed in the literature before, although various statistical methods have been proposed for testing the symmetry of unconditional density functions, which include, among others, Butler (1969), Hollander (1971), Rothman and Woodroofe (1972), Srinivasan and Godio (1974), Doksum et al (1977), Hill and Rao (1977), Lockhart and McLaren (1985), Csörgö and Heathcote (1987), Zhu (1998) and Diks and Tong (1998). In addition to the interest on symmetric distributions addressed in the aforementioned literature, the symmetry of conditional distribution is particular relevant in modelling time series data in business and finance (Brännäs and De Gooijer, 1992) and in constructing predictive regions for nonlinear time series (Polonik and Yao, 1998).

The paper is organized as follows: we propose the two new estimators for conditional densities in Section 2. The asymptotic normality of the estimators are presented under some mixing conditions. Section 3 addresses the issue of choosing bandwidths. The bootstrap tests for the symmetry are discussed in Section 4. Numerical illustration through two simulated examples and a real data set is reported in Section 5.

2 Estimation of conditional densities

We assume that data are available in the form of a strictly stationary stochastic process $\{(X_i, Y_i)\}$, where Y_i is a scalar and X_i is a d-dimensional vector. Naturally, this includes the case where the pairs (X_i, Y_i) are independent and identically distributed. In the time series context, X_i typically denotes a vector of lagged values of Y_i. Let $g(y|x)$ be the conditional density of Y_i given $X_i = x$, which we assume to be smooth in both x and y. We are interested in estimating $g(y|x)$ and its derivatives from the data $\{(X_i, Y_i), 1 \le i \le n\}$.

Let $K(.)$ be a symmetric density function on $\mathbb{R}$ and $K_b(u) = b^{-1}K(u/b)$. Note that as $b \to 0$,

$$\mathrm{E}\{K_b(Y_i - y)|X_i = x\} = g(y|x) + O(b^2).$$

This suggests that $g(y|x)$ can be regarded as a regression of $K_b(Y_i - y)$ on X_i. For example, if $d = 1$, Nadaraya-Watson kernel regression yields the **kernel estimator**

$$\tilde{g}(y|x) = \sum_{i=1}^{n} w_i(x) K_b(Y_i - y) \tag{2.1}$$

where

$$w_i(x) = \frac{W_h(X_i - x)}{\sum_{j=1}^{n} W_h(X_j - x)},$$

$W_h(u) = h^{-1}W(u/h)$, $W(\cdot)$ is a kernel function and $h > 0$ is a bandwidth. This estimator was proposed by Rosenblatt (1969) and its properties were further explored by Hyndman, Bashtannyk and Grunwald (1996). Bashtannyk and Hyndman (1998) explore bandwidth selection rules. Note that there are two smoothing parameters: h controls the smoothness between conditional densities in the x direction (the smoothing parameter for the regression) and b controls the smoothness of each conditional density in the y direction.

If local polynomial regression is used we obtain the **local polynomial estimator** proposed by Fan, Yao and Tong (1996). Let

$$R(\theta;x,y) = \sum_{i=1}^{n}\{K_b(Y_i - y) - \sum_{j=0}^{r}\theta_j(X_i - x)^j\}^2 W_h(X_i - x). \tag{2.2}$$

Then $\widehat{g}(y|x) = \widehat{\theta}_0$ is a local rth order polynomial estimator where $\widehat{\theta}_{xy} = (\widehat{\theta}_0, \widehat{\theta}_1, \ldots, \widehat{\theta}_r)'$ is that value of θ which minimizes $R(\theta;x,y)$. For $r = 0$, this estimator is identical to (2.1). While this estimator has some nice properties, it is not restricted to be non-negative for $r > 0$. In this paper, we propose two new estimators to overcome this shortcoming. To simplify discussion we introduce our methods and develop theory in the case where X_i is a scalar (i.e., $d = 1$).

2.1 Two new non-negative estimators

We replace $R(\theta;x,y)$ by

$$R_1(\theta;x,y) = \sum_{i=1}^{n}\{K_b(Y_i - y) - A(X_i - x, \theta)\}^2 W_h(X_i - x) \tag{2.3}$$

where

$$A(x,\theta) = \ell\Big(\sum_{j=0}^{p}\theta_j x^j\Big)$$

and $\ell(\cdot)$ is a monotonic function mapping $\mathbb{R} \to \mathbb{R}^+$. Using $\ell(u) = \exp(u)$ seems a reasonable choice. Then $\widehat{g}_1(y|x) \equiv A(0,\widehat{\theta}_{xy}) = \ell(\widehat{\theta}_0)$ where $\widehat{\theta}_{xy}$ minimizes $R_1(\theta;x,y)$.

We call this the **local parametric estimator**. It is in the same spirit as the local logistic estimator for a conditional distribution function proposed by Hall, Wolff and Yao (1999), and is a conditional version of the density estimator proposed by Loader (1996). Further, it is equivalent to using local likelihood estimation (Tibshirani and Hastie, 1987) for the regression of $K_b(Y_i - y)$ against X_i with the Gaussian likelihood and link function ℓ^{-1}. Consequently, $\widehat{\theta}_{xy}$ may be easily computed using local likelihood estimation software such as `locfit` (Loader, 1997). (Note that the `gam` function in S-Plus will not allow a non-identity link function with the Gaussian likelihood.) If an identity link is used ($\ell(u) = u$), we obtain the local polynomial estimator as a special case.

An alternative estimator is obtained by modifying the local linear estimator for $g(y|x)$ directly to force it to be positive. Let $\theta_0 = \ell(\alpha)$ in (2.2). We shall denote this estimator by $\widehat{g}_2(y|x)$ and refer to it as the **constrained local polynomial estimator**. Obviously, this idea can also be applied to the problem of estimation of a conditional distribution function, addressed by Hall, Wolff and Yao (1999).

Depending on bandwidth choice, both of these estimators also furnishes consistent estimators of the derivatives of the conditional density. Let

$$g^{(i)}(y|x) \equiv (\tfrac{\partial}{\partial y})^i g(y|x), \quad g^{(|j)}(y|x) \equiv (\tfrac{\partial}{\partial x})^j g(y|x), \quad g^{(i|j)}(y|x) \equiv (\tfrac{\partial}{\partial y})^i(\tfrac{\partial}{\partial x})^j g(y|x),$$

$$\ell^{(j)}(u) \equiv (\tfrac{\partial}{\partial u})^j \ell(u) \qquad \text{and} \qquad A^{(j)}(x,\theta) \equiv (\tfrac{\partial}{\partial x})^j A(x,\theta).$$

For $j = 1, 2, \ldots, r$ we can estimate the density derivatives:

$$\widehat{g}_1^{(|j)}(y|x) = A^{(j)}(0,\widehat{\theta}_{xy}) = \sum_{k=1}^{j}\widehat{\theta}_k \tbinom{j-1}{k-1}\ell^{(k)}(\widehat{\theta}_0) \qquad \text{and} \qquad \widehat{g}_2^{(|j)}(y|x) = j!\widehat{\theta}_j\,.$$

If $K(u)$ is at least q-times differentiable, then for $i = 1,2,\ldots,q$ we can also estimate the density derivatives $\widehat{g}_1^{(i)}(y|x)$ and $\widehat{g}_2^{(i)}(y|x)$. These are unavailable in closed form but they are easily obtained using numerical differentiation.

In practice, we rescale $\widehat{g}(y|x)$, $\widehat{g}_1(y|x)$ and $\widehat{g}_2(y|x)$ to ensure they integrate to 1. Note that there is no need to rescale the kernel estimator $\tilde{g}(y|x)$.

2.2 Asymptotic properties

For the local parametric estimator $\widehat{g}_1(y|x)$ we only consider functions A of type $A(x,\theta) = \exp(\theta_0 + \theta_1 x + \ldots + \theta_r x^r)$, with $r \geq 1$. Let f denote the marginal density of X_i. We impose the following regularity conditions:

(C1) For fixed y and x, $f(x) > 0$, $g(y|x) > 0$, f is continuous at x, and $g(y|\cdot)$ has $2[r/2]+2$ continuous derivatives in a neighbourhood of x, where $[t]$ denotes the integer part of t.

(C2) The kernel K and W are symmetric, compactly supported probability density functions. Further, $|W(x_1) - W(x_2)| \leq C|x_1 - x_2|$ for any x_1, x_2.

(C3) The process $\{(X_i, Y_i)\}$ is absolutely regular, that is

$$\beta(j) \equiv \sup_{i\geq 1} \mathrm{E}\Big\{ \sup_{A\in\mathcal{F}_{i+j}^{\infty}} |P(A|\mathcal{F}_1^i) - P(A)| \Big\} \to 0 \quad \text{as} \quad j \to \infty,$$

where $\mathcal{F}_i^j$ denotes the σ-field generated by $\{(X_k, Y_k) : i \leq k \leq j\}$. Furthermore, $\sum_{j\geq 1} j^2 \beta(j)^{\delta/(1+\delta)} < \infty$ for some $\delta \in [0,1)$. (We define $a^b = 0$ when $a = b = 0$.)

(C4) As $n \to \infty$, $h \to 0$, $b \to 0$, $nbh \to \infty$ and $\liminf_{n\to\infty} nh^{2(r+1)} > 0$.

Condition (C3) holds with $\delta = 0$ if and only if the process $\{(X_i, Y_i)\}$ is m-dependent for some $m \geq 1$. The requirement of the kernels being compactly supported is imposed for the sake of brevity of proofs. In particular, the Gaussian kernel is allowed. The assumption on the mixing conditions is also not the weakest possible.

Theorem 1 below presents the asymptotic normality of the estimators. The asymptotic expressions for biases and variances are useful in development of the bandwidth selection procedures described in Section 3. The proof of the theorem is largely the same as the proof of Theorem 1 in Hall, Wolff and Yao (1999) (see also the proof of Theorem 1 of Fan, Yao and Tong, 1996), which will be omitted.

We introduce some notation first. Define

$$\kappa_j = \int u^j W(u)du, \quad \nu_j = \int u^j W^2(u)du, \quad \mu_j = \int u^j K(u)du, \quad \text{and} \quad \lambda_j = \int u^j K^2(u)du.$$

Let S denote the $(r+1)\times(r+1)$ matrix with (i,j)-th element κ_{i+j-2}, and $\kappa^{(i,j)}$ be the (i,j)-th element of S^{-1}. Let $r_1 = 2[r/2]+2$,

$$\tau_r^2 = \lambda_0 \int \left(\sum_{i=1}^{r+1} \kappa^{(1,i)} v^{i-1} \right)^2 W^2(v)dv, \qquad \eta_r = \frac{1}{(r+1)!} \sum_{i=1}^{r+1} \kappa^{(1,i)} \kappa_{r_1+i-1},$$

and let θ_{xy} be uniquely defined by

$$g(y|x) = A(0, \theta_{xy}), \quad \text{and} \quad g^{(|j)}(y|x) = A^{(j)}(0, \theta_{xy}) \quad j = 1, \dots, r.$$

Let N_{n1} and N_{n2} denote random variables with the standard Normal distribution.

Theorem 1. (i) Suppose $r \geq 1$ and conditions (C1) – (C4) hold. Then as $n \to \infty$,

$$\begin{aligned} \widehat{g}_1(y|x) - g(y|x) &= (nhb)^{-1/2} \left\{ \frac{g(y|x)}{f(x)} \right\}^{1/2} \tau_r N_{n1} + h^{r_1} \eta_r \{ g^{(|r_1)}(y|x) - A^{(r_1)}(0, \theta_{xy}) \} \\ &+ b^2 \tfrac{\mu_2}{2} g^{(2)}(y|x) + o\{(nhb)^{-1/2} + h^{r_1} + b^2\}. \end{aligned} \quad (2.4)$$

(ii) Assume conditions (C1) – (C4) with $r = 1$. Then as $n \to \infty$,

$$\begin{aligned} &\widehat{g}_2(y|x) - g(y|x) = (nhb)^{-1/2} \left\{ \frac{\lambda_0 \nu_0 g(y|x)}{f(x)} \right\}^{1/2} N_{n2} \\ &+ h^2 \tfrac{\kappa_2}{2} g^{(|2)}(y|x) + b^2 \tfrac{\mu_2}{2} g^{(2)}(y|x) + o\{(nhb)^{-1/2} + h^2 + b^2\}. \end{aligned} \quad (2.5)$$

Remark 1. To the first order, the asymptotic variance of $\widehat{g}_1(y|x)$ is exactly the same as in the case of local polynomial estimator $\widehat{g}(y|x)$ of order r. This similarity extends also to the bias term, to the extent that for both $\widehat{g}_1$ and local polynomial estimators the bias is of order $O(h^{r+1} + b^2)$ for odd r and $O(h^{r+2} + b^2)$ for even r. However, the form of bias as functionals of the 'regression mean' g are quite different. This is a consequence of the fact that $\widehat{g}_1(y|x)$ is constrained to be non-negative. In fact, (2.4) would also hold for the local polynomial estimator with order r if we replace the term $A^{(r_1)}(0, \theta_{xy})$ by 0. See Fan and Gijbels (1996) §6.2 or Fan, Yao and Tong (1996). Note, however, that neither reference gives explicitly the bias term in the order h^{r_1} and that the expression they give for τ_2^2 contains some typographical errors.

Remark 2. For the linear case ($r = 1$) we have $\tau_1^2 = \lambda_0 \nu_0$ and $\eta_1 = \kappa_2/2$. Because of the above remark, (2.5) also holds for the standard local linear estimator. On the other hand, when $\ell(u) = \exp(u)$ and $r = 1$, $A^{(r_1)}(0, \theta_{xy}) = \left[g^{(|1)}(y|x) \right]^2 / g(y|x)$.

Remark 3. For the quadratic case ($r = 2$), we have

$$\tau_2^2 = \frac{\lambda_0 (\kappa_4^2 \nu_0 - 2\kappa_2 \kappa_4 \nu_2 + \kappa_2^2 \nu_4)}{(\kappa_4 - \kappa_2^2)^2} \qquad \text{and} \qquad \eta_2 = \frac{\kappa_4^2 - \kappa_6 \kappa_2}{6(\kappa_4 - \kappa_2^2)}.$$

Remark 4. It may be proved that, under conditions (C1) – (C4) and $r \geq 1$, $\widehat{\theta}_{xy} \to \theta_{xy}$. Consequently, we may prove that $\widehat{g}_1(y|x)$ is a consistent estimator. Similarly, $\widehat{g}_2(y|x)$ is also consistent.

3 Bandwidth selection

Using (2.4), we find the asymptotic mean square error of $\widehat{g}_1(y|x)$ is

$$\mathrm{E}\{\widehat{g}_1(y|x)-g(y|x)\}^2 \approx \frac{\tau_r^2 g(y|x)}{nhbf(x)} + \left\{h^{r_1}\eta_r\left[g^{(|r_1)}(y|x)-A^{(r_1)}(0,\theta_{xy})\right]+b^2\frac{\mu_2}{2}g^{(2)}(y|x)\right\}^2,$$

and so the weighted integrated MSE is

$$\begin{aligned}\text{IMSE} &= \int\int \mathrm{E}\{\widehat{g}_1(y|x)-g(y|x)\}^2 f^2(x)\,dx\,dy \\ &= \left\{\frac{\tau_r^2}{nhb}+\alpha_r h^{2r_1}+\beta_r h^{r_1}b^2+\gamma b^4\right\}\{1+o(1)\} \qquad (3.1)\end{aligned}$$

$$\text{where}\quad \alpha_r = \eta_r^2\int\int\left[g^{(|r_1)}(y|x)-A^{(r_1)}(0,\theta_{xy})\right]^2 f^2(x)\,dx\,dy \qquad (3.2)$$

$$\beta_r = \mu_2\eta_r\int\int g^{(2)}(y|x)\left[g^{(|r_1)}(y|x)-A^{(r_1)}(0,\theta_{xy})\right] f^2(x)\,dx\,dy \qquad (3.3)$$

$$\text{and}\quad \gamma = \frac{\mu_2^2}{4}\int\int\left(g^{(2)}(y|x)\right)^2 f^2(x)\,dx\,dy. \qquad (3.4)$$

Bashtannyk and Hyndman (1998) used a similar weighted IMSE to derive bandwidths for the estimator (2.1).

Optimal bandwidths for $\widehat{g}_1(y|x)$ can be derived by differentiating (3.1) with respect to h and b and setting the derivatives to zero. Solving the resulting equations gives

$$h^* = \left(\frac{\tau_r^2}{nc_r r_1(2\alpha_r+\beta_r c_r^2)}\right)^{\frac{2}{5r_1+2}} \qquad \text{and} \qquad b^* = c_r(h^*)^{r_1/2} \qquad (3.5)$$

where

$$c_r = \sqrt{\frac{(r_1-2)\beta_r+\sqrt{(r_1-2)^2\beta_r^2+32r_1\alpha_r\gamma}}{8\gamma}},.$$

When $r=1$, this simplifies to $c_1=(\alpha_1/\gamma)^{1/4}$. (Because of Remark 2, h^* and b^* in (3.5) are also optimal for $\widehat{g}_2(y|x)$ with $r=1$.) Substituting these optimal bandwidths into (3.1) shows that the IMSE is of order $n^{-4r_1/(5r_1+2)}$. Note that the optimal bandwidth h^* is different from that in standard kernel regression estimation. For example, $h^*=O(n^{-1/6})$ when $r=1$ while the optimal bandwidth for local linear regression estimation is of order $n^{-1/5}$. Intuitively we need a larger bandwidth (in the order $n^{-1/6}$) to compensate the sparseness of data points due to the smoothing in the y-direction. Similarly, the optimal bandwidth b^* is of order $O(n^{-1/6})$ when for unconditional density estimation, the optimal order is $O(n^{-1/5})$. The larger bandwidth for the conditional estimator is because of the local estimation due to smoothing in the x-direction.

3.1 Normal reference rules

We shall derive the optimal bandwidths for the estimator where the conditional distribution is normal with quadratic conditional mean and constant variance σ^2, and the marginal distribution of X is normal with mean μ and variance v^2. Then we can write

$$g(y|x)=\tfrac{1}{\sigma}\,\phi\left(\tfrac{y-d_0-d_1(x-\mu)-d_2(x-\mu)^2}{\sigma}\right) \qquad \text{and} \qquad f(x)=\tfrac{1}{v}\,\phi\left(\tfrac{y-\mu}{v}\right).$$

Substituting these into (3.2)–(3.4), we obtain

$$\gamma = \frac{3\mu_2^2}{64\pi\sigma^5 v}, \quad \alpha_1 = \frac{\kappa_2^2(2d_2^2\sigma^2 + d_1^4 + 12d_2^2v^2(d_1^2+d_2^2v^2))}{16\pi\sigma^5 v}, \quad \beta_1 = \frac{\mu_2\kappa_2(d_1^2+2d_2^2v^2)}{16\pi\sigma^5 v}$$

and $c_1 = (\alpha/\gamma)^{1/4}$ when the log link ($\ell(u) = \exp(u)$) is used. For the local linear estimator ($\ell(u) = u$), we obtain the same γ and c_1 values, with

$$\alpha_1 = \frac{\kappa_2^2(8d_2^2\sigma^2 + 3d_1^4 + 36d_2^2v^2(d_1^2+d_2^2v^2))}{64\pi\sigma^5 v} \quad \text{and} \quad \beta_1 = \frac{3\mu_2\kappa_2(d_1^2+2d_2^2v^2)}{32\pi\sigma^5 v}.$$

The local quadratic estimator is more difficult and we only give the bandwidths for the identity link ($\ell(u) = u$) assuming the conditional mean is linear (i.e., $d_2 = 0$). Then we obtain the same γ with

$$\alpha_2 = \frac{105\eta_2^2 d_1^8}{64\pi\sigma^9 v}, \quad \beta_2 = \frac{-15\eta_2\mu_2 d_1^4}{32\pi\sigma^7 v} \quad \text{and} \quad c_2^2 = \frac{|\eta_2|d_1^4(\sqrt{305} - 5\,\mathrm{sign}(\eta_2))}{2\mu_2\sigma^2}$$

where $\mathrm{sign}(u) = u/|u|$.

In the special case where both $W(u)$ and $K(u)$ denote a standard normal kernel, and the conditional mean is linear ($d_2 = 0$), we substitute the above values into (3.5) to obtain the following simple rules:

- When $r = 1$ and $\ell(u) = \exp(u)$, $h^* \approx 0.916\left(\frac{v\sigma^5}{n|d|^5}\right)^{1/6}$ and $b^* = 1.05|d|h^*$.
- When $r = 1$ and $\ell(u) = u$, $h^* \approx 0.935\left(\frac{v\sigma^5}{n|d|^5}\right)^{1/6}$ and $b^* = |d|h^*$.
- When $r = 2$ and $\ell(u) = u$, $h^* \approx 0.703\left(\frac{v\sigma^{10}}{nd^{10}}\right)^{1/11}$ and $b^* \approx \frac{2.37d^2}{\sigma}(h^*)^2$.

3.2 A bandwidth selection algorithm

For a given bandwidth b and a given value y, finding $\hat{g}(y|x)$ is a standard nonparametric problem of regressing $K_b(Y_i - y)$ on X_i. Therefore, we can adapt bandwidth selection methods used in regression for use in this problem. Let $M_b(h;y)$ denote a goodness-of-fit statistic for the regression of $K_b(Y_i - y)$ on X_i with bandwidth h. For example, $M_b(h;y)$ may denote the generalized cross-validation statistic (Fan and Gijbels, 1996, p.45). We then define

$$M_b(h) = \sum_{j=1}^{N} M_b(h;y'_j)$$

where $\boldsymbol{y} = \{y'_1, \ldots, y'_N\}$ are equally spaced in the sample space of Y. For a given value of b, $M_b(h)$ may be minimized to select a value of h. This approach was suggested by Bashtannyk and Hyndman (1998) for the kernel estimator with $M_b(h;y)$ denoting the penalized average square prediction error (see, for example, Härdle, 1991). Fan, Yao and Tong (1996) suggested a similar approach for the local polynomial estimator with $M_b(h)$ denoting the Residual Squares Criterion proposed by Fan and Gijbels (1995).

When this approach is combined with the normal reference rules, we have a useful algorithm for selecting the bandwidth parameters.

1 Find an initial value for the smoothing parameter b using the normal reference rule.

2 Given this value of b, minimize $M_b(h)$ to find a value for h.

An alternative bandwidth selection strategy is to use a parametric bootstrap method similar to that proposed by Bashtannyk and Hyndman (1998). We fit a parametric model

$$Y_i = d_0 + d_1 X_i + \cdots + d_k X_i^k + \sigma\varepsilon_i$$

where ε_i is standard normal, $d_0, \ldots, d_k$ and σ are estimated from the data and k is determined by AIC. Then we form a parametric estimator $\tilde{g}(y\,|\,x)$ based on the model. We simulate a bootstrap data set $\boldsymbol{Y}^* = \{Y_1^*, \ldots, Y_n^*\}$ from the fitted parametric model based on the observations $\boldsymbol{X} = \{X_1, \ldots, X_n\}$. We form the bootstrap estimator $\widehat{g}^*(.|.)$ based on the sample $\{(X_i, Y_i^*)\}$. We choose h and b to minimize

$$\sum_{i=1}^{n} E[\{\widehat{g}^*(Y_i|X_i) - \tilde{g}(Y_i|X_i)\}^2|(X_j, Y_j),\, 1 \le j \le n]. \tag{3.6}$$

Unfortunately, this bootstrap method is extremely slow, even on modern computing equipment, and we do not use it in the numerical examples.

4 Bootstrap tests for symmetry

For fixed x with $f(x) > 0$, we are interested in testing the hypothesis that the conditional distribution $g(\cdot|x)$ is symmetric, that is

$$H_0 : g(y|x) = g(2u(x) - y|x) \quad \text{for any } y,$$

where $u(x)$ is the centre of the conditional distribution of $g(.|x)$. Under hypothesis H_0, we would expect that the above equality also holds approximately for a good estimator of g, say $\widehat{g}$. Therefore, we define the test statistic

$$T(x) = \min_u \int \{\widehat{g}(y|x) - \widehat{g}(2u - y|x)\}^2 dy$$

and reject H_0 for large values of T.

To derive the asymptotic distribution of T (under H_0) is a tedious matter. Typically the sample size n must be very large to ensure asymptotic results are adequately accurate in nonparametric tests (see, for example, Hjellvik, Yao and Tjøstheim, 1998). Therefore we adopt a bootstrap approach in this paper.

Note all the estimators described in Section 2 can be written as linear forms of $\{K_b(Y_i - y)\}$ as follows

$$\widehat{g}(y|x) = \sum_{i=1}^{n} m_i(x) K_b(Y_i - y),$$

where the weight $m_j(x)$ depends on $\{X_i\}$ and x only. Note the kernel function $K(.)$ is symmetric. It is easy to see that

$$\widehat{g}(2u(x) - y|x) = \sum_{i=1}^{n} m_i(x) K_b(2u(x) - Y_i - y).$$

This means that the mirror reflection of the estimator $\widehat{g}(\cdot|x)$ with respect to $u(x)$ is $\widehat{g}$ itself obtained with the sample $\{(Y_i, X_i)\}$ replaced by $\{(2u(x) - Y_i, X_i)\}$. This motivates the following resampling scheme.

1 We calculate

$$u(x) = \arg\min_u \int \{\widehat{g}(y|x) - \widehat{g}(2u - y|x)\}^2 dy. \tag{4.1}$$

2 We sample n independent observations $\{X_i^*, 1 \le i \le n\}$ from $\{X_i, 1 \le i \le n\}$ with replacement.

3 Suppose $X_i^* = X_{i_j}$. For each $1 \le i \le n$, sample Y_i^* from the uniform distribution on the two symmetric points Y_{i_j} and $2u(x) - Y_{i_j}$.

4 Form the statistic T^* in the same way as T with $\{X_i, Y_i\}$ replaced by $\{X_i^*, Y_i^*\}$.

We reject H_0 if T is greater than the upper α-point of the conditional distribution of T^* given $\{X_i, Y_i\}$. In fact, the p-value is the relative frequency of the event $\{T^* \ge T\}$ in the bootstrap replications.

We may let $\widehat{g}$ be the local parametric estimator $\widehat{g}_1$ with $r = 1$ or the constrained local linear estimator $\widehat{g}_2$. We keep the bandwidths unchanged in the bootstrap stage.

Since we only test the symmetry of $g(\cdot|x)$ at fixed x, one would expect that we only sample Y_i^* from a symmetric distribution when X_i^* is close to x. This is effectively achieved in the nonparametric estimation of $g(\cdot|x)$, since the estimation is localized by the kernel function.

When we generate the bootstrap samples, we largely ignore the possible dependence in the data. Note that under the mixing condition (C3), the dependence does not enter the major terms (i.e., first order terms) in the asymptotic expansions in Theorem 1. This is due to the fact that in nonparametric regression (with random design), we only use effectively the nh nearest neighbours in the *state space*, which are unlikely to be the neighbours in the *time space* under the mixing condition (C3). Those points could be regarded as asymptotically independent when $n \to \infty$. In fact we may prove that it holds almost surely that the conditional distribution of T^* given $\{X_i, Y_i\}$ is asymptotically equal to the null-hypothesis distribution of T (cf. Kreiss, Neumann and Yao, 1998).

Remark 5. Note that since $f(x) > 0$, the null hypothesis can be expressed equivalently as $H_0: g(y|x) f(x) = g(2u(x) - y|x) f(x)$ for any y. Furthermore, the joint density function $p(x,y) \equiv g(y|x) f(x)$ can be easily estimated. For example, the simple product kernel estimator is $\widehat{p}(x,y) = \frac{1}{n}\sum_{i=1}^n W_h(X_i - x) K_b(Y_i - x)$. Therefore, an alternative test statistic can be defined as $T_1(x) = \min_u \int \{\widehat{p}(x,y) - \widehat{p}(x, 2u - y)\}^2 dy$. The bootstrap procedure described above can be applied to facilitate this alternative test.

Remark 6. When the a density is symmetric, a symmetric estimator may be obtained as

$$\widehat{\widehat{g}}(y|x) = \frac{1}{2}\left(\widehat{g}(y|x) + \widehat{g}(2u(x) - y|x)\right). \tag{4.2}$$

See Kraft, Lepage and van Eeden (1985) and Meloche (1991) for further discussion on estimation of symmetric densities. In the numerical examples, we estimate the density by (4.2) if $\widehat{g}(y|x)$ passes the symmetry test.

5 Numerical examples

We illustrate the symmetry tests through simulations and by application to some real data. In all cases, we have used the Gaussian kernel, $K(u) = W(u) = \phi(u) = \exp(-u^2/2)/\sqrt{2\pi}$.

Example 1

Consider the model $Y_i = 5 + (1 + W_i)X_i + \varepsilon_i$ where $\{X_i\}$, $\{W_i\}$ and $\{\varepsilon_i\}$ are all independent with X_i uniformly distributed on $[0, 12]$, ε_i normally distributed with zero mean and variance 9, and W_i is a binary variable with $\Pr(W_i = 1) = 1 - \Pr(W_i = 0) = 0.3$. Figure 1 shows a scatterplot of 500 observations from this model. The line through the points is $u(x)$ calculated from (4.1). When $x = 0$, the density is symmetric, and it increases in skewness as x increases.

We computed the p-value of the bootstrap test for symmetry for $0 \leq x \leq 12$ at steps of 0.5. For these tests, we used the local parametric estimator of $g(y|x)$ with $r = 1$ and bandwidths chosen using the algorithm of Section 3.2 to be $h = 1.35$ and $b = 1.59$. (For this example, the true optimal bandwidths calculated using (3.5) are $h^* = 0.87$ and $b^* = 1.25$.) Figure 2 shows the p-values. Each test involved 100 replications. The skewness is clearly detected by the tests for $x > 6$.

Example 2

We next consider a quadratic AR(1) time series model

$$Y_t = 0.23Y_{t-1}(16 - Y_{t-1}) + 0.4\varepsilon_t \tag{5.1}$$

where $\{\varepsilon_t\}$ is a sequence of independent random variables each with the standard normal distribution truncated in the interval $[-12, 12]$. The conditional distribution of Y_t given $X_t \equiv Y_{t-m}$ is symmetric for $m = 1$ but not necessarily so for $m > 1$. Figure 3 shows a lagged scatterplot of 600 observations from this model with $m = 3$. The line through the points is $u(x)$ calculated from (4.1) where $\widehat{g}(y|x)$ is the local parametric estimate with $r = 1$. Bandwidths were chosen using the algorithm to be $h = 0.4$ and $b = 1.2$. For each of the bootstrap tests, 100 replications were performed. The p-values from the bootstrap test for symmetry are shown in Figure 4. There is a clear evidence that the conditional distribution is not symmetric for x between 6.5 and 8.5.

Old Faithful Geyser data

Azzalini and Bowman (1990) give data on the waiting time between the starts of successive eruptions and the duration of the subsequent eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming. The data were collected continuously between 1–15 August

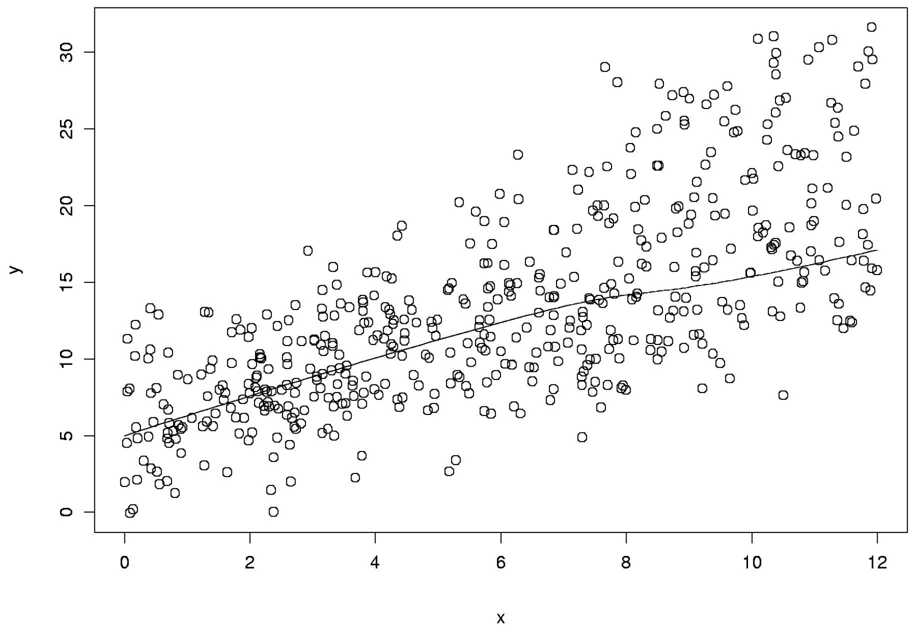

Figure 1: *Scatterplot of 500 observations from Example 1. The line through the points is $u(x)$, the estimated centre of symmetry, calculated from (4.1).*

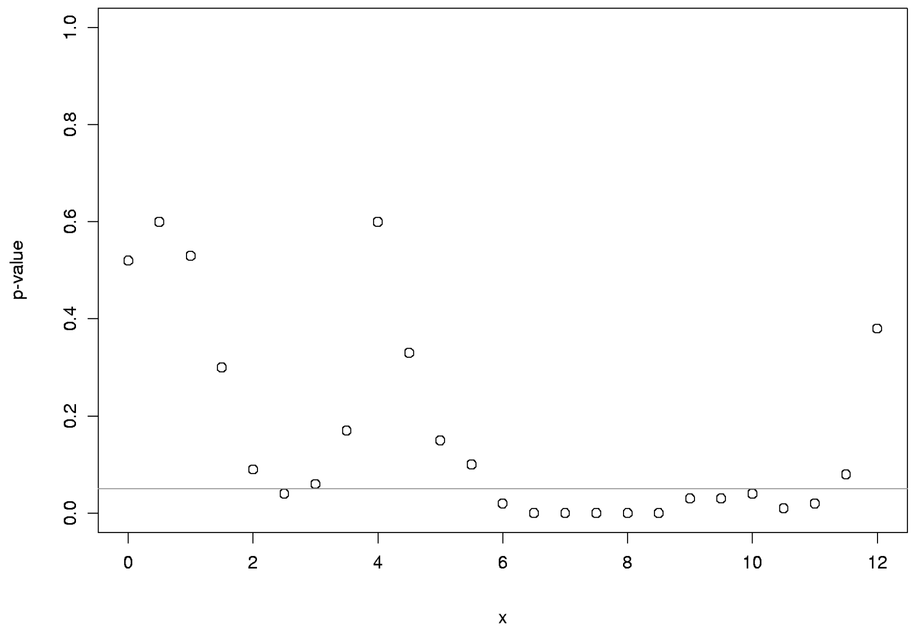

Figure 2: *The p-values of the bootstrap test for symmetry of the conditional density $g(y|x)$ in Example 1. Here $\widehat{g}(y|x)$ is the local parametric estimate with $r = 1$ and bandwidths chosen using the normal reference rules to be $h = 1.1$ and $b = 1.6$. The horizontal line shows the 0.05 level.*

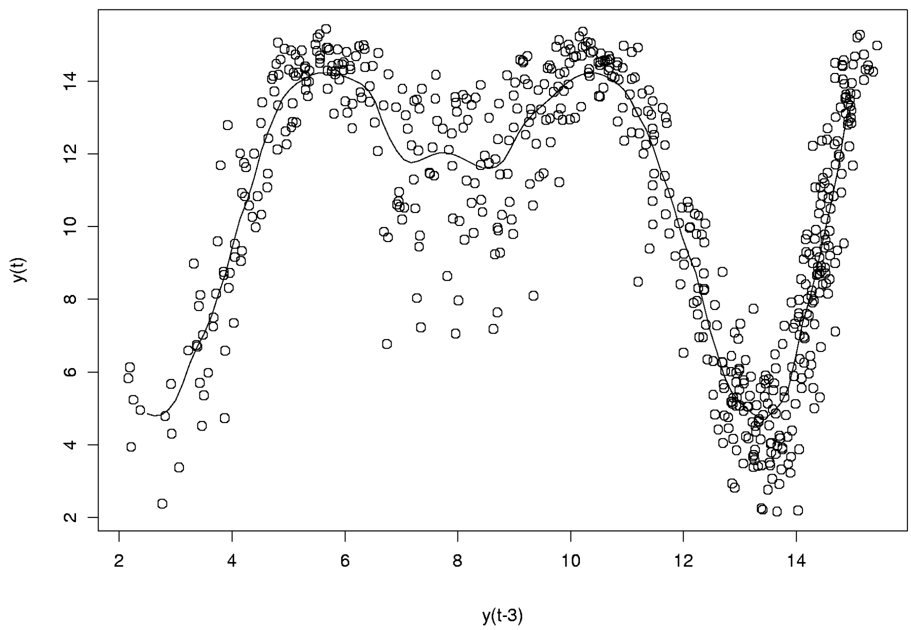

Figure 3: *Scatterplot of 600 observations from Example 2. The line through the points is $u(x)$, the estimated centre of symmetry, calculated from (4.1).*

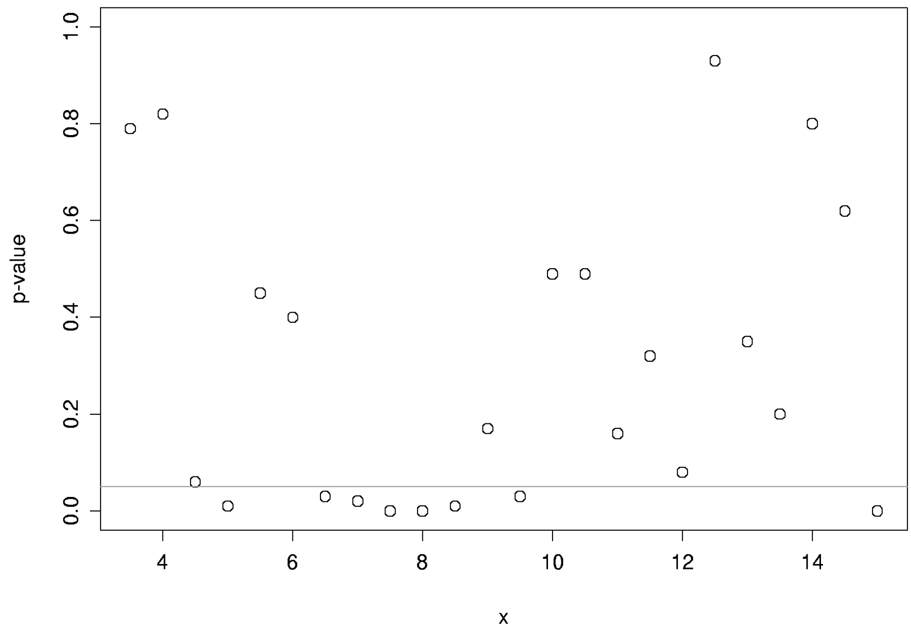

Figure 4: *The p-values of the bootstrap test for symmetry of the conditional density $g(y|x)$ in Example 2. Here $\widehat{g}(y|x)$ is the kernel estimate with bandwidths $h = b = 0.5$. The horizontal line shows the 0.05 level.*

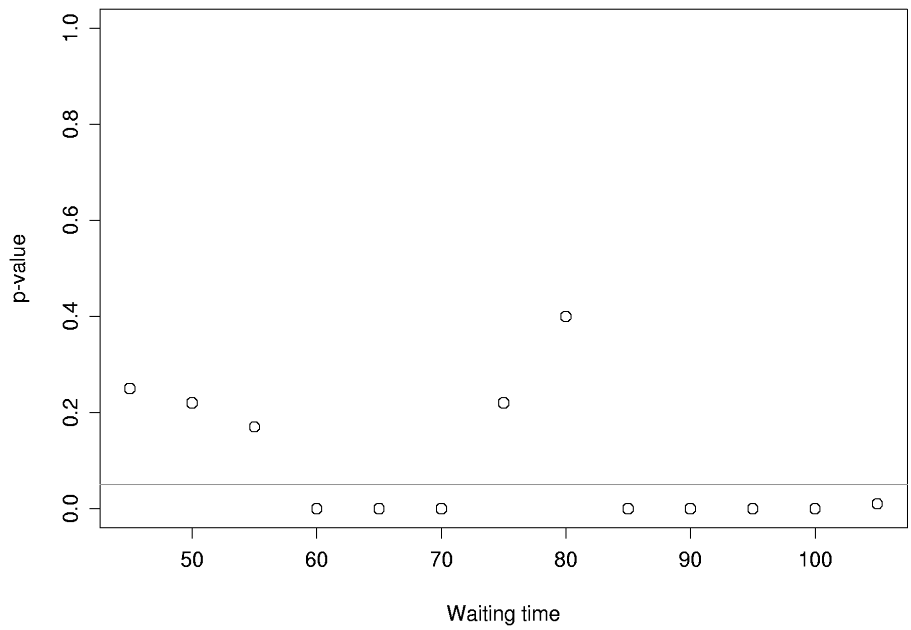

Figure 5: *The p-values of the bootstrap test for symmetry of the density of the Old Faithful Geyser eruption duration conditional on the waiting time between eruptions. Here* $\hat{g}(y|x)$ *is the kernel estimate with bandwidths* $h = 7.2$ *and* $b = 0.41$ *chosen using the bandwidth selection algorithm. The horizontal line shows the 0.05 level.*

1985. There are a total of 299 observations. The times are measured in minutes. Some duration measurements, taken at night, were originally recorded as S (short), M (medium), and L (long). These values have been coded as 2, 3 and 4 minutes respectively. This data set is also distributed with S-Plus.

We are interested in the distribution of duration time conditional on the previous waiting time. The bandwidth selection algorithm gives bandwidths $h = 8.1$ and $b = 0.33$. Using these values we test the symmetry of the conditional densities (again using the kernel estimator (2.1)) with 100 replications per test. The p-values from the bootstrap test for symmetry are shown in Figure 5. Where the p-value is greater than 0.05, we replace $\hat{g}(y|x)$ by the symmetric estimator (4.2). The resulting estimates are shown in Figure 6 using the stacked density visualization method of Hyndman, Bashtannyk and Grunwald (1996).

6 Acknowledgements

This work was carried out while Rob Hyndman was a visitor to the Department of Statistics, Colorado State University and Qiwei Yao was a visitor to the Australian National University. Rob Hyndman was supported in part by an Australian Research Council grant. Qiwei Yao was supported partially by EPSRC Grant L16385 and BBSRC/EPSRC Grant 96/MMI09785. The authors would like to thank Clive Loader for making his locfit software available, and Chris Jones for some helpful comments.

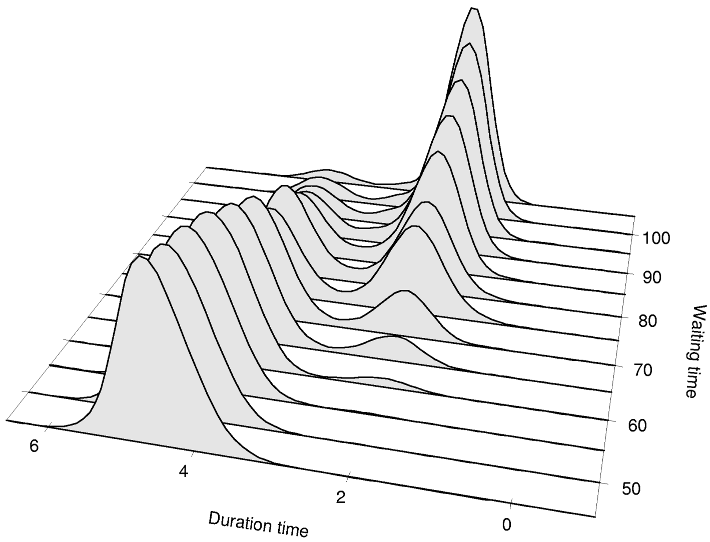

Figure 6: *Estimated conditional density of eruption duration conditional on waiting time to the eruption. The densities have been symmetrized if the p-values in Figure 5 are greater than 0.05. Bandwidths were chosen using the selection algorithm.*

7 References

AZZALINI, A. and BOWMAN, A.W. (1990) A look at some data on the Old Faithful geyser. *Applied Statistics* **39**, 357–365.

BASHTANNYK, D.M. and HYNDMAN, R.J. (1998) *Bandwidth selection for kernel conditional density estimation*. Technical report, Department of Econometrics and Business Statistics, Monash University.

BRÄNNÄS, K. and DE GOOIJER, J.G. (1992) Modelling business cycle data using autoregressive-asymmetric moving average models. *ASA Proceedings of Business and Economic Statistics Section*, 331–336.

BUTLER, C. (1969) A test for symmetry using the sample distribution function. *Ann. Math. Statist.* **40**, 2209–2210.

COPAS, J.B. (1995) Local likelihood based on kernel censoring. *J. Roy. Statist. Soc. Ser. B* **57**, 221–235.

CSÖRGÖ, S. and HEATHCOTE, C.R. (1987) Testing for symmetry. *Biometrika* **74**, 177–184.

DIKS, C. and TONG, H. (1998) A test for symmetries of multivariate probability distributions. (Submitted.)

DOKSUM, K.A., FENSTAD, G. and AABERGE, R. (1977) Plots and tests for symmetry. *Biometrika* **64**, 473–487.

FAN, J. and GIJBELS, I. (1995) Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *J. R. Statist. Soc.* B **57**, 371–394.

FAN, J. and GIJBELS, I. (1996) *Local polynomial estimation*. Chapman and Hall: New York.

FAN, J., YAO, Q. and TONG, H. (1996) Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika* **83**, 189–206.

HALL, P., WOLFF, R. and YAO, Q. (1999) Methods for estimating a conditional distribution function. *J. Amer. Statist. Assoc.* To appear March 1999.

HÄRDLE, W. (1991) *Smoothing techniques with implementation in S*. Springer-Verlag: New York.

HILL, D.L. and RAO, P.V. (1977) Tests for symmetry based on Cramér-von Mises statistics. *Biometrika* **64**, 489–494.

HJELLVIK, V., YAO, Q. and TJØSTHEIM, D. (1998) Linearity testing using local polynomial approximation. *J. Statist. Plann. Infer.* **68** (2), 295–321.

HJORT, N.L. and JONES, M.C. (1996) Locally parametric nonparametric density estimation. *Ann. Statist.*, **24**, 1619–1647.

HOLLANDER, M. (1971) A nonparametric test for bivariate symmetry. *Biometrika* **71**, 203–212.

HYNDMAN, R.J., BASHTANNYK, D.M. and GRUNWALD, G.K. (1996) Estimating and visualizing conditional densities. *J. Comp. Graph. Statist.* **5**(4), 315–336.

KRAFT, C.H., LEPAGE, Y. and VAN EEDEN, C. (1985) Estimation of a symmetric density function. *Communications in Statistics: Theory and Methods*, **14**, 273–288.

KREISS, J.P., NEUMANN, M. and YAO, Q. (1998) Bootstrap tests for simple structures in nonparametric time series regression. (Submitted.)

LOADER, C.R. (1996) Local likelihood density estimation. *Ann. Statist.*, **24**, 1602–1618.

LOADER, C. (1997) Locfit: an introduction. *Statistical Computing and Graphics Newsletter*, **8**(1), 11–17.

LOCKHART, R.A. and MCLAREN, C.G. (1985) Asymptotic points for a test of symmetry about a specified mean. *Biometrika* **85**, 208–210.

MELOCHE, J. (1991) Estimation of a symmetric density. *Canadian J. Statist.*, **19**, 151–164

POLONIK, W. and YAO, Q. (1998) Conditional minimum volume predictive regions for stochastic processes. (Submitted.)

ROSENBLATT, M. (1969) "Conditional probability density and regression estimators", in P. Krishnaiah, ed., *Multivariate Analysis II*. Academic Press: New York, pp. 25–31.

ROTHMAN, E.N.D. and WOODROOFE, M.A. (1972) A Cramér-von Mises type statistic for testing symmetry. *Ann. Math. Statist.* **43**, 2035–2038.

SIMONOFF, J.S. (1996) *Smoothing Methods in Statistics.* Springer, New York.

SRINIVASAN, R. and GODIO, L.B. (1974) A Cramér-von Mises type statistic for testing symmetry. *Biometrika*, **61**, 196–198.

TIBSHIRANI, R. and HASTIE, T. (1987) Local likelihood estimation. *J. Amer. Statist. Assoc.* **82**, 559–567.

YU, K. and JONES, M.C. (1998) Local linear quantile regression. *J. Amer. Statist. Assoc.* **93**, 228–237.

ZHU, L.-X. (1998) Assessing elliptical symmetry via a computer-assisted test procedure. *J. Amer. Statist. Assoc.*, to appear.