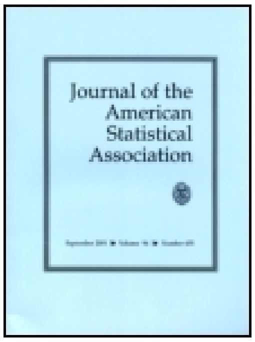

Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Comment

Jianqing Fan & Yang Feng

To cite this article: Jianqing Fan & Yang Feng (2009) Comment, Journal of the American Statistical Association, 104:487, 1003-1007, DOI: 10.1198/jasa.2009.tm09188

To link to this article: https://doi.org/10.1198/jasa.2009.tm09188

Published online: 01 Jan 2012.

Kim, J., and Gleser, L. J. (2000), "SIMEX Approaches to Measurement Error in ROC Studies," *Communications in Statistics. Theory and Methods*, 29, 2473–2491.
Kulathinal, S. B., Kuulasmaa, K., and Gasbarra, D. (2002), "Estimation of an Errors-in-Variables Regression Model When the Variances of the Measurement Errors Vary Between the Observations," *Statistics in Medicine*, 21, 1089–1101.
Li, T., and Vuong, Q. (1998), "Nonparametric Estimation of the Measurement Error Model Using Multiple Indicators," *Journal of Multivariate Analysis*, 65, 139–165.
Linton, O., and Whang, Y. J. (2002), "Nonparametric Estimation With Aggregated Data," *Econometric Theory*, 18, 420–468.
Marron, J. S., and Wand, M. P. (1992), "Exact Mean Integrated Squared Error," *The Annals of Statistics*, 20, 712–736.
National Research Council, Committee on Pesticides in the Diets of Infants and Children (1993), *Pesticides in the Diets of Infants and Children*, Washington, DC: National Academies Press.
Schennach, S. M. (2004a), "Estimation of Nonlinear Models With Measurement Error," *Econometrica*, 72, 33–75.
——— (2004b), "Nonparametric Regression in the Presence of Measurement Error," *Econometric Theory*, 20, 1046–1093.
Simonoff, J. S. (1996), *Smoothing Methods in Statistics*, New York: Springer.
Staudenmayer, J., and Ruppert, D. (2004), "Local Polynomial Regression and Simulation–Extrapolation," *Journal of the Royal Statistical Society, Ser. B*, 66, 17–30.
Staudenmayer, J., Ruppert, D., and Buonaccorsi, J. (2008), "Density Estimation in the Presence of Heteroskedastic Measurement Error," *Journal of the American Statistical Association*, 103, 726–736.
Stefanski, L. A. (2000), "Measurement Error Models," *Journal of the American Statistical Association*, 95, 1353–1358.
Stefanski, L. A., and Carroll, R. J. (1990), "Deconvoluting Kernel Density Estimators," *Statistics*, 21, 165–184.
Taupin, M. L. (2001), "Semi-Parametric Estimation in the Nonlinear Structural Errors-in-Variables Model," *The Annals of Statistics*, 29, 66–93.
Thamerus, M. (2003), "Fitting a Mixture Distribution to a Variable Subject to Heteroscedastic Measurement Errors," *Computational Statistics*, 18, 1–17.
Wand, M. P., and Jones, M. C. (1995), *Kernel Smoothing*, London: Chapman & Hall.

Comment

Jianqing FAN and Yang FENG

We would like to congratulate Carroll, Delaigle, and Hall on their important and stimulating contributions to nonparametric prediction in measurement error models. The paper deals with nonparametric prediction with measurements of different quality (contaminated errors). They also studied the difficulty of nonparametric prediction, which depends on the relative order magnitude of tail behavior of characteristic functions. They observed an interesting phenomenon that the convergence rate of the estimator can be as fast as $O(n^{-1/2})$ in certain cases, and considered the case when the error densities are unknown. The simulation studies and the real data example show the advantages of their estimator over the Nadaraya–Watson estimator. We appreciate the opportunity to discuss the paper and provide additional insights.

1. BEST LINEAR UNBIASED ESTIMATOR

When the measurement errors are heterogeneous across samples, the essential idea is to aggregate the information from each observation to estimate the characteristic function. The question arises naturally how to optimally aggregate them. We appeal to the best linear unbiased estimator.

To simplify the notation, instead of following the main paper, we let $\phi_V(t) = \mathrm{E}e^{itV}$ be the characteristic function of a random variable V. Following the other notations and assumptions in the main paper, we have

$$\phi_T(t) = \phi_X(t)\phi_{U^F}(t)$$

and we need to estimate $\phi_X(t)$ from the observable data $\{W_j\}$, which is related to ϕ_X through

$$\phi_{W_j}(t) = \phi_X(t)\phi_{U_j}(t).$$

Jianqing Fan is Frederick L. Moore Professor of Finance (E-mail: *jqfan@princeton.edu*) and Yang Feng is Ph.D. Candidate (E-mail: *yangfeng@princeton.edu*), Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08544. The work was supported by the NIH grant R01-GM072611 and NSF grants DMS-0704337 and DMS-0714554.

It is easy to calculate that

$$\mathrm{E}e^{itW_j} = \phi_X(t)\phi_{U_j}(t)$$

and that

$$\begin{aligned}\mathrm{Var}(e^{itW_j}) &= \mathrm{E}\big(e^{itW_j} - \phi_{W_j}(t)\big)\big(e^{-itW_j} - \phi_{W_j}(-t)\big)\\ &= 1 - \big|\phi_X(t)\phi_{U_j}(t)\big|^2.\end{aligned}$$

Hence, a sequence of unbiased estimators of $\phi_X(t)$ is $\{e^{itW_j}/\phi_{U_j}(t)\}$ with heterogeneous variance

$$V_j(t) = \big|\phi_{U_j}(t)\big|^{-2} - |\phi_X(t)|^2.$$

Thus, $\phi_X(t)$ can be estimated by the following linear unbiased estimator:

$$\hat{\phi}_X(t) = \sum_{j=1}^{n} a_j(t)e^{itW_j}/\phi_{U_j}(t) \tag{1.1}$$

for some vector $\mathbf{a}(t) = (a_1(t), a_2(t), \ldots, a_n(t))^T$ with $\sum_{j=1}^n a_j(t) = 1$. The corresponding variance of the estimator will be

$$V_{\mathbf{a}}(t) = \sum_{j=1}^{n} a_j(t)^2 V_j(t).$$

The weights of the Best Linear Unbiased Estimator (BLUE) are given by

$$a_{j,0}(t) = \big(\big|\phi_{U_j}(t)\big|^{-2} - |\phi_X(t)|^2\big)^{-1}/A(t), \quad \text{for } j = 1, 2, \ldots, n, \tag{1.2}$$

Journal of the American Statistical Association
September 2009, Vol. 104, No. 487, Theory and Methods
DOI: 10.1198/jasa.2009.tm09188

where $A(t) = \sum_j (|\phi_{U_j}(t)|^{-2} - |\phi_X(t)|^2)^{-1}$. The corresponding minimum variance is $1/A(t)$ and the BLUE of the characteristic function $\phi_X(t)$ is

$$\hat{\phi}_{X,0}(t) = \frac{\sum_{j=1}^n (|\phi_{U_j}(t)|^{-2} - |\phi_X(t)|^2)^{-1} e^{itW_j}/\phi_{U_j}(t)}{\sum_{j=1}^n (|\phi_{U_j}(t)|^{-2} - |\phi_X(t)|^2)^{-1}}. \quad (1.3)$$

The quantity ϕ_X in the ideal weight (1.2) is unknown, but this can be easily estimated consistently by a simple plug-in procedure. A simple approach is to drop the unknown term ϕ_X out from (1.2), which results in the weight

$$a_{j,1}(t) = |\phi_{U_j}(t)|^2 \Big/ \sum_{j=1}^n |\phi_{U_j}(t)|^2, \quad (1.4)$$

and the estimate

$$\hat{\phi}_{X,1}(t) = \frac{\sum_{j=1}^n \phi_{U_j}(-t) e^{itW_j}}{\sum_{j=1}^n |\phi_{U_j}(t)|^2}. \quad (1.5)$$

This is exactly the same as the one used in the main paper under discussion. Clearly, our estimator (1.3) is more efficient, at least in terms of estimating characteristic functions.

