NON-PARAMETRIC REGRESSION FOR ESTIMATING TOTALS IN FINITE POPULATIONS

Alan H. Dorfman, Bureau of Labor Statistics
OSMR, 2 Massachusetts Ave., NE, Washington, DC 20212

Key Words: Best Linear Unbiased Estimation, Calibration Estimator, Kernel Estimation, Weighted Balanced Samples

Overview
Non-parametric regression based on a simple kernel estimator is reviewed, and applied to get $\hat{T}_{np}$, a non-parametric regression based estimator of totals in finite populations. Expressions for the asymptotic bias and variance of $\hat{T}_{np}$ are given, and implications drawn. For example, there is a difference between the ideal bandwidth for constructing $\hat{T}_{np}$ and that required by standard non-parametric regression. An important fact is that var($\hat{T}_{np}$) approximates the variance of a class of best linear unbiased (BLU) parametric estimators of total (those based on 'columnar models') under ideal sample conditions ('weighted balance'). We can usefully combine non-parametric and parametric approaches in a composite estimator, the *non-parametric calibration* estimator $\hat{T}_{cal}$. Its bias properties are noted. A simulation study on a messy data set (the classic Beef Population), suggests that $\hat{T}_{cal}$ has an advantage over other estimators. We also note an anomalous property of the well-known GREG estimator. Conclusions are drawn, other relevant work described, and suggestions made for further work.

Non-parametric Regression

Non-parametric regression has its origins in data exploration. Given a data set $s = \{x_i, y_i\}$, i = 1, 2, ..., n – a "cloud of points" – we want, without detailed data modelling, to get an idea of the relation of y to x, "beneath the cloud of points". Basically, we want to draw a line in the $x - y$ plane through the cloud that shows the essential features of y's dependency on x. [Note: throughout this paper, y and x both will be understood to be scalars.]

To do this, we suppose the expectation of Y is a smooth function of x, that is, we suppose

$$Y = m(x) + \sigma(x)e,$$

where e is white noise, and $m()$ is smooth (continuously differentiable of order at least 2.)

To construct our line, we take a fine, uniform grid of points $r = \{x_j\}$ spanning s, and estimate $m(x_j),\ j \in r$ by $\hat{m}(x_j) = \sum_{i \in s} w_{ij} y_i$, where $\sum_{i \in s} w_{ij} = 1$, and w_{ij} is larger, the closer x_i is to x_j. Connecting the values $\hat{m}(x_j)$ in order of increasing x_j gives a "smooth" line that can give us a good idea of the relationship of y on x.

There are a variety of ways to form the weights w_{ij}. We here focus on what is probably the simplest way, leading to so-called kernel regression smoothing. Let $K(u)$ be a symmetric (about 0) density function, a so-called "kernel" function, preferably with finite support. Examples are the uniform density $K(u) = 0.5I\{-1 \le u \le 1\}$, the bi-square $K(u) = \frac{15}{16}\left(1-u^2\right)^2 I\{-1 \le u \le 1\}$, and the Epanichnikov $K(u) = \frac{3}{4}\left(1-u^2\right)I\{-1 \le u \le 1\}$. (The simulation work discussed below uses the bi-square.) Whatever our choice of basic kernel, we can get a family of densities from $K(u)$ by scale transformation: $K_b(u) = b^{-1}K(u/b)$. The scale parameter b is commonly referred to as the "bandwidth". Often in the literature the bandwidth is symbolized by h, rather than b. Since, in the sampling context, h is often used to refer to strata in stratified sampling, we here prefer b. The kernel based weights are taken as

$w_{ij} = K_b\left(x_i - x_j\right) \Big/ \sum_{i \in s} K_b\left(x_i - x_j\right)$. Note that these weights satisfy the above conditions: they add to 1, and they are larger, the closer x_i is to the "target" x_j. In addition, kernel weights are non-negative. The amount of smoothing depends on the size of b. The smaller b is, the more wiggly the resulting graph. Proper choice of bandwidth is a major issue.

How close is $\hat{m}(x_j)$ to $m(x_j)$? The following theory is well established.

Suppose the x's in s are realizations of independent random variables each with density $d_s(x)$. Then it can be shown (Härdle, 1991) that

$$E(\hat{m}(x_j) - m(x_j)) \approx c_1 \frac{b^2 \beta(x)}{d_s(x_j)},$$

and

$$\mathrm{var}(\hat{m}(x_j)) \approx c_2 \frac{\sigma^2(x_j)}{nbd_s(x_j)},$$

where n is the number of units in s

$c_1 = (1/2)\int u^2 K(u)du$,

$c_2 = \int K^2(u)du$, and

and $\beta(x_j) = m''(x_j)d_s(x_j) + 2m'(x_j)d_s'(x_j)$. From this one can draw that the best bandwidth (giving the minimal mean square error) is $b = k(x_j)n^{-1/5}$, with

$$k(x_j) = \left(\frac{c_2 \sigma^2(x_j)}{4c_1^2 d_s(x_j)[\beta(x_j)/d_s(x_j)]^2} \right)^{1/5}.$$

This means, for example, that, the number of points in s needs to increase 32-fold, other things being equal, merely to halve the optimal bandwidth. This formula by itself does not allow us to determine the best bandwidth, since it depends on unknown quantities.

Estimation of totals in finite populations

We now change context a bit. We consider a finite population P of size N. Values of the variable x are known for the units of the population, and s is an ignorable sample of size n from P, for which y values are known. ("Ignorable" means that, given information on x, knowledge of how the sample was taken provides no additional information about y.)

Suppose we want to estimate the total $T = \sum_P Y_i = \sum_s Y_i + \sum_r Y_i$. Since y values are available to us on s, the problem is essentially to get a reasonable estimate on r, where $r = P - s$ is the "remainder" of the population, outside s. That is, we want the second sum in T above.

A natural idea is to use non-parametric regression to get estimates $\hat{m}(x_j)$, for $j \in r$ and add these up, to get an estimate of $T_r = \sum_r Y_j$. (Note that r is no longer necessarily a nice even gridwork of x's. To save notation in what follows we will consistently use "i" to refer to units *in* the sample s, and "j" for values in r, *i.e.* for units *just* not in sample.) This gives us the *non-parametric (kernel) estimator of total*

$$\hat{T}_{np} = \sum_s Y_i + \sum_{P-s} \hat{m}(x_j) = \sum_s Y_i + \sum_{P-s} \sum_s w_{ij} Y_i$$
$$= \sum_s Y_i + \sum_s w_i Y_i = \sum_s (1 + w_i) Y_i$$

where $w_i = \sum_{j \in P-s} w_{ij}$.

We can make some simple observations:

1) $\hat{T}_{np}$ is linear in the Y's.

2) The estimator is data intensive both in that it requires us to know the values of x for all the units i the population, and requires intensive calculation. The former is the more serious restriction these days.

3) If we compare $\hat{T}_{np}$ to $\hat{T}_\pi = \sum_s Y_i / \pi_i$, the classic design-based expansion estimator, where the π_i's are inclusion probabilities in a randomization based sample, we can see that the $w_i + 1$'s *replace* the π_i's, in the following sense:

If the sampling design is well constructed, π_i represents the *a priori* effective number of population units that are near the ith unit. The $w_i + 1$ give the *de facto* number of such points for the particular sample in hand, using x as the measure of nearness. Thus use of inclusion probabilities in non-parametric based estimation of totals is gratuitous.

We gain further clarity from the following theorem.

Theorem Suppose $Y_i = m(x_i) + \sigma(x_i)e_i$, i = $1,\ldots,N$, with e_i~(0,1) independent. Suppose a sample s of size n is taken, and let $d_s(x)$, $d_{P-s}(x)$ represent the density of sample and non-sample x's respectively. Set $\beta(x) = m''(x)d_s(x) + 2m'(x)d_s'(x)$.

Then

$E(\hat{T}_{np} - T \mid X_P) =$

$$c_1 b^2 (N-n) \int \beta(x) d_s(x)^{-1} d_{P-s}(x) dx + O_p\left((N-n)b^3 + (N-n)n^{-1/2} b^{1/2}\right)$$

and

$$\operatorname{var}\left(\hat{T}_{np} - T \mid X_P\right) =$$
$$(N-n)^2 n^{-1} \int \sigma^2(x) d_s(x)^{-1} [d_{P-s}(x)]^2 dx + c_2 (N-n) n^{-1} b^{-1} \int \sigma^2(x) d_s^{-1}(x) d_{P-s}(x) dx + (N-n)^2 n^{-1} b^2 c_1^2 \int c^*(x) d_s(x) dx + (N-n) \int \sigma^2(x) d_{P-s}(x) dx + O_p\left((N-n)^2 n^{-1} b^3 + (N-n)^2 n^{-3/2} b^{-1/2}\right),$$

where $c^*(x) =$

$$\left\{-2 \frac{d_s''(x) d_{P-s}(x) + d_s'(x)^2}{d_s(x)^2} + d_{P-s}''(x)\right\} d_s(x)^{-1} d_{P-s}(x).$$

Observations:

(i) The conditional relative bias $E\left(\hat{T}_{np} - T\right)/T = O_p\left(b^2 + n^{-1/2} b^{1/2}\right) \to 0$, if $b \to 0$.

(ii) The expression for the variance has a bandwidth independent term, which dominates. Thus the variance is $O_p\left((N-n)^2/n\right)$, for $b \to 0, nb \to \infty$. In this respect it is typical of variances of estimators of total in general.

(iii) Suppose $b = Cn^{\varepsilon}$. Then $\varepsilon < -1/4$ implies that the bias relative to the standard deviation $E\left(\hat{T}_{np} - T \mid X_P\right)/\operatorname{var}^{1/2}\left(\hat{T}_{np} - T \mid X_P\right) \to 0$ in probability. This result is desirable from the point of view of constructing confidence intervals based on estimates of variance, and suggests that for given sample size *n*, the bandwidth should be narrower than would optimally be the case for standard non-parametric regression described in the previous section.

(iv) The order of the bias is minimal when $b = Cn^{-1/3}$ (However, we need to be a bit cautious in relying on the order properties. For example, let $N - n = 358$, $n = 52$; then the "order of variance" is $358/\sqrt{52} = 49.6$, and the "order of bias" is $358/52^{2/3} = 25.7$, *not* seriously lower. This suggests we need to look at explicit expressions.)

(v) Low sample density can be a problem for bias, especially where $m(x)$ is steep since the integrand $\beta(x) d_s^{-1}(x) = m''(x) + 2m'(x) d_s'(x)/d_s(x)$.

(vi) In the variance, low sample density can also be a problem, paricularly where $\sigma^2(x)$ is large

(vii) Suppose $N >> n$, so that $d_{P-s}(x) \approx d_P(x)$ and suppose the sample is selected so that

$$d_s(x) = \sigma(x) d_P(x) / \int \sigma(x) d_P(x) dx. \qquad (1)$$

Then the lead term of the variance becomes

$$(N^2/n)\left\{\int \sigma(x) d_P(x) dx\right\}^2$$

This establishes an important connection to estimation of totals based on parametric regression models. To see this we remind ourselves of a result of R. Royall.

Theorem (Royall 1992). Suppose

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon, \qquad (2)$$

$\operatorname{var}(\varepsilon) = \mathbf{V}$, with $\mathbf{V} = diag\left\{\sigma^2(x_1), \ldots, \sigma^2(x_N)\right\}$, and both $\mathbf{V}\mathbf{1}_N$ and $\mathbf{V}^{1/2}\mathbf{1}_N \in \mathcal{M}(\mathbf{X})$. That is, both the vector of variances and of standard deviations are in the column space of $\mathbf{X}$ – we can refer to such models as ***columnar models***. Then the best linear unbiased estimator is of the form $\hat{T}_{BLU} = 1'\mathbf{X}\hat{\beta}_s$, with $\hat{\beta}_s = \left(\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s\right)\mathbf{X}_s'\mathbf{V}_s^{-1}Y_s$, and satisfies

$$\operatorname{var}_M\left[\hat{T}_{BLU} - T\right] \geq n^{-1}\left(\mathbf{1}_N'\mathbf{V}^{1/2}\mathbf{1}_N\right)^2 - \mathbf{1}_N'\mathbf{V}\mathbf{1}_N$$
$$= N^2 n^{-1}\left(N^{-1}\textstyle\sum_P \sigma(x_i)\right)^2 - \sum_P \sigma^2(x_i)$$
$$\approx (N^2/n)\left\{\int \sigma(x) d_P(x) dx\right\}^2.$$

This is the same expression as for $\hat{T}_{np}$ noted in (vii) above.

The bound is achieved if and only if we have *weighted balance* with respect to the standard deviations, which is to say

$$\frac{1}{n}\mathbf{1}_s'\mathbf{V}_s^{-1/2}\mathbf{X}_s = \frac{\mathbf{1}_N'\mathbf{X}}{\mathbf{1}_N'\mathbf{V}^{1/2}\mathbf{1}_N}.$$

Furthermore, $\hat{T}_{BLU}$ remains unbiased if the truth is that $\mathbf{y} = \mathbf{Z}\gamma + \varepsilon$, so long as $\frac{1}{n}\mathbf{1}_s'\mathbf{V}_s^{-1/2}\mathbf{Z}_s = \frac{\mathbf{1}_N'\mathbf{Z}}{\mathbf{1}_N'\mathbf{V}^{1/2}\mathbf{1}_N}$.

This just says that for each column vector *z* in **Z**,

$$\frac{\sum_s \sigma(x_i)^{-1} z_i}{n} = \frac{\sum_P z_l}{\sum_P \sigma(x_l)}.$$

We refer specifically to $\sigma(x)$- weighted balance or just "$\sigma(x)$-balance." In the present context, with *x* taken as a scalar, **X** is an elementary polynomial model in *x*, and **Z** might be a complex polynomial model, more closely approximating the underlying truth of the relation between *Y* and *x*.

A common way to get $\sigma(x)$-balance is by a two step process: (a) do $pp\sigma(x)$ sampling many times from the population to get several $pp\sigma(x)$ samples; (b) choose the sample among these most closely matching the balance conditions. This works because in design-expectation, such sampling gives a $\sigma(x)$-balanced sample. (See Valliant, Dorfman, and Royall (2000) for a detailed account.) Now the condition (1) above can be shown to be a smooth, stochastic approximation to conditions for samples arising out of $pp\sigma(x)$ sampling, including samples having $\sigma(x)$-balance. Thus we have the following comparison:

Estimator based on *parametric* (columnar) model gives exact unbiasedness and minimal variance, for a particular sample—a *weighted balanced sample*.

Estimator based on *non-parametric* model gives approximate unbiasedness (with degree of unbiasedness dependent on bandwidth), with no condition on balance, and gives approximate minimal variance, under approximate weighted balance.

Can we combine best of both? We might use a parametric model meeting conditions (*columnar model*), but then adjust estimate using non-parametric regression to protect against bias if the sample does not meet the weighted balance condition.

The non-parametric regression calibration estimator
Basic idea: Suppose the parametric working model (2) is used to construct $\hat{T}_{BLU}$, and the truth is $Y = m(x) + \sigma(x)e$.

Then the bias of $\hat{T}_{BLU}$ is $E(\hat{T}_{BLU} - T \mid \mathbf{X}_P) = \sum_{P-s}\delta(x_j)$,

where the deviations $\delta(x_j) = x_j' E(\hat{\beta}) - m(x_j)$.

Sample residuals $r_i = y_i - x_i'\hat{\beta}$ are unbiased for $-\delta(x_i)$ and we can estimate the non-sample deviations non-parametrically by $-\hat{\delta}(x_j) = \sum_i w_{ij} r_i$, to yield the estimator $\hat{T}_{cal} = \hat{T}_{BLU} + \sum_{P-s}\hat{\delta}(x_j)$.

This is the non-parametric regression calibration estimator first described in (Chambers, Dorfman, Wehrly 1993).

Bias of $\hat{T}_{cal}$
Suppose the working model is a (*p*th order) *polynomial model* in a single variable, then we have the following asymptotic expression for the bias of $\hat{T}_{cal}$:

$$E(\hat{T}_{cal} - T \mid X_P) =$$
$$c_1 b^2 (N-n) \int \beta^*(x) d_s(x)^{-1} d_{P-s}(x) dx$$
$$+ O_p\left((N-n)b^3 + (N-n)n^{-1/2}b^{1/2}\right), \text{ where}$$

$$\beta^*(x) = \beta(x) - \sum_{l=1}^{p} \beta_l(x) \text{ and}$$

$$\beta_l(x) = l(l-1)x^{l-2} d_s(x) + 2l x^{l-1} d_s'(x),\ l = 1,\ldots,p$$

The lead term reduces to zero if $m(x)$ is actually a *p*th order polynomial. This suggests that if the working model is nearly correct, then wider bandwidths can be used with $\hat{T}_{cal}$ than with $\hat{T}_{np}$, with a possible reduction in variance.

Simulation Study
We investigate the behavior of several estimators of total on (a mildly trimmed version of) the Beef Population (Chambers and Dunstan 1986). We have $N = 410$. The auxiliary variable x = herd size, and the variable of interest y = beef income. The population is quite messy, but some detailed examination under transformations suggests that a good model for the population is given by

$$E(Y \mid x) = \exp(1.74 + 5.33 \log\log x),$$
$$\mathrm{var}(Y \mid x) = 1.4 \exp(1.51 + 5.33 \log\log x).$$

Sampling was carried out $ppx^{3/4}$, with the sample size taken as $n = 52$. In addition to non-parametric regression estimation, the following two parametric models were used for inference:

$$Y_i = \alpha + \beta x_i^{3/4} + \gamma x_i^{3/2} + x_i^{3/4}\varepsilon_i \quad (3)$$

$$Y_i = \alpha + \beta x_i + x_i^{3/4}\varepsilon_i, \quad (4)$$

as well as non-parametric calibration estimators based on these. Note that (3) is a columnar model, (4) is not. Also, for comparison, Generalized Regression Estimators (GREG) were calculated using linear and quadratic models that, like (3) and (4), assumed $\mathrm{var}(Y_i \mid x_i) = x_i^{3/2}\sigma^2$. The GREG is of the form

$\hat{T}_{LU,\pi_i^{-1}\sigma^{-2}} + \sum_s \pi_i^{-1} r_i$,

with $\hat{T}_{LU,\pi^{-1}\sigma^{-2}} = \mathbf{1}'\mathbf{X}\hat{\beta}_{s,\pi^{-1}\sigma^{-2}}$, where $\hat{\beta}_{\pi^{-1}\sigma^{-2}} = \left(\mathbf{X}_s'\tilde{\mathbf{V}}_s^{-1}\mathbf{X}_s\right)^{-1}\mathbf{X}_s'\tilde{\mathbf{V}}_s^{-1}Y_s$, with $\tilde{\mathbf{V}} = \mathbf{V}diag\{\pi_1,\ldots,\pi_N\}$ (Sarndal, Swenson, Wretman 1992, etns. 12.2.1 and 12.2.2).

A thousand samples were generated. Empirical Bias, standard error, and root mean square error for the several estimators.. Non-parametric estimation and calibration was done on the log(log(x)) scale. Using a constant bandwidth for all estimates of the components $m(x_j)$ makes more sense on this scale. (However, we do not use a strictly constant bandwidth; in cases where a target x_j has less than a set minimal number of sample x's within the bandwidth interval, the interval is enlarged to guarantee the minimum. This idea goes back to Cleveland (1979).) We note the degree of flexibility the calibration estimator affords: we can construct the parametric component on one scale, and make the non-parametric adjustment on another.

The**Table** attached to the end of this paper gives results of the simulation. Some **observations**:
[The italicized Roman numerals in the Table indicate the first row of results (possibly the only row) corresponding to the particular observation.]

(i) The *BLU* estimator based on the *columnar* model does well, with low bias and variance [row 1].
(ii) The *calibration* estimator based on the *columnar* model does slightly better than the *BLUE* at high bandwidth [last several rows of table]. In general this estimator appears robust to changes in bandwidth.
(iii) The *non-parametric* kernel estimator is weak for this population, and sensitive to bandwidth selection – large b yield extreme biases
(iv) (a)The variance of the *np* estimator is U-shaped on b; (b, c) for the *calibration* estimators, variance steadily decreases with b
(v) The *BLU* estimator based on the *linear* model with "correct" weights (which is not recommended, not being columnar) has low variance, but large bias, yielding large *rmse* [row2]
(vi) The *calibration* estimator based on the *linear* model is better than the corresponding *BLU* over a wide range of bandwidths, and can be considerably better [cf. b = 0.18, ...,0.48]. At large b, however, we get large bias, leading to very bad *rmse*.
(vii) Biases of (a) *np* and (b, c) *calibration* estimators are opposite in sign
(viii) the *GREG* based on the *quadratic* model and "correct" variances does well, but the *GREG* based on the linear model with "correct" weights has a huge bias, and consequent high *rmse* . (The explanation is as follows: the population is concave. Fitting a straight line with severe downweighting on the right creates many large negative residuals for large x. If one were to sum the residuals using the same weights as those used in the regression, the result would be zero. But the adjustment is done using only the inverse π - weights, which, relatively speaking, gives large weight to the large x residuals. Hence an extreme negative adjustment.)
(ix) Straight BLU estimation using the same models and effective weights as were used for the GREG is included for comparison (i.e. this is GREG minus the residual adjustment term.) We note that the adjustment improves estimation in the case of the near columnar model, but makes it severely worse in the case of the linear model.

Conclusion: A Sampling "Meta-Strategy"
The following seems to be a reasonable overall sampling strategy:

(i) The most straightforward approach is to use a *BLU* estimator based on an appropriate columnar model, having selected a corresponding weighted balanced sample.
(ii) Failing a weighted balanced sample (and possibly even if one has it) use a *non-parametric* calibration estimator based on the appropriate columnar model, using moderate to large bandwidth.
(iii) In the rare case where modelling is hopeless, use straight non-parametric regression estimator.

Related Work
The following is intended to give an idea of what work has been done related to the application of non-parametric regression to sampling, but is not intended to be comprehensive.

Recent books on non-parametric regression are Wand, and Jones (1995) and Fan and Gibjels (1996).

Kuo (1988) applied non-parametric regression to sample data to estimate the finite population distribution function. Dorfman and Hall (1993) and Kuk (1993) developed further methods and theory for this.

Dorfman (1992; 1994) applied non-parametric regression to sample data to estimate the finite population total. The calibration paper of Chambers, Dorfman, and Wehrly (1993) was meant to comprehend estimation of any finite population "parameter". Chambers (1996) describes using nonparametric regression calibration successfully on multi-variate data, in combination with ridge regression methods. Breidt and Opsomer (2000a, 2000b) focus on estimating totals using local linear regression and a twicing procedure which parallels the GREG.

Non-parametric regression for purposes of data exploration and analysis has been carried out by Smith and Njenga (1992), Korn and Graubard (1998, 1999), Scott and Whitaker (1996), Bellhouse and Stafford (2000), Chambers, Dorfman, and Sverchkov (2000).

Further Research

The most obvious omisssion from the present study is the use of local linear regression (Cleveland 1979; Fan, 1992; Ruppert and Wand, 1994). The expression for the asymptotic bias of this version of a non-parametric regression estimator of total will not include division by the sample density, and so the bias of a local linear regression based estimator should be less sensitive to sparse x regions in the sample data (J. Opsomer, personal communciation). It can be shown that the local linear estimator of total shares the property of the calibration estimator of having zero bias, if the model used for local linear regression is the correct one. We would expect it to perform in intermediate fashion between the kernel based estimator, and the calibration estimator, if the model used is at all close to the truth. The calibration estimator itself could use local regression to make the non-parametric adjustment.

Except in the situation of non-ignorable sampling, where not all the information about y is contained in the auxiliary variable, the weights used in non-parametric regression effectively *supercede* the inclusion probability weights customarily associated with survey sampling. As a rule, incorporation of both non-parametric *and* sampling into the process seems tautological, and likely to lead to inefficiencies. However, Breidt and Opsomer (2000a) report a *loss* of efficiency using pure model-based nonparametric regression, relative to a twiced design-based local regression estimator. Probably additional comparative study is in order. One point to note is that different non-parametric regression estimators are likely to be their best under different bandwidths. In particular, estimators using twicing (as in the calibration version, or the GREG version of Opsomer and Breidt) tend to do better at larger bandwidths than their un-twiced kin (cf. Chambers, Dorfman, and Sverchkov 2000).

A "pure twicing" non-parametric estimator, using non-parametric regression weights both for the original fit, and for the residual adjustment, would be worthy of investigation.

Dorfman (1994) suggests a variance estiamtion procedure for the nonparametric estimate of total, but further work is in order.

The calibration estimator seemed fairly immune to variations in bandwidth in the present simulation study. Chambers, Dorfman, and Wehrly (1993) suggest a method for choosing bandwidth in the calibration case. Nonetheless, probably the most pressing need is for some automatic way of selecting bandwidth in the case of non-parametric regression for estimating totals.

Any opinions expressed in this paper are those of the author and do not constitute policy of the Bureau of Labor Statistics.

REFERENCES

Bellhouse, D.R. and Stafford, J.E. (2000), Local Polynomial Regression in Complex Surveys, in *Analysis of Survey Data*, edited by C. Skinner and R.L. Chambers, Chichester: John Wiley (to appear).

Breidt, F.J. and Opsomer, J.D. (2000a), Local Polynomial Regression Estimators in Survey Sampling, *Annals of Statistics*, to appear

Breidt, F.J. and Opsomer, J.D. (2000b), Local Polynomial Regression Estimation for Complex Surveys, This *Proceedings of the Section on Survey Research Methods*, American Statistical Association

Chambers, R. L (1996), Robust Case-Weighting for Multipurpose Establishment Surveys, *Journal of Official Statistics* 12, 3-32

Chambers, R. L., Dorfman, A. H., and Hall, P. (1992), Properties of Estimators of the Finite Distribution Function, *Biometrika* 79, 577-582.

Chambers, R. L., Dorfman, A. H., and Sverchkov, M (2000), Nonparametric regression with Complex Survey Data, in *Analysis of Survey Data*, edited by C. Skinner and R.L. Chambers, Chichester: John Wiley (to appear).

Chambers, R. L., Dorfman, A. H. &Wehrly, T. E. (1993), Bias Robust Estimation in Finite Populations using Nonparametric Calibration, *J. Am Statist. Assoc.*88, 268-277.

Cleveland, W.S. (1979), Robust Locally Weighted Regression and Smoothing Scatterplots, *Journal of the American Statistical Association* 74, 268-277

Cochran, W. G. (1977), *Sampling Techniