and Marron's (A.1)–(A.5),

$$\text{(3)} \quad \operatorname{Var}(\hat{m}_T(x)) = \frac{\sigma^2}{nh} \int K^2 + o((nh)^{-1}).$$

This implies that $\hat{m}_T$ with bandwidth h has the same asymptotic variance as $\hat{m}_E$ with the bandwidth $h_x = h/f(x)$. In particular, the limiting variances of $\hat{m}_T$ and $\hat{m}_E$ are the same in a case highlighted by Chu and Marron, that is, when $X_1, \ldots, X_n$ are a random sample from a $U(0,1)$ distribution.

The bias of $\hat{m}_T(x)$ has the representation (again under assumptions akin to (A.1)–(A.5))

$$\text{(4)} \quad \begin{aligned} &\operatorname{Bias}(\hat{m}_T(x)) \\ &\quad = \frac{h^2}{2}(mQ)''(F(x)) \int u^2K + o(h^2) \\ &\quad = \frac{h^2}{2}\left\{\frac{m''(x)f(x) - m'(x)f'(x)}{f^3(x)}\right\} \int u^2K \\ &\qquad + o(h^2). \end{aligned}$$

In general, $\operatorname{Bias}(\hat{m}_T)$ is different from both $\operatorname{Bias}(\hat{m}_E)$ and $\operatorname{Bias}(\hat{m}_C)$; this is true even if one allows the bandwidths of $\hat{m}_E$ and $\hat{m}_C$ to vary with x a la $h_x = h/(f(x))^{\alpha}$. By considering (3) and (4) above, and Sections 3 and 4 of Chu and Marron, one finds, not surprisingly, that $\operatorname{MSE}(\hat{m}_T)$ is not comparable with either $\operatorname{MSE}(\hat{m}_C)$ or $\operatorname{MSE}(\hat{m}_E)$. It is worth noting, though, that when $X_1, \ldots, X_n$ are iid $U(0,1)$, the asymptotic MSEs of $\hat{m}_T$ and $\hat{m}_E$ are identical when the two estimators use the same bandwidth.

Introducing the estimator $\hat{m}_T$ certainly does not settle the mean squared error issue. However, $\hat{m}_T$ is attractive in that it avoids both the random denominator problem of $\hat{m}_E$ and the down weighting pathology of $\hat{m}_C$. Another nice feature of $\hat{m}_T$ is that, like $\hat{m}_C$, it has a convenient form for estimating m', so long as $\hat{F}$ is differentiable. Considering $\hat{m}_T$ also brings into light the question of estimating the regression-quantile function mQ, an object whose importance has been stressed by Parzen (1981). Since it is natural to use a fixed, evenly spaced design on $[0,1]$ to estimate mQ, the convolution estimator seems ideally suited for estimating regression-quantile functions.

My final point concerns the use of kernel methods to test the adequacy of linear models. I was glad that Chu and Marron mentioned the problem of testing for linearity, and the attendant importance of how $\hat{m}_C$ and $\hat{m}_E$ perform when m is a straight line. I prefer $\hat{m}_C$ over $\hat{m}_E$ for purposes of testing linearity, since, as Chu and Marron point out, $\hat{m}_C$ has smaller bias than $\hat{m}_E$ in the straight line case. Indeed, Hart and Wehrly (1991) show that a boundary-corrected version of $\hat{m}_C$ (with bandwidth h) tends to a straight line as h tends to infinity. The limiting line is a consistent estimator of m when $m(x) = \beta_0 + \beta_1 x$. Higher-order kernels can be used to obtain kernel estimates that are polynomials (of any given degree) for large h. Such kernel estimates are a crucial part of a test proposed by Hart and Wehrly (1991) for checking the fit of a polynomial.

Comment

M. C. Jones

It is a great pleasure to congratulate the authors on a most informative, thought-provoking and, above all, *balanced* investigation of the issues involved in choosing between versions of the kernel regression estimator.

Chu and Marron (henceforth C&M) understandably concentrate on comparing and contrasting the two kernel estimators probably most widely employed in the literature: the Nadaraya-Watson (N-W) estimator, $\hat{m}_E$, and the Gasser-Müller

M. C. Jones is Lecturer, Department of Statistics, The Open University, Milton Keynes MK7 6AA, United Kingdom.