The plug-in version of (1.3) is now given by

$$\hat{\phi}_{X,2}(t) = \frac{\sum_{j=1}^n (|\phi_{U_j}(t)|^{-2} - |\hat{\phi}_{X,1}(t)|^2)^{-1} e^{itW_j}/\phi_{U_j}(t)}{\sum_{j=1}^n (|\phi_{U_j}(t)|^{-2} - |\hat{\phi}_{X,1}(t)|^2)^{-1}}. \quad (1.6)$$

We can iterate the plug-in process until convergence and this leads to the estimator $\hat{\phi}_{X,3}(t)$, which solves

$$\phi_{X,3}(t) = \frac{\sum_{j=1}^n (|\phi_{U_j}(t)|^{-2} - |\phi_{X,3}(t)|^2)^{-1} e^{itW_j}/\phi_{U_j}(t)}{\sum_{j=1}^n (|\phi_{U_j}(t)|^{-2} - |\phi_{X,3}(t)|^2)^{-1}}. \quad (1.7)$$

As anticipated, in Section 3, we will show both $\hat{\phi}_{X,2}$ and $\hat{\phi}_{X,3}$ perform similarly to $\hat{\phi}_{X,0}$. Thus, $\hat{\phi}_{X,2}(t)$ suffices for practice.

2. ESTIMATORS BASED ON BLUE

With an estimate of the characteristic function $\hat{\phi}_X(t)$ as in (1.1), we can estimate the marginal density $f_T(x)$ of $T = X + U$ and the conditional mean function $\mu(x) = E(Y|T = x)$ in the same manner as the main paper. More specifically, for a given kernel function K and a bandwidth h, by the Fourier inversion theorem, we define

$$\hat{f}_T(x) = (2\pi)^{-1} \int e^{-itx} \phi_U(t)\hat{\phi}_X(t)\phi_K(th)\, dt$$
$$= \sum_{j=1}^n K_{h,j}(x - W_j), \quad (2.1)$$

where the induced kernel $K_{h,j}(z)$ is given by

$$K_{h,j}(z) = (2\pi)^{-1} \int e^{-itz} \phi_K(th) \frac{a_j(t)\phi_U(t)}{\phi_{U_j}(t)}\, dt.$$

Using the above induced kernel, we can get an estimator of $\mu(x)$ as

$$\hat{\mu}(x) = \frac{\sum_{j=1}^n Y_j K_{h,j}(x - W_j)}{\sum_{j=1}^n K_{h,j}(x - W_j)}. \quad (2.2)$$

Note that the characteristic function $\phi_K(th)$ is introduced to damp down the estimate $\hat{\phi}_T(t) = \phi_U(t)\hat{\phi}_X(t)$ when $|t|$ is large, whether the function $|\phi_U(t)\hat{\phi}_X(t)|$ is integrable or not. In particular, when the function $|\phi_U(t)\hat{\phi}_X(t)|$ is integrable, we can take h as small as zero. In this case, the estimator is independent of K and h.

If we take

$$a_j(t) = n^{-1}\phi_{U_j}(t)/\phi_U(t),$$

the estimators (2.1) and (2.2) reduce respectively to the kernel density estimator and the Nadaraya–Watson estimator. The total weight is not necessarily one and the estimator is not necessarily consistent. This estimator does not take into account heterogeneity in the measurement error and thus is not optimal in general. Our new estimator is based on the ideal weight (1.2) or its estimated version. The estimator introduced by Carroll, Delaigle, and Hall takes weights (1.4). In other words, different estimators lie in choosing different weight functions $\{a_j(t)\}$.

3. PERFORMANCE COMPARISON

We now study the efficiency gain by using the BLUE for estimating the characteristic function. To facilitate the performance comparison, we consider the two-error model as in the paper, where we assume that the first m observations are contaminated by an error with density $f_{U^{(1)}}$, and the last $n - m$ observations are contaminated by an error with density $f_{U^{(2)}}$, and the error in the future observations is $U_F \sim f_{U^{(1)}}$. In this case, the optimal weights in (1.2) take only two values. For this two-error model, the variance function for $\hat{\phi}_{X,0}(t)$ in (1.3) can be simplified as

$$Z_0(t) = 1/\big(m\big(|\phi_{U^{(1)}}(t)|^{-2} - |\phi_X(t)|^2\big)^{-1} + (n-m)\big(|\phi_{U^{(2)}}(t)|^{-2} - |\phi_X(t)|^2\big)^{-1}\big)$$

and the variance function for $\hat{\phi}_{X,1}(t)$ in (1.5) equals to

$$Z_1(t) = \big(m|\phi_{U^{(1)}}(t)|^4\big(|\phi_{U^{(1)}}(t)|^{-2} - |\phi_X(t)|^2\big) + (n-m)|\phi_{U^{(2)}}(t)|^4\big(|\phi_{U^{(2)}}(t)|^{-2} - |\phi_X(t)|^2\big)\big) \big/ \big(m|\phi_{U^{(1)}}(t)|^2 + (n-m)|\phi_{U^{(2)}}(t)|^2\big)^2.$$

It is clear that $Z_0(t) \le Z_1(t)$.

To gauge the gain by using the BLUE, we calculate the ratio $Z_1(t)/Z_0(t)$ for four different error density settings with $m = n/2$. The results are shown in Figure 1. Notice that in setting A, where the two different error densities are chosen to be the same as the first example in the paper, the ratio is approximately equal to 1. In setting B, we increase the variances. In this case, the ratios are greater than 1 for a range of t. Similar results can be found for error density settings C and D. If the error distribution is more extreme, then the ratios will be even larger. In general, the method of Carroll, Delaigle, and Hall is nearly BLUE, when the contaminated noise is relatively small.

To evaluate the effectiveness of the plug-in version of BLUE, we conduct a simulation experiment to compare the relative performance of $\hat{\phi}_{X,1}(t)$, $\hat{\phi}_{X,2}(t)$, and $\hat{\phi}_{X,3}(t)$ with the oracle estimator $\hat{\phi}_{X,0}(t)$. The variance of each estimator is computed based on 1000 simulations with $m = n/2 = 250$. Again, we define the corresponding variance functions to be $Z_1(t), Z_2(t)$, and $Z_3(t)$ and render the ratios Z_1/Z_0, Z_2/Z_0, and Z_3/Z_0 in Figure 2 under the same settings as in Figure 1. It is worth noticing that both the ratios Z_2/Z_0 and Z_3/Z_0 almost equal to 1. This shows that the

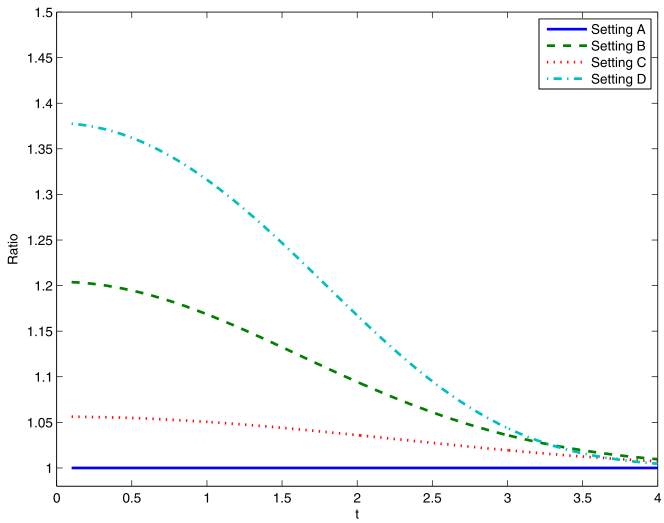

Figure 1. Ratio $Z_1(t)/Z_0(t)$ when setting A: $U^{(1)} \sim$ Normal and $U^{(2)} \sim$ Laplace, with $\sigma^2_{U^{(1)}} = \sigma^2_{U^{(2)}} = 0.2\,\mathrm{Var}(X)$; setting B: $U^{(1)} \sim$ Normal and $U^{(2)} \sim$ Laplace, with $\sigma^2_{U^{(1)}} = 0.25\,\mathrm{Var}(X)$ and $\sigma^2_{U^{(2)}} = 2\,\mathrm{Var}(X)$; setting C: $U^{(1)} \sim$ Normal and $U^{(2)} \sim$ Normal, with $\sigma^2_{U^{(1)}} = 0.25\,\mathrm{Var}(X)$ and $\sigma^2_{U^{(2)}} = \mathrm{Var}(X)$; setting D: $U^{(1)} \sim$ Normal and $U^{(2)} \sim$ Normal, with $\sigma^2_{U^{(1)}} = 3\,\mathrm{Var}(X)$ and $\sigma^2_{U^{(2)}} = 0.25\,\mathrm{Var}(X)$.

plug-in BLUE $\hat{\phi}_{X,2}(t)$, the fully iterated estimator $\hat{\phi}_{X,3}(t)$, and the oracle estimator have a similar performance. For this reason, the plug-in BLUE suffices for practical purpose, which outperforms $\hat{\phi}_{X,1}(t)$ in the main paper. The question is how much this gain will translate into the gain in estimating $\hat{\mu}(x)$ as defined in (1.2). We appeal to the asymptotic analysis.

