Nonparametric Estimation of Smooth Conditional Distributions

Bruce E. Hansen*
University of Wisconsin†

www.ssc.wisc.edu/~bhansen

May 2004

Abstract

This paper considers nonparametric estimation of smooth conditional distribution functions (CDFs) using kernel smoothing methods. We propose estimation using a new smoothed local linear (SLL) estimator. Estimation bias is reduced through the use of a local linear estimator rather than local averaging. Estimation variance is reduced through the use of smoothing. Asymptotic analysis of mean integrated squared error (MISE) reveals the form of these efficiency gains, and their magnitudes are demonstrated in numerical simulations. Considerable attention is devoted to the development of a plug-in rule for bandwidths which minimize estimates of the asymptotic MISE. We illustrate the estimation method with an application to the U.S. quarterly GDP growth rate.

*Research supported by the National Science Foundation.
†Department of Economics, 1180 Observatory Drive, University of Wisconsin, Madison, WI 53706

1 Introduction

This paper considers nonparametric estimation of smooth conditional distribution functions (CDFs). As the CDF is the conditional expectation of an indicator function, it may be estimated by nonparametric regression estimation methods. A recent paper which exploits advances in nonparametric regression for estimation of the CDF is Hall, Wolff and Yao (1999). See also Fan and Yao (2003).

Standard nonparametric regression estimates of the CDF, however, are not smooth and therefore are inefficient. If the dependent variable in these regressions – the indicator function – is replaced by a smooth function, estimation variance can be reduced at the cost of an increase in bias. If the degree of smoothing is carefully chosen, it can result in a lowered finite sample mean integrated squared error (MISE).

The advantages of smoothed estimation of distribution functions has been discussed in the statistics literature. For references see Hansen (2004). The extension to CDF estimation considered here is new.

Our analysis concerns a pair of random variables $(Y, X) \in R \times R$ that have a smooth joint distribution. Let $f(y)$ and $g(x)$ denote the marginal densities of Y and X, and let $f(y \mid x)$ denote the conditional density of Y given $X = x$. The conditional distribution function is

$$F(y \mid x) = \int_{-\infty}^{y} f(u \mid x)\, du.$$

Define as well the derivatives

$$\begin{aligned} F^{(s)}(y \mid x) &= \frac{\partial^s}{\partial x^s} F(y \mid x) \\ f^{(s)}(y \mid x) &= \frac{\partial^s}{\partial x^s} f(y \mid x) \\ f'(y \mid x) &= \frac{\partial}{\partial y} f(y \mid x). \end{aligned}$$

Given a random sample $\{Y_1, X_1, ..., Y_n, X_n\}$ from this distribution, the goal is to estimate $F(y \mid x)$ at a fixed value x. To simplify the notation we will sometimes suppress dependence on x.

Let $\Phi(s)$ and $\phi(s)$ denote the standard normal distribution and density functions and let $\Phi_\sigma(s) = \Phi(s/\sigma)$ and $\phi_\sigma(s) = \sigma^{-1}\phi(s/\sigma)$. Let $\phi^{(m)}(s) = (d^m/ds^m)\,\phi(s)$ and $\phi_\sigma^{(m)}(s) = (d^m/ds^m)\,\phi_\sigma(s) = \sigma^{-(m+1)}\phi^{(m)}(s/\sigma)$ denote the derivatives of $\phi(s)$ and $\phi_\sigma(s)$. In particular, $\phi^{(1)}(s) = -s\phi(s)$ and $\phi^{(2)}(s) = \left(s^2 - 1\right)\phi(s)$.

Our estimators will make use of two univariate kernel functions $w(s)$ and $k(s)$. We assume that both are symmetric and second-order, are normalized to have a unit variance (e.g. $\int_{-\infty}^{\infty} s^2 w(s)ds = 1$), and have a finite sixth moment (e.g. $\int_{-\infty}^{\infty} s^6 w(s)ds < \infty$). Let $R = \int_{-\infty}^{\infty} w(s)^2 ds$ denote the roughness of w. Let

$$K(s) = \int_{-\infty}^{s} k(u)du$$

denote the integrated kernel of k and define the constant

$$\psi = 2\int_{-\infty}^{\infty} sK(s)k(s)ds > 0. \tag{1}$$

For our applications, we set both $k(s)$ and $w(s)$ to equal the normal kernel $\phi(s)$, in which case $R = 1/2\sqrt{\pi}$, $\psi = 1/\sqrt{\pi}$ and $K(s) = \Phi(s)$.

Section 2 presents unsmoothed estimators. Section 3 introduces our smoothed estimators. Section 4 develops plug-in bandwidth rules for the local linear and smoothed local linear estimator. Section 5 presents a numerical simulation of performance. Section 6 is a simple application to the U.S. quarterly real GDP growth rate. Section 7 concludes. The proof of Theorem 1 is presented in the Appendix.

Gauss code which implements the estimators and bandwidth methods is available on the author's webpage http://www.ssc.wisc.edu/~bhansen/

2 Unsmoothed Estimators

Note that $F(y \mid x) = E(1(Y_i \leq y) \mid X_i = x)$, and thus it is a regression function. Thus the CDF may be estimated by regression of $1(Y_i \leq y)$ on X_i. A simple nonparametric estimator is the Nadaraya-Watson (NW) estimator, and takes the form of a local average

$$\hat{F}_b(y \mid x) = \frac{\sum_{i=1}^n w_i 1(Y_i \leq y)}{\sum_{i=1}^n w_i} \tag{2}$$

where

$$w_i = \frac{1}{b} w\left(\frac{x - X_i}{b}\right) \tag{3}$$

are kernel weights and b is a bandwidth.

Reduced bias may be achieved using a local linear (LL) estimator. Local linear estimation is a special case of local polynomial regression, and is typically recommended for estimation mean. For a general treatment see Fan and Gijbels (1996). The estimator is the intercept from a weighted least-squares regression of $1(Y_i \leq y)$ on $(x - X_i)$ using the weights w_i, and can be written as

$$\begin{aligned}
\tilde{F}_b(y \mid x) &= \frac{\sum_{i=1}^n w_i^* 1(Y_i \leq y)}{\sum_{i=1}^n w_i^*} && (4)\\
w_i^* &= w_i\left(1 - \hat{\beta}(x - X_i)\right) && (5)\\
\hat{\beta} &= \left(\sum_{i=1}^n w_i (x - X_i)^2\right)^{-1}\left(\sum_{i=1}^n w_i (x - X_i)\right).
\end{aligned}$$

As shown in Theorem 1(b) of Hall, Wolff and Yao (1999), if $b = cn^{-1/5}$ then

$$E\left(\tilde{F}_b(y \mid x) - F(y \mid x)\right)^2 \simeq \frac{R(F(y \mid x)(1 - F(y \mid x)))}{g(x)nb} + \frac{b^4}{4} F^{(2)}(y \mid x)^2 + O\left(n^{-6/5}\right).$$

Integrating over y we obtain the MISE

$$\int_{-\infty}^{\infty} E\left(\tilde{F}_b\left(y \mid x\right) - F\left(y \mid x\right)\right)^2 dy = \frac{RV}{gnb} + \frac{b^4 V_1}{4} + O\left(n^{-6/5}\right) \tag{6}$$

where

$$\begin{aligned} g &= g(x) \\ V &= \int_{-\infty}^{\infty} F\left(y \mid x\right)\left(1 - F\left(y \mid x\right)\right) dy \\ V_1 &= \int_{-\infty}^{\infty} F^{(2)}\left(y \mid x\right)^2 dy. \end{aligned}$$

The first term on the right side of (6) is the integrated asymptotic variance, and the second is the integrated squared asymptotic bias. This bias term is simplified relative to the NW estimator, and this is the source of the improved performance of $\tilde{F}$ over $\hat{F}$.

Hall, Wolff and Yao (1999) point out that the local linear estimator $\tilde{F}$ has two undesirable properties. It may be nonmonotonic in y, and is not constrained to lie in $[0, 1]$. Thus the estimator $\tilde{F}$ is not a CDF. The problem occurs when the weights (5) are negative, which occurs when $\hat{\beta}\left(x - X_i\right) > 1$. A simple solution[1] is to modify w_i^* to exclude negative values. Thus we redefine the weights as follows

$$w_i^* = \begin{cases} 0 & \hat{\beta}\left(x - X_i\right) > 1 \\ w_i\left(1 - \hat{\beta}\left(x - X_i\right)\right) & \hat{\beta}\left(x - X_i\right) \leq 1 \end{cases} \tag{7}$$

This modification is asymptotically negligible, but constrains $\tilde{F}$ to be a valid CDF. Using the modified weights (7) our estimator is a slight modification of the local linear estimator, however we will continue to refer to it as the local linear (LL) estimator for simplicity.

3 Smoothed Estimators

For some bandwidth $h > 0$ let $K_h(s) = K(s/h)$ and consider the integrated kernel smooth $K_h\left(y - Y_i\right)$. For small h it can be viewed as a smooth approximation to the indicator function $1\left(Y_i \leq y\right)$. Smoothed CDF estimators replace the indicator function $1\left(Y_i \leq y\right)$ in (2) and (4) with $K_h\left(y - Y_i\right)$, yielding the smoothed Nadaraya-Watson (SNW) estimator

$$\hat{F}_{h,b}\left(y \mid x\right) = \frac{\sum_{i=1}^n w_i K_h\left(y - Y_i\right)}{\sum_{i=1}^n w_i}$$

[1] Hall, Wolff and Yao (1999) proposed an alternative solution, introducing weights to the Nadaraya-Watson estimator which are selected by an empirical likelihood criterion. Their estimator is asymptotically equivalent to ours, but computationally more complicated.

and the smoothed local linear (SLL) estimator

$$\tilde{F}_{h,b}(y \mid x) = \frac{\sum_{i=1}^{n} w_i^* K_h(y - Y_i)}{\sum_{i=1}^{n} w_i^*}.$$

The weights w_i and w_i^* are defined in (3) and (7), respectively.

Theorem 1 *Assume that $F(y \mid x)$ and $g(x)$ are continuously differentiable up to the fourth order in both y and x. If $b = cn^{-1/5}$ and $h = O(b)$ as $n \to \infty$, then*

$$\int_{-\infty}^{\infty} E\left(\tilde{F}_{h,b}(y \mid x) - F(y \mid x)\right)^2 dy = \frac{R}{gnb}(V - h\psi) + \frac{b^4 V_1}{4} - \frac{h^2 b^2 V_2}{2} + \frac{h^4 V_3}{4} + O\left(n^{-6/5}\right) \quad (8)$$

where

$$\begin{aligned} g &= g(x) \\ V &= \int_{-\infty}^{\infty} F(y \mid x)(1 - F(y \mid x))\, dy \\ V_1 &= \int_{-\infty}^{\infty} F^{(2)}(y \mid x)^2\, dy \\ V_2 &= \int_{-\infty}^{\infty} f(y \mid x) f^{(2)}(y \mid x)\, dy \\ V_3 &= \int_{-\infty}^{\infty} f'(y \mid x)^2\, dy. \end{aligned}$$

The MISE in (8) is a generalization of (6) to allow for $h \geq 0$. The MISE is strictly decreasing at $h = 0$ (since $\psi > 0$) demonstrating that the asymptotic MISE is minimized for strictly positive h. Thus there are (at least asymptotic) efficiency gains from smoothing.

4 Plug-In Bandwidth Selection

We suggest selecting the bandwidths (h, b) by the plug-in method. Plug-in bandwidths are the values which minimize an estimate of the asymptotic MISE (8). Estimation of the latter requires estimation of the parameters g, V, V_1, V_2, and V_2. In this section we discuss estimation of these components, and implementation of the plug-in bandwidth rule.

Readers who are interested in the plug-in algorithm but not the motivational details can jump directly to Section 4.5, which summarizes the implementation details.

4.1 Kernel Estimation of g

The normal kernel density estimator is

$$\hat{g} = \frac{1}{n} \sum_{j=1}^{n} \phi_r(x - X_i) \quad (9)$$

where r is a bandwidth. An estimate of the asymptotic MSE-minimizing choice for r is

$$\hat{r} = \left(\frac{\hat{g}_s(x)}{2\sqrt{\pi}\left(\hat{g}_t^{(2)}(x)\right)^2 n} \right)^{1/5} \tag{10}$$

where

$$\begin{aligned}
\hat{g}_s(x) &= \frac{1}{n}\sum_{j=1}^{n} \phi_s \left(x - X_i\right) \\
\hat{g}_t^{(2)}(x) &= \frac{1}{n}\sum_{j=1}^{n} \phi_t^{(2)} \left(x - X_i\right)
\end{aligned}$$

and s and t are preliminary bandwidths. Letting $\hat{\sigma}_x$ denote the sample standard deviation for X_i, we suggest the Gaussian reference rules $s = 1.06\hat{\sigma}_x n^{-1/5}$ and $t = .94\hat{\sigma}_x n^{-1/9}$, which are MISE-optimal for estimation of the density and its second derivative when g is Gaussian.

4.2 Estimation of V

The parameter V is a function of the conditional distribution $F(y \mid x)$. A parametric estimator of the latter is the intercept from a p'th order polynomial regression of $1(Y_i \le y)$ on $(X_i - x)$. For $p \ge 0$ let

$$X_p = \begin{pmatrix} 1 & X_1 - x & \cdots & (X_1 - x)^p \\ \vdots & \vdots & & \vdots \\ 1 & X_n - x & \cdots & (X_n - x)^p \end{pmatrix}, \tag{11}$$

and let $1(Y \le y)$ denote the stacked $1(Y_i \le y)$. Then the estimator can be written as

$$\hat{F}(y \mid x) = \delta_1' \left(X_p' X_p\right)^{-1} X_p' 1 (Y \le y) \tag{12}$$

where $\delta_1 = (1\ 0\ \cdots\ 0)'$ is the first unit vector. Given (12), an estimate of V is

$$\begin{aligned}
\hat{V} &= \int_{-\infty}^{\infty} \hat{F}(y \mid x)\left(1 - \hat{F}(y \mid x)\right) dy \\
&= \delta_1' \left(X_p' X_p\right)^{-1} X_p' \int_{-\infty}^{\infty} 1(Y > y)\, 1(Y \le y)'\, dy X_p \left(X_p' X_p\right)^{-1} \delta_1 \\
&= \delta_1' \left(X_p' X_p\right)^{-1} X_p' \left(Y - Y'\right)^{+} X_p \left(X_p' X_p\right)^{-1} \delta_1
\end{aligned} \tag{13}$$

where $(a)^+ = a1(a \ge 0)$ for scalar a, and is an element-by-element operation for matrices.

A nonparametric estimator of V can be based on a local linear regression. Given a bandwidth q_0, let $w_{0i} = \phi_{q_0}(x - X_i)$ denote Gaussian kernel weights and set $W_0 = diag\{w_{0i}\}$. Then the estimator is

$$\tilde{F}(y \mid x) = \delta_1' \left(X_1' W_0 X_1\right)^{-1} X_1' W_0 1 (Y \le y)$$

and the estimate of V is

$$\tilde{V} = \delta_1' \left(X_1' W_0 X_1\right)^{-1} X_1' W_0 \left(Y - Y'\right)^+ W_0 X_1 \left(X_1' W_0 X_1\right)^{-1} \delta_1. \tag{14}$$

The bandwidth q_0 which minimizes the asymptotic mean integrated squared error of $\tilde{F}(y \mid x)$ is

$$q_0 = \left(\frac{V}{2\sqrt{\pi} g V_1 n}\right)^{1/5}$$

which depends on the unknowns g, V and V_1. (See equation (8) with $h = 0$.) Using the estimates $\hat{g}$ from (9), $\hat{V}$ from (13), and $\hat{V}_1$ from (17) below, we obtain a plug-in bandwidth

$$\hat{q}_0 = .78 \left(\frac{\hat{V}}{\hat{g}\hat{V}_1 n}\right)^{1/5}. \tag{15}$$

4.3 Estimation of V_1

The parameter V_1 is a function of the second derivative $F^{(2)}(y \mid x)$ which can be estimated by polynomial regression. For some $p \geq 2$ let X_p be defined as in (11). The estimator is

$$\hat{F}^{(2)}(y \mid x) = \delta_3' \left(X_p' X_p\right)^{-1} X_p' 1 (Y \leq y) \tag{16}$$

where δ_3 is the third unit vector (a one in the third place, and zeros elsewhere). To calculate V_2 from (16), observe that for any $c > \max Y_i$

$$\int_{-\infty}^{c} 1(Y \leq y) 1(Y \leq y)' dy = Dc - \max\left(Y, Y'\right)$$

where D is an $n \times n$ matrix of ones, and observe that $\delta_3' \left(X_p' X_p\right)^{-1} X_p' D = 0$. Thus

$$\begin{aligned}
\int_{-\infty}^{c} \hat{F}^{(2)}(y \mid x)^2 dy &= \delta_3' \left(X_p' X_p\right)^{-1} X_p' \int_{-\infty}^{c} 1(Y \leq y) 1(Y \leq y)' dy X_p \left(X_p' X_p\right)^{-1} \delta_3 \\
&= \delta_3' \left(X_p' X_p\right)^{-1} X_p' \left(Dc - \max\left(Y, Y'\right)\right) X_p \left(X_p' X_p\right)^{-1} \delta_3 \\
&= -\delta_3' \left(X_p' X_p\right)^{-1} X_p' \max\left(Y, Y'\right) X_p \left(X_p' X_p\right)^{-1} \delta_3
\end{aligned}$$

where the max operator is element-by-element. Note that this expression is invariant to $c > \max Y_i$, thus we define

$$\hat{V}_1 = -\delta_3' \left(X_p' X_p\right)^{-1} X_p' \max\left(Y, Y'\right) X_p \left(X_p' X_p\right)^{-1} \delta_3. \tag{17}$$

This estimator was also used in (15) to construct a plug-in rule for the nonparametric estimate $\tilde{V}$.

A nonparametric estimator of V_1 can be based on a local cubic regression. (While any local polynomial regression of order two or above is consistent, local cubic regression is recommend for estimation of a regression second derivative.) Given a bandwidth q_1, let $w_{1i} = \phi_{q_1}(x - X_i)$ denote Gaussian kernel weights and set $W_1 = diag\{w_{1i}\}$. The estimator of $F^{(2)}(y \mid x)$ is

$$\tilde{F}^{(2)}(y \mid x) = \delta_3' \left(X_3' W_1 X_3\right)^{-1} X_3' W_1 1 (Y \leq y) \tag{18}$$

and the estimate of V_1 is

$$\tilde{V}_1 = \delta_3' \left(X_3' W_1 X_3\right)^{-1} X_3' W_1 \max\left(Y, Y'\right) W_1 X_3 \left(X_3' W_1 X_3\right)^{-1} \delta_3. \tag{19}$$

Using Fan and Gijbels (1996, Theorem 3.1 and equation (3.20)) the bandwidth which minimizes the asymptotic mean integrated squared error of $\hat{F}^{(2)}(y \mid x)$ is

$$q_1 = 1.01 \left(\frac{V}{g V_4 n}\right)^{1/9}$$

where

$$V_4 = \int_{-\infty}^{\infty} F^{(4)}(y \mid x)^2 \, dy.$$

The constant V_4 may be estimated by

$$\hat{V}_4 = -\delta_5' \left(X_p' X_p\right)^{-1} X_p' \max\left(Y, Y'\right) X_p \left(X_p' X_p\right)^{-1} \delta_5 \tag{20}$$

where $p \geq 4$ and δ_5 is the fifth unit vector, suggesting the plug-in bandwidth

$$\hat{q}_1 = 1.01 \left(\frac{\hat{V}}{\hat{g} \hat{V}_4 n}\right)^{1/9}. \tag{21}$$

4.4 Estimation of V_2 and V_3

If $f'(y \mid x)$ does not depend on x then we have the simplification $V_3 = \int_{-\infty}^{\infty} f'(y)^2 \, dy$ which is a well-studied estimation problem. For a bandwidth a^* let X_0 be the $n \times 1$ vector of ones and

$$\hat{f}_{a^*}'(y) = \frac{1}{n} \sum_{j=1}^{n} \phi_{a^*}^{(1)}(y - Y_i) = \phi_{a^*}^{(1)}\left(y - Y'\right) X_0 \left(X_0' X_0\right)^{-1}$$

be a Gaussian kernel estimate of $f'(y)$. An estimate of V_3 in this case is

$$\begin{aligned} \int_{-\infty}^{\infty} \hat{f}_{a^*}'(y)^2 \, dy &= \left(X_0' X_0\right)^{-1} X_0' \int_{-\infty}^{\infty} \phi_{a^*}^{(1)}(y - Y) \phi_{a^*}^{(1)}\left(y - Y'\right) dy X_0 \left(X_0' X_0\right)^{-1} \\ &= -\left(X_0' X_0\right)^{-1} X_0' \phi_a^{(2)}\left(Y - Y'\right) X_0 \left(X_0' X_0\right)^{-1} \end{aligned} \tag{22}$$

where $a = \sqrt{2} a^*$. The second equality is equation (7.1) of Marron and Wand (1992):

$$\int_{-\infty}^{\infty} \phi_a^{(r)}(y - Y_i) \phi_a^{(s)}(y - Y_j) \, dy = (-1)^r \phi_{\sqrt{2}a}^{(r+s)}(Y_i - Y_j). \tag{23}$$

The estimator (22) is the Jones and Sheather (1991) estimator. Hansen (2004) describes a multiple plug-in method for selection of a. We recommend this choice and hold it fixed for the remainder of this sub-section and the next.

Now consider estimation of V_3 allowing for dependence on x. Note that the estimate of $f'(y)$ in (24) is a regression of $\phi_{a^*}^{(1)}(y - Y_i)$ on a constant. Similarly, we can estimate $f'(y \mid x)$ as the intercept in a polynomial regression on $x_i - x$. For X_p defined in (11) let

$$\hat{f}_a'(y \mid x) = \delta_1' \left(X_p' X_p\right)^{-1} X_p' \phi_{a^*}^{(1)}(y - Y). \tag{24}$$

The associated estimate of V_3 is

$$\begin{aligned}
\hat{V}_3 &= \int_{-\infty}^{\infty} \hat{f}_{a^*}'(y \mid x)^2 \, dy \\
&= \delta_1' \left(X_p' X_p\right)^{-1} X_p' \int_{-\infty}^{\infty} \phi_{a^*}^{(1)}(y - Y) \phi_{a^*}^{(1)}(y - Y') \, dy X_p \left(X_p' X_p\right)^{-1} \delta_1 \\
&= -\delta_1' \left(X_p' X_p\right)^{-1} X_p' \phi_a^{(2)}(Y - Y') X_p \left(X_p' X_p\right)^{-1} \delta_1. \qquad (25)
\end{aligned}$$

Alternatively, we could use a fully nonparametric estimator such as a local linear estimator. However, the optimal bandwidths depend on derivatives of the conditional density that are difficult to estimate. We thus do not consider further complications.

In addition, using an analog of (16) and (24), a simple estimator of V_2 is

$$\begin{aligned}
\hat{V}_2 &= \int_{-\infty}^{\infty} \hat{f}_{a^*}^{(2)}(y \mid x) \hat{f}_{a^*}(y) \, dy \\
&= \delta_3' \left(X_p' X_p\right)^{-1} X_p' \int_{-\infty}^{\infty} \phi_{a^*}(y - Y) \phi_{a^*}(y - Y') \, dy X_p \left(X_p' X_p\right)^{-1} \delta_1 \\
&= \delta_3' \left(X_p' X_p\right)^{-1} X_p' \phi_a (Y - Y') X_p \left(X_p' X_p\right)^{-1} \delta_1. \qquad (26)
\end{aligned}$$

where the third equality uses (23).

4.5 Plug-In Bandwidth

Our plug-in estimate of (8) is

$$\tilde{M}(h, b) = \frac{R}{\hat{g} n b} \left(\tilde{V} - h\psi\right) + \frac{b^4 \tilde{V}_1}{4} - 2\frac{h^2 b^2 \hat{V}_2}{2} + \frac{h^4 \hat{V}_3}{4}. \tag{27}$$

The plug-in bandwidth pair $(\tilde{h}, \tilde{b})$ jointly minimize $\tilde{M}(h, b)$. A closed-form solution is not available, so the function needs to be numerically minimized. One caveat is that we must constrain $h < \psi/\tilde{V}$ to ensure a sensible solution. (When $h > \psi/\tilde{V}$ then $\tilde{M}$ is strictly increasing in b so unconstrained minimization sets $\tilde{b} = 0$ which is not sensible.) As a practical matter it makes sense to bound h more substantially, we suggest the constraint $h \leq \psi/2\tilde{V}$, and this is done in our numerical applications.

The plug-in bandwidth for the unsmoothed estimator $\tilde{F}_b$ is found by setting $h = 0$ and minimizing to find $\hat{b} = \left(R\tilde{V}/\hat{g}\tilde{V}_1 n\right)^{1/6}$.

To summarize, the computation algorithm is as follows. First, $\hat{g}$ is calculated from (9) using (10). Second, for some $p \geq 4$, $\hat{V}$, $\hat{V}_1$ and $\hat{V}_4$ are calculated using (13), (17) and (20). Third, using the bandwidth (15), $\tilde{V}$ is calculated using (14). Fourth, using the bandwidth (21), $\tilde{V}_1$ is calculated using (19). Fifth, the bandwidth a is calculated using the method of Hansen (2004), allowing $\hat{V}_2$ and $\hat{V}_3$ to be calculated using (26) and (25). Finally, $\hat{g}$, $\tilde{V}$, $\tilde{V}_1$, $\hat{V}_2$, $\hat{V}_3$ are set into (27) which is numerically minimized over (h, b) to obtain $(\tilde{h}, \tilde{b})$.

5 Simulation

The performance of the CDF estimators is investigated in a simple simulation experiment using bivariate data. The variables X_i are generated as iid $N(0,1)$. The variable Y_i is generated as $Y_i = X_i + e_i$ where e_i is iid, independent of X_i. The error e_i is distributed as G, one of the first nine mixture-normal distributions from Marron and Wand (1992). Consequently, the true CDF for Y_i is $G(y - x)$. For each choice of G we generated 10,000 independent samples of size $n = 50$, 100 and 300.

For each sample we estimated the CDF at $x = 1$ and $y = \{-2.0, -1.5, -1.0, -0.5, 0.0, 0.5, 1.0, 1.5, 2.0\}$, using the NW, SNW, LL and SLL estimators. Our measure of precision is the exact mean-squared error, averaged over these values of y (as an approximation to the MISE). The standard normal kernel was used for all estimates.

We first abstracted from the issue of bandwidth selection and evaluated the precision of the four estimates using the infeasible finite-sample-optimal bandwidth, calculated by quasi-Newton minimization. The results are presented in the first three columns of Table 1. We have taken the unsmoothed NW estimator as a baseline, and have divided the MISE of the other estimators by that of the NW estimator. Thus the numbers in Table 1 represent the relative improvement attained relative to the NW estimator.

The results show that across models and sample sizes, the SLL estimator achieves the lowest MISE. In each case, using local linear estimation rather than local averaging reduces the MISE, and using smoothing lowers the MISE.

We next investigated the performance of our plug-in bandwidths $\hat{b}$ for the LL estimator and $(\tilde{h}, \tilde{b})$ for the SLL estimator. On each sample described above we calculated these plug-in bandwidths and CDF estimators, and evaluated the MISE across the 10,000 samples. The results are presented in the final two columns of Table 1. As before, the MISE is normalized by the MISE of the infeasible NW estimator. Thus the numbers represented the relative improvement attained by the feasible plug-in estimators relative to this infeasible estimator.

In each case, the plug-in SLL estimator has lower MISE than the plug-in LL estimator. In some cases the difference is minor, in other case it is quite substantial (the largest gains are for Model #5, where the MISE is reduced by about 25%). In each case the MISE is higher than if the infeasible optimal bandwidth had been used, in some cases substantially higher, suggesting that further improvements in bandwidth choice may be possible.

Overall, the simulation evidence confirms the implications of the asymptotic analysis. We can recommend use of the SLL estimator (with our proposed plug-in bandwidth) for empirical practice.

6 GDP Growth Forecasting

We illustrate the method with a simple application to the U.S. real GDP growth rate. Let GDP_i denote the level of quarterly real GDP and let $y_i = 400\left(\ln\left(GDP_i\right) - \ln\left(GDP_{i-1}\right)\right)$ denote the annualized quarterly growth rate. Set $x_i = y_{i-1}$.

We restrict the sample to run from first third quarter of 1984 to the first quarter of 2004 (79 observations). This choice is consistent with the decline in output volatility documented by McConnell and Perez-Quiros (2000). We estimated the CDF using the local linear (LL) and smoothed local linear (SLL) estimators for y_i given $x_i = x$, fixing x to equal 2% and 4.5%, which are approximately the 25'th and 75'th quantiles of the unconditional distribution. The CDF was evaluated at 501 evenly spaced increments between -2 and 7, and the bandwidths were calculated by the plug-in method described in Section 4.5. To provide a contrast, we also estimated the CDF using a homoskedastic Gaussian AR(1).

Figure 1 displays the three CDFs for the case of $x = 2.0$ and Figure 2 for the case $x = 4.5$. In both cases, the three estimates are meaningfully different from one another. The parametric estimate has a fairly different location and variance. The two nonparametric estimates have the same basic shape, but the unsmoothed local linear estimate is visually quite erratic, and the smoothed local linear estimator appears to apply an intuitively correct degree of smoothing.

It is also interesting to note that the LL estimator uses a plug-in bandwidth of $\hat{b} = 1.59$ in Figure 1 and $\hat{b} = 2.13$ in Figure 2, while the SLL estimator uses $\tilde{b} = 1.44$, $\tilde{h} = 0.53$ and $\tilde{b} = 2.0$ and $\tilde{h} = 0.23$ respectively. In both cases, the plug-in rule selects similar values for b for the two estimators (with that for SLL slightly smaller), and selects a considerably smaller value for h than b.

7 Conclusion

We have shown how to combine local linear methods and smoothing to produce good nonparametric estimates of the conditional distribution function for bivariate data. We have shown that meaningful improvements in estimation efficiency can be achieved by using these techniques. Furthermore, we have derived data-based plug-in rules for bandwidth selection which appear to work in practice. These tools should prove useful for empirical application.

This analysis has been confined to bivariate data. An important extension would to the case of multivariate data. In particular, this extension would require the study and development of feasible bandwidth selection rules.

8 Appendix: Proof of Theorem 1

The proof of Theorem 1 is based on the following set of expansions.

Lemma 1 *Let* $g^{(s)} = \frac{d^s}{dx^s} g(x)$ *and* $K_h^*(Y_i) = K_h(y - Y_i) - F(y \mid x)$.

$$
\begin{aligned}
Ew_i &= g(x) + \frac{b^2}{2} g^{(2)}(x) + O(b^4) && (28)\\
E\left(w_i (x - X_i)\right) &= -b^2 g^{(1)}(x) + O(b^4) && (29)\\
E\left(w_i (x - X_i)^2\right) &= b^2 g(x) + O(b^4) && (30)\\
E\left(w_i K_h^*(Y_i)\right) &= \frac{b^2}{2} F^{(2)}(y \mid x) g(x) + b^2 F^{(1)}(y \mid x) g^{(1)}(x) && (31)\\
&\quad + \frac{h^2}{2} f'(y \mid x) g(x) + O\left(b^4\right) \\
E\left(w_i (x - X_i) K_h^*(Y_i)\right) &= -b^2 F^{(1)}(y \mid x) g(x) + O\left(b^4\right) && (32)\\
E\left(w_i^2 K_h^*(Y_i)^2\right) &= \frac{Rg(x)\left(F(y \mid x)\left(1 - F(y \mid x)\right) - h\psi f(y \mid x)\right)}{b} && (33)\\
&\quad + O(b) \\
E\left(w_i^2 (x - X_i)^2 K_h^*(Y_i)^2\right) &= O(b) && (34)
\end{aligned}
$$

Proof. Throughout, we make use of the assumption that $h = O(b)$ to simplify the bounds. The derivations make use of the following Taylor series expansions which are valid under the assumption that the functions are fourth order differentiable:

$$
\begin{aligned}
g(x - bv) &= g(x) - vbg^{(1)}(x) + \frac{v^2 b^2}{2} g^{(2)}(x) - \frac{v^3 b^3}{6} g^{(3)}(x) + O(b^4) && (35)\\
F(y - hu \mid v) &= F(y \mid v) - huf(y \mid v) + \frac{h^2 u^2}{2} f'(y \mid v) - \frac{h^3 u^3}{6} f''(y \mid v) + O(h^4) && (36)\\
F(y \mid x - bv) &= F(y \mid x) - bvF^{(1)}(y \mid x) + \frac{b^2 v^2}{2} F^{(2)}(y \mid x) - \frac{b^3 v^3}{6} F^{(3)}(y \mid x) + O(b^4) && (37)\\
f'(y \mid x - bv) &= f'(y \mid x) - bv \frac{\partial}{\partial y} f^{(1)}(y \mid x) + O(b^2). && (38)
\end{aligned}
$$

To show (28), by a change of variables and (35)

$$
\begin{aligned}
Ew_i &= \int_{-\infty}^{\infty} \frac{1}{b} w\left(\frac{x - v}{b}\right) g(v) dv \\
&= \int_{-\infty}^{\infty} w(v) g(x - bv) dv \\
&= \int_{-\infty}^{\infty} w(v) \left(g(x) - vbg^{(1)}(x) + \frac{v^2 b^2}{2} g^{(2)}(x) - \frac{v^3 b^3}{6} g^{(3)}(x)\right) dv + O(b^4) \\
&= g(x) + \frac{b^2}{2} g^{(2)}(x) + O(b^4).
\end{aligned}
$$

Note that this makes use of the facts that $\int_{-\infty}^{\infty} w\left(v\right) dy = 1$, $\int_{-\infty}^{\infty} w\left(v\right) v = 0$, $\int_{-\infty}^{\infty} w\left(v\right) v^2 = 1$ and $\int_{-\infty}^{\infty} w\left(v\right) v^3 = 0$.

Similarly

$$\begin{aligned}
E\left(w_i\left(x - X_i\right)\right) &= \int_{-\infty}^{\infty} \left(x - v\right) \frac{1}{b} w\left(\frac{x - v}{b}\right) g(v) dv \\
&= b \int_{-\infty}^{\infty} v w\left(v\right) g(x - bv) dv \\
&= b \int_{-\infty}^{\infty} v w\left(v\right) \left(g(x) - vbg^{(1)}(x) + \frac{v^2 b^2}{2} g^{(2)}(x)\right) dv + O(b^4) \\
&= -b^2 g^{(1)}(x) + O(b^4)
\end{aligned}$$

yielding (29) and

$$\begin{aligned}
E\left(w_i\left(x - X_i\right)^2\right) &= \int_{-\infty}^{\infty} \left(x - v\right)^2 \frac{1}{b} w\left(\frac{x - v}{b}\right) g(v) dv \\
&= b^2 \int_{-\infty}^{\infty} v^2 w\left(v\right) g\left(x - bv\right) dv \\
&= b^2 g(x) + O(b^4)
\end{aligned}$$

yielding (30).

To show (31) and (32), first observe that by integration by parts, a change of variables and (36)

$$\begin{aligned}
\int_{-\infty}^{\infty} K_h\left(y - u\right) f\left(u \mid v\right) du &= \int_{-\infty}^{\infty} k_h\left(y - u\right) F\left(u \mid v\right) du \\
&= \int_{-\infty}^{\infty} k\left(u\right) F\left(y - hu \mid v\right) du \\
&= \int_{-\infty}^{\infty} k\left(u\right) \left(F\left(y \mid v\right) - huf\left(y \mid v\right) + \frac{h^2 u^2}{2} f'\left(y \mid v\right) - \frac{h^3 u^3}{6} f''\left(y \mid v\right)\right) du \\
&\quad + O(h^4) \\
&= F\left(y \mid v\right) + \frac{h^2}{2} f'\left(y \mid v\right) + O(h^4). \qquad (39)
\end{aligned}$$

By a change of variables, and using (35) and (37),

$$\begin{aligned}
&\int_{-\infty}^{\infty} F(y \mid v) \frac{1}{b} w\left(\frac{x-v}{b}\right) g(v) dv \\
= &\int_{-\infty}^{\infty} F(y \mid x-bv) g(x-bv) w(v) dv \\
= &\int_{-\infty}^{\infty} \left(F(y \mid x) - bvF^{(1)}(y \mid x) + \frac{b^2v^2}{2} F^{(2)}(y \mid x) - \frac{b^3v^3}{6} F^{(3)}(y \mid x)\right) \\
&\cdot \left(g(x) - bvg^{(1)}(x) + \frac{b^2v^2}{2} g^{(2)}(x) - \frac{v^3b^3}{6} g^{(3)}(x)\right) w(v) + O(b^4) \\
= &\; F(y \mid x) g(x) + \frac{b^2}{2} F(y \mid x) g^{(2)}(x) + \frac{b^2}{2} F^{(2)}(y \mid x) g(x) + b^2 F^{(1)}(y \mid x) g^{(1)}(x) + O(b^4). \quad (40)
\end{aligned}$$

Similarly, by a change of variables and (35) and (38),

$$\begin{aligned}
\int_{-\infty}^{\infty} f'(y \mid v) \frac{1}{b} w\left(\frac{x-v}{b}\right) g(v) dv &= \int_{-\infty}^{\infty} f'(y \mid x-bv) g(x-bv) w(v) dv \\
&= f'(y \mid x) g(x) + O(b^2). \quad (41)
\end{aligned}$$

Together, (28), (39), (40) and (41) show that

$$\begin{aligned}
E\left(w_i K_h^*(Y_i)\right) &= E\left(w_i K_h(Y_i)\right) - E(w_i) F(y \mid x) \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{b} w\left(\frac{x-v}{b}\right) K_h(y-u) f(u \mid v) g(v) du dv \\
&\quad - \left(g(x) + \frac{b^2}{2} g^{(2)}(x)\right) F(y \mid x) + O(b^4) \\
&= \int_{-\infty}^{\infty} F(y \mid v) \frac{1}{b} w\left(\frac{x-v}{b}\right) g(v) dv - \left(g(x) + \frac{b^2}{2} g^{(2)}(x)\right) F(y \mid x) \\
&\quad + \frac{h^2}{2} \int_{-\infty}^{\infty} f'(y \mid v) \frac{1}{b} w\left(\frac{x-v}{b}\right) g(v) dv + O(b^4) \\
&= \frac{h^2}{2} f'(y \mid x) g(x) + \frac{b^2}{2} F^{(2)}(y \mid x) g(x) + b^2 F^{(1)}(y \mid x) g^{(1)}(x) \\
&\quad + O\left(b^4\right)
\end{aligned}$$

which is (31).

Similarly, using (29), (39) and similar derivations,

$$
\begin{aligned}
E\left(w_i\left(x-X_i\right)K_h^*(Y_i)\right) &= E\left(w_i\left(x-X_i\right)K_h(Y_i)\right) - E\left(w_i\left(x-X_i\right)\right)F\left(y \mid x\right) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(x-v\right)\frac{1}{b}w\left(\frac{x-v}{b}\right)K_h\left(y-u\right)f\left(u \mid v\right)g(v)dvdu \\
&\quad +b^2 g^{(1)}(x)F\left(y \mid x\right) + O(b^4) \\
&= b\int_{-\infty}^{\infty} vF\left(y \mid x-bv\right)g(x-bv)w\left(v\right)dv \\
&\quad +b\frac{h^2}{2}\int_{-\infty}^{\infty} vf'\left(y \mid x-bv\right)g(x-bv)w\left(v\right)dv \\
&\quad +b^2 g^{(1)}(x)F\left(y \mid x\right) + O(b^4) \\
&= -b^2 F^{(1)}\left(y \mid x\right)g(x) + O\left(b^4\right)
\end{aligned}
$$

which is (32).

To show (33) first observe that by a change of variables and (35),

$$
\begin{aligned}
E\left(w_i^2\right) &= \int_{-\infty}^{\infty}\frac{1}{b^2}w\left(\frac{x-v}{b}\right)^2 g\left(v\right)dv \\
&= \frac{1}{b}\int_{-\infty}^{\infty} w\left(v\right)^2 g\left(x-bv\right)dv \\
&= \frac{1}{b}\int_{-\infty}^{\infty} w\left(v\right)^2\left(g(x)-vbg^{(1)}(x)\right)dv + O(b) \\
&= \frac{Rg(x)}{b} + O(b).
\end{aligned}
\tag{42}
$$

Second, using (39), a change of variables, (35) and (37),

$$
\begin{aligned}
E\left(w_i^2 K_h(Y_i)\right) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{b}w\left(\frac{x-v}{b}\right)^2 K_h\left(y-u\right)f\left(u \mid v\right)g(v)dvdu \\
&= \int_{-\infty}^{\infty}\frac{1}{b^2}w\left(\frac{x-v}{b}\right)^2 g(v)F\left(y \mid v\right)dv + O\left(\frac{h^2}{b}\right) \\
&= \frac{1}{b}\int_{-\infty}^{\infty} w\left(v\right)^2 g(x-vb)F\left(y \mid x-vb\right)dv + O\left(b\right) \\
&= \frac{1}{b}\int_{-\infty}^{\infty} w\left(v\right)^2\left(g(x)-vbg^{(1)}(x)\right)\left(F\left(y \mid x\right)-vbF^{(1)}\left(y \mid x\right)\right)dv + O(b^2) \\
&= \frac{Rg(x)F\left(y \mid x\right)}{b} + O(b).
\end{aligned}
\tag{43}
$$

Third, by integration by parts, a change of variables, and (36)

$$\begin{aligned}\int_{-\infty}^{\infty} K_h(y-u)^2 f(u \mid v)\,du &= 2\int_{-\infty}^{\infty} K_h(y-u)\,k_h(y-u)\,F(u \mid v)\,du \\ &= 2\int_{-\infty}^{\infty} K(u)\,k(u)\,F(y-hu \mid v)\,du \\ &= 2\int_{-\infty}^{\infty} K(u)\,k(u)\,(F(y \mid v) - huf(y \mid v))\,du + O(h^2) \\ &= F(y \mid v) - h\psi f(y \mid v) + O(h^2). \qquad (44)\end{aligned}$$

The final equality uses the definition (1) and the fact that the symmetry of $k(u)$ about zero implies

$$\int_{-\infty}^{\infty} 2K(u)\,k(u)\,du = 0.$$

Fourth, using (44), (35), (37) and (38),

$$\begin{aligned}E\left(w_i^2 K_h(Y_i)^2\right) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{b^2} w\left(\frac{x-v}{b}\right)^2 K_h(y-u)^2 f(u \mid v)\,g(v)dvdu \\ &= \int_{-\infty}^{\infty} \frac{1}{b^2} w\left(\frac{x-v}{b}\right)^2 (F(y \mid v) - h\psi f(y \mid v))\,g(v)dv + O\left(\frac{h^2}{b}\right) \\ &= \frac{1}{b}\int_{-\infty}^{\infty} w(v)^2 (F(y \mid x-vb) - h\psi f(y \mid x-vb))\,g(x-vb)dv + O(b) \\ &= \frac{Rg(x)\,(F(y \mid x) - h\psi f(y \mid x))}{b} + O(b) \qquad (45)\end{aligned}$$

Together, (42), (43) and (45) show

$$\begin{aligned}E\left(w_i^2 K_h^*(Y_i)^2\right) &= E\left(w_i^2 K_h(Y_i)^2\right) - 2E\left(w_i^2 K_h(Y_i)\right) F(y \mid x) + E\left(w_i^2\right) F(y \mid x)^2 \\ &= \frac{Rg(x)\,(F(y \mid x) - h\psi f(y \mid x)) - 2Rg(x)F(y \mid x)^2 + Rg(x)F(y \mid x)^2}{b} + O(b) \\ &= \frac{Rg(x)\,(F(y \mid x)\,(1 - F(y \mid x)) - h\psi f(y \mid x))}{b} + O(b)\end{aligned}$$

which is (33). (34) is shown similarly.

Proof of Theorem 1: From equations (29)-(30),

$$\begin{aligned}E\hat{\beta}_x &= \left(E\left(w_i(x-X_i)^2\right)\right)^{-1}(E(w_i(x-X_i))) + O\left(b^2\right) \\ &= -g(x)^{-1}g^{(1)}(x) + O\left(b^2\right).\end{aligned}$$

Combined with (28), this yields

$$\begin{aligned} Ew_i^* &= Ew_i - E\hat{\beta}E\left(w_i\left(x-X_i\right)\right)+O(b^2) \\ &= f(x)+O(b^2). \end{aligned} \tag{46}$$

Next, using equations (31) and (32),

$$\begin{aligned} E\left(w_i^*K_h^*(Y_i)\right) &= E\left(w_iK_h^*(Y_i)\right)-E\hat{\beta}E\left(w_i\left(x-X_i\right)K_h^*(Y_i)\right)+O\left(b^4\right) \\ &= \frac{b^2}{2}F^{(2)}\left(y\mid x\right)g(x)+\frac{h^2}{2}f'\left(y\mid x\right)g(x)+O\left(b^4\right). \end{aligned}$$

Together with (46) we find

$$\begin{aligned} E\left(\tilde{F}_{h,b}\left(y\mid x\right)-F\left(y\mid x\right)\right) &= \frac{Ew_i^*K_h^*\left(y-Y_i\right)}{Ew_i^*}+O\left(b^4\right) \\ &= \frac{1}{2}\left(b^2F^{(2)}\left(y\mid x\right)+h^2f'\left(y\mid x\right)\right)+O\left(b^4\right). \end{aligned}$$

Similarly, using (33) and (34)

$$\begin{aligned} Var\left(\tilde{F}_{h,b}\left(y\mid x\right)\right) &= \frac{1}{n}\frac{Var\left(w_i^*K_h^*\left(y-Y_i\right)\right)}{\left(Ew_i^*\right)^2} \\ &= \frac{1}{n}\frac{E\left(w_iK_h^*\left(y-Y_i\right)\right)^2}{g(x)^2}+O\left(\frac{b}{n}\right) \\ &= \frac{R\left(F\left(y\mid x\right)\left(1-F\left(y\mid x\right)\right)-h\psi f\left(y\mid x\right)\right)}{g(x)nb}+O\left(\frac{b}{n}\right)O\left(n^{-6/5}\right). \end{aligned}$$

Together, these expressions and the assumption that $b=cn^{-1/5}$ establish that

$$\begin{aligned} E\left(\tilde{F}_{h,b}\left(y\mid x\right)-F\left(y\mid x\right)\right)^2 &= \frac{R\left(F\left(y\mid x\right)\left(1-F\left(y\mid x\right)\right)-\psi hf\left(y\mid x\right)f(x)\right)}{g(x)nb}+\frac{h^4f'\left(y\mid x\right)^2}{4} \\ &\quad +\frac{h^2b^2f'\left(y\mid x\right)F^{(2)}\left(y\mid x\right)}{2}+\frac{b^4f'\left(y\mid x\right)F^{(2)}\left(y\mid x\right)^2}{4}+O\left(n^{-6/5}\right). \end{aligned}$$

Integrating over y we obtain

$$\begin{aligned} \int_{-\infty}^{\infty}E\left(\tilde{F}_{h,b}\left(y\mid x\right)-F\left(y\mid x\right)\right)^2dy &= \frac{R\left(\int_{-\infty}^{\infty}F\left(y\mid x\right)\left(1-F\left(y\mid x\right)\right)dy-\psi h\int_{-\infty}^{\infty}f\left(y\mid x\right)dyf(x)\right)}{g(x)nb} \\ &\quad +\frac{h^4}{4}\int_{-\infty}^{\infty}f'\left(y\mid x\right)^2dy+\frac{h^2b^2}{2}\int_{-\infty}^{\infty}f'\left(y\mid x\right)F^{(2)}\left(y\mid x\right)dy \\ &\quad +\frac{b^4}{4}\int_{-\infty}^{\infty}F^{(2)}\left(y\mid x\right)^2dy+O\left(n^{-6/5}\right) \end{aligned}$$

which equals the expression in (8) using the equivalence

$$\int_{-\infty}^{\infty} f'(y \mid x)\, F^{(2)}(y \mid x)\, dy = -\int_{-\infty}^{\infty} f(y \mid x)\, f^{(2)}(y \mid x)\, dy$$

which holds by integration by parts. This completes the proof.

Table 1:
Normalized MISE
Conditional Distribution Function Estimators

$n = 50$	Optimal Bandwidth			Plug-In Bandwidths	
Error Density	SNW	LL	SLL	LL	SLL
Gaussian	.76	.70	.56	.79	.67
Skewed Unimodal	.80	.64	.53	.69	.60
Strongly Skewed	.89	.85	.78	.88	.82
Kurtotic Unimodal	.82	.77	.66	.81	.72
Outlier	.65	.69	.49	.99	.76
Bimodal	.76	.71	.55	.84	.72
Separated Bimodal	.93	.82	.77	.95	.92
Asymmetric Bimodal	.76	.68	.52	.80	.68
Trimodal	.78	.72	.56	.86	.74

$n = 100$	Optimal Bandwidth			Plug-In Bandwidths	
Error Density	SNW	LL	SLL	LL	SLL
Gaussian	.80	.67	.58	.71	.63
Skewed Unimodal	.83	.63	.54	.66	.58
Strongly Skewed	.84	.89	.84	.92	.88
Kurtotic Unimodal	.85	.76	.68	.81	.73
Outlier	.61	.61	.44	1.06	.80
Bimodal	.90	.66	.55	.72	.64
Separated Bimodal	.96	.80	.77	.84	.81
Asymmetric Bimodal	.81	.64	.52	.69	.62
Trimodal	.83	.68	.57	.74	.66

$n = 300$	Optimal Bandwidth			Plug-In Bandwidths	
Error Density	SNW	LL	SLL	LL	SLL
Gaussian	.86	.64	.59	.66	.61
Skewed Unimodal	.88	.60	.55	.62	.58
Strongly Skewed	.97	.93	.91	1.20	1.16
Kurtotic Unimodal	.87	.73	.68	.85	.79
Outlier	.58	.51	.39	.97	.75
Bimodal	.88	.61	.56	.63	.59
Separated Bimodal	.97	.78	.76	.81	.80
Asymmetric Bimodal	.89	.59	.52	.61	.57
Trimodal	.95	.62	.65	.65	.61

References

[1] Fan, Jianqing and Irene Gijbels (1996): *Local Polynomial Modelling and Its Applications*: New York: Chapman and Hall.

[2] Fan, Jianqing and Qiwei Yao (2003): *Nonlinear Time Series: Nonparametric and Parametric Methods.* New York: Springer.

[3] Hall, Peter, Rodney C. L. Wolff, and Qiwei Yao (1999): "Methods for Estimating a Conditional Distribution Function," *Journal of the American Statistical Association*, 94, 154-163.

[4] Hansen, Bruce E. (2004): "Bandwidth Selection for Nonparametric Distribution Estimation", manuscript, University of Wisconsin.

[5] Jones, M.C. and S. J. Sheather (1991): "Using non-stochastic terms to advantage in kernel-based estimation of integrated squared density derivatives," *Statistics and Probability Letters*, 11, 511-514.

[6] Marron, J.S. and M.P. Wand (1992): "Exact mean integrated squared error," *Annals of Statistics*, 20, 712-736.

[7] McConnell, M.M. and G. Perez-Quiros (2000): "Output fluctuations in the United States: What has changed since the early 1980s?" *American Economic Review*, 90, 1464-1476.

Figure 1:
Estimated Conditional Distribution of U.S. GDP Growth
Conditional on Current Rate = 2%

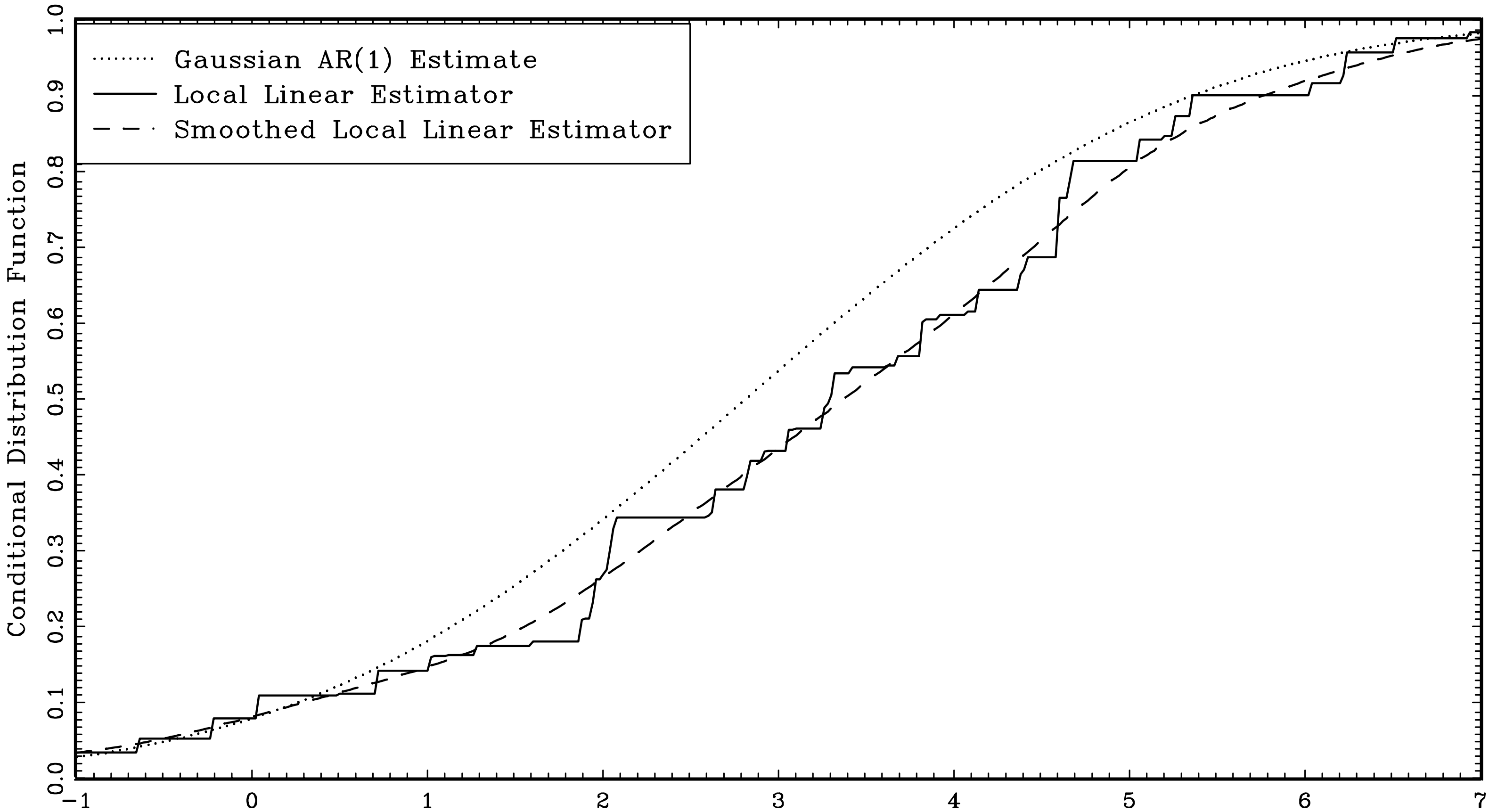

Figure 2:
Estimated Conditional Distribution of U.S. GDP Growth
Conditional on Current Rate = 4.5%