4. ASYMPTOTIC PROPERTIES

Following the same lines of the proof for Theorem 4.1 in the main paper, we can derive the asymptotic property for the corresponding estimator $\hat{\mu}(x)$ when we use the weights $\mathbf{a}$ as in (1.1). The upper bound of the asymptotic variance is given by

$$n \int |\phi_K(th)|^2 |\phi_{U^F}(t)|^2 V_{\mathbf{a}}(t)\, dt, \tag{4.1}$$

where $V_{\mathbf{a}}(t)$ is the variance of estimated characteristic function

$$V_{\mathbf{a}}(t) = \sum_{j=1}^{n} |a_j(t)|^2 \left(\left|\phi_{U_j}(t)\right|^{-2} - |\phi_X(t)|^2 \right).$$

In other words, the weighted integral of $V_{\mathbf{a}}(t)$ contributes to the asymptotic variance. Minimizing (4.1) would give the same solution $a_{j,0}(t)$ as given by (1.2). The finite sample gain over the estimator proposed by Carroll, Delaigle, and Hall remains to be seen.

From (4.1), it reveals that the individual data contributes to the asymptotic variance reflected in the expression

$$\int |\phi_K(th)|^2 \left|\phi_{U^F}(t)/\phi_{U_j}(t)\right|^2 dt,$$

the smaller the more informative. When the function $|\phi_{U^F}(t)/\phi_{U_j}(t)|^2$ is integrable, the variance is bounded. If it is not integrable, it depends on the tail behavior of $|\phi_{U^F}(t)/\phi_{U_j}(t)|$. For the two-error model, the first m data points are of usual quality and the last $n-m$ can be of much higher or lower quality, depending on the tail behavior of $|\phi_{U^{(1)}}(t)/\phi_{U^{(2)}}(t)|$.

A simple way to understand the rates of convergence for two-error case is as follows. Let $\hat{\mu}_1(t)$ be the Nadaraya–Watson based on the first part of homogeneous sample $(W_1, Y_1), \ldots, (W_m, Y_m)$. Then, it is well known

$$\hat{\mu}_1(t) = \mu(t) + O_P\left(h^k + (mh)^{-1/2}\right). \tag{4.2}$$

Based on the remaining $(n-m)$ homogeneous data points, we have an independent estimator

$$\hat{\mu}_2(t) = \sum_{j=m+1}^{n} Y_j K_{h,1}(x - W_j) \Bigg/ \sum_{j=m+1}^{n} K_{h,1}(x - W_j), \tag{4.3}$$

where the induced kernel is now given by

$$K_{h,1}(z) = ((n-m)2\pi)^{-1} \times \int \exp(itz)\phi_K(th)\phi_{U^{(1)}}(t)/\phi_{U^{(2)}}(t)\, dt. \tag{4.4}$$

This estimator has been shown in the main paper to have the following asymptotic representation [see also (4.1)]:

$$\hat{\mu}_2(t) = \mu(t) + O_P\left(h^k + (n-m)^{-1/2} v_1(h)^{1/2}\right), \tag{4.5}$$

where $v_1(h) = \int |\phi_K(th)|^2 |\phi_{U^{(1)}}(t)|^2 / |\phi_{U^{(2)}}(t)|^2\, dt$. Taking the best linear combination of the estimators from the two parts results in

$$\hat{\mu}_*(t) = \lambda(t)\hat{\mu}_1(t) + (1 - \lambda(t))\hat{\mu}_2(t), \tag{4.6}$$

where $\lambda(t)$ is inversely proportional to the ratio of the asymptotic variances of the two estimators. Then, it is clear that the estimator has the rate of convergence as stated in Theorem 4.2 of the paper under discussion, namely

$$\hat{\mu}_*(t) = \mu(t) + O_P\left(h^k + \min\left\{(mh)^{-1/2}, (n-m)^{-1/2} v_1(h)^{1/2}\right\}\right).$$

The above heuristic also provides the intuition why the optimal rate is given in Theorem 4.3 and also renders an alternative procedure (4.6).

5. PARAMETRIC RATES OF CONVERGENCE

We now give an intuitive explanation of the parametric rate of convergence for the two-error model. This is only possible when the second part of $(n-m)$ data points are of much higher quality (with less contaminated errors) than the first part of data and are of nonnegligible sample size. In this case, we can regard $U^{(1)}$ as $U^{(2)}$ with additional source of idiosyncratic noise ε:

$$U^{(1)} = U^{(2)} + \varepsilon. \tag{5.1}$$

Under (5.1), it is well known that the density of $T = X + U^{(1)}$ is given

$$f_T(x) = \int f_\varepsilon(x - u) f_{W^{(2)}}(u)\, du, \tag{5.2}$$

where $W^{(2)} = X + U^{(2)}$ and $f_V(\cdot)$ represents the density of a random variable V. An obvious root-n consistent estimator is

$$\hat{f}_T(x) = (n-m)^{-1} \sum_{i=m+1}^{n-m} f_\varepsilon(x - W_i)\, du, \tag{5.3}$$

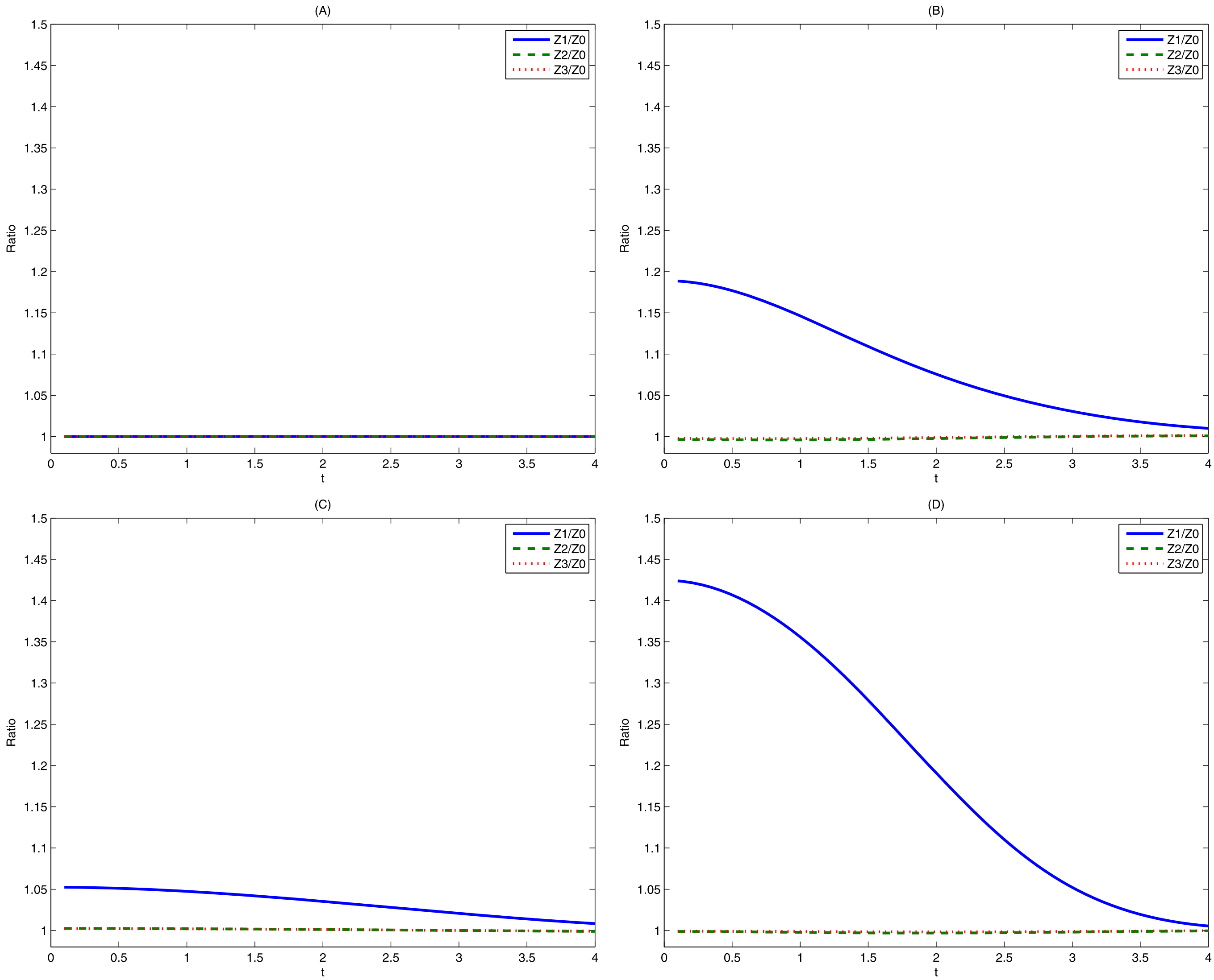

Figure 2. Ratios of the variance function of three estimated characteristic functions with that using the oracle weight. The four error density settings are the same as in Figure 1.

which does not require any smoothing.

Similarly, the function $D(x) = \int y f_{Y,T}(y,x)\,dy$ can be expressed as

$$D(x) = \int f_\varepsilon(x-u) D_2(u)\,du$$
$$= \int\int f_\varepsilon(x-u) y f_{Y,W^{(2)}}(y,u)\,du\,dy, \qquad (5.4)$$

where $D_2(x) = \int y f_{Y,W^{(2)}}(y,x)\,dy$ can be directly estimated from second half of the data. This leads to the estimate

$$\hat{D}(x) = (n-m)^{-1} \sum_{i=m+1}^{n} f_\varepsilon(x - W_i) Y_i, \qquad (5.5)$$

which is of the parametric rate.

The above argument is based upon the assumption (5.1). Even when (5.1) does not hold, (5.2) and (5.4) can still hold, as long as $|\phi_{U^{(1)}}(t)/\phi_{U^{(2)}}(t)|$ is integrable. In this case, we can define $f_\varepsilon(x)$ through the Fourier inversion

$$f_\varepsilon(x) = (2\pi)^{-1} \int \exp(-itx) \phi_{U^{(1)}}(t)/\phi_{U^{(2)}}(t)\,dt.$$

Note that $f_\varepsilon(x)$ is not necessarily a density function, but it is a known weight function. Through direct estimate of $\hat{D}(x)$ and $\hat{f}_T(x)$ in (5.3) and (5.5), it follows easily that $\mu(x) = D(x)/f_T(x)$ can be estimated at the parametric rate. The estimator $\hat{\mu}_3 = \hat{D}(x)/\hat{f}_T(x)$ can be combined with the Nadaraya–Watson estimator from the first part to yield a simple and new estimator as in (4.6).

6. ADDING NOISE

Carroll, Delaigle, and Hall posed a provoking question in Remark 4.3 whether the efficiency of prediction can be gained by adding a noise. They did not elaborate how to predict Y based on observed data. It is true that $\mu_T(x) = E(Y|T = x)$ can be estimated more accurately as discussed in Section 5. However,

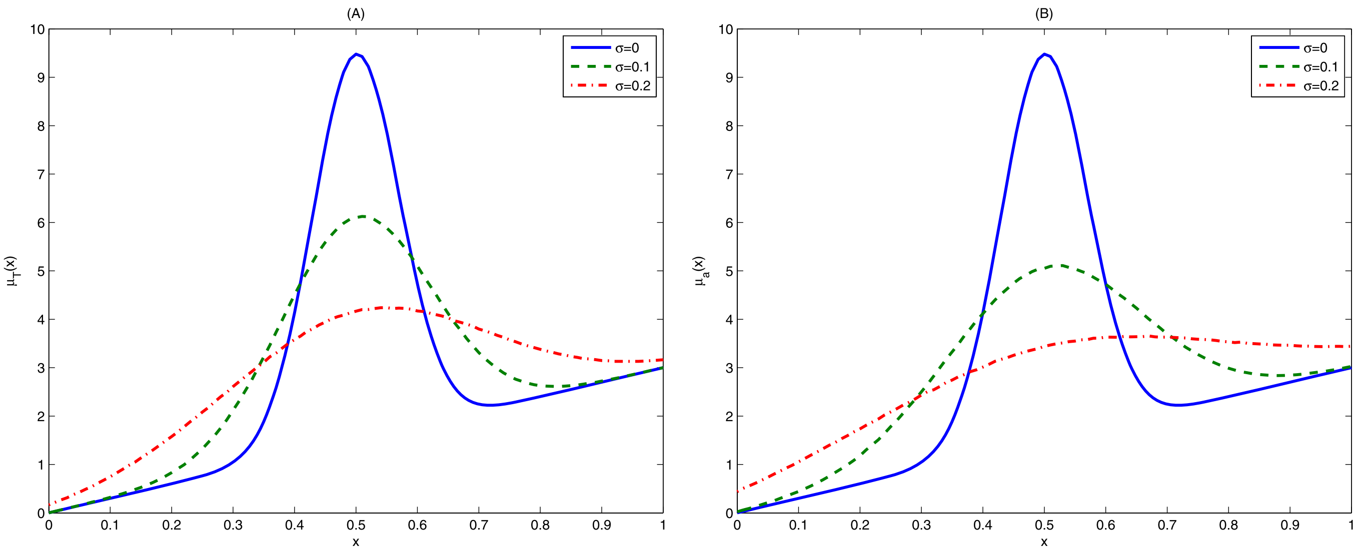

Figure 3. Comparisons of $\mu_T(x) = E(Y|T = x)$ (left panel) and $\mu_a(x) = E\mu_T(x + \epsilon)$ (right panel) for different values of σ. Here $\mu_X(x) = -3x + 20\exp\{-100(x - 0.5)^2\}/\sqrt{2\pi}$ is taken the same function as the first example in the paper.

$\mu_T(x)$ can be very different from $\mu_X(x) = E(Y|X = x)$, causing a lot of biases. Left panel of Figure 3 shows the difference between the two functions for three noise level $\sigma = 0$, $\sigma = 0.1$, and $\sigma = 0.2$ with normal distribution.

Suppose that we wish to predict Y based on $X = x$ and only the function $\mu_T(x)$ is available. One naturally generates many $T_i = x + \varepsilon_i$, gets prediction $\mu_T(x + \varepsilon_i)$ and uses their average as the prediction of Y. This approach essentially computes

$$\mu_a(x) = E\mu_T(x + \varepsilon).$$

This prediction can be seriously biased as shown in the right panel of Figure 3. The authors are welcome to suggest an alternative method of prediction. In general, we suspect that adding noise will improve the prediction.

Comment

Susanne M. Schennach

1. OVERVIEW

The paper "Nonparametric Prediction in Measurement Error Models" by R. J. Carroll, A. Delaigle, and P. Hall considers the general problem of estimating a predictor function $E[Y|T]$, where Y is the variable to be predicted and T is an explanatory variable. T is generated through $T = X + U^F$, where X is the true unobserved explanatory variable and U^F a measurement error. Given a sample $(Y_j, T_j)_{j=1}^n$, this problem would be simple, since the measurement error would play no role in this inference problem—the best predictor could be estimated by a (perhaps nonparametric) least-square regression of Y on T.

However, this problem acquires a highly nontrivial nature when the measurement error U_j of the mismeasured variable W_j used to construct the predictor function (with $W_j = X + U_j$) has a distribution that differs from the measurement error distribution of the explanatory variable T that is actually used to make the prediction. Thanks to this paper, this important setting finally receives the attention it deserves, as it frequently occurs in practice whenever data from different sources are combined. One dataset could contain only a proxy T for the variable of interest Y, while another separate dataset could contain variable both the variable of interest Y and a proxy W, which attempts to measure the same quantity as T, but, due to some difference in data collection or measurement procedures, does so with a different form of measurement error. This paper casts and solves the problem in its most general form, allowing for full hereroscedasticity in the measurement error U_j.

In this comment, we first provide some intuition regarding the origin of the exceptionally fast convergence rates obtained. We then informally suggest a series of modest extensions of the authors' results: (i) a slightly more efficient estimator, (ii) uniform convergence results, and (iii) a novel way to combine the results of the paper with the regression calibration technique in order to further reduce measurement error-induced bias.

2. PARAMETRIC RATES—SOME INTUITION

One of the most captivating results of the paper is the possibility of obtaining parametric convergence rates ($n^{-1/2}$, where

Susanne M. Schennach is Professor, Department of Economics, University of Chicago, Chicago, IL 60637 (E-mail: *smschenn@uchicago.edu*). This work was made possible in part through financial support from the National Science Foundation via grant SES-0752699.

Journal of the American Statistical Association
September 2009, Vol. 104, No. 487, Theory and Methods
DOI: 10.1198/jasa.2009.tm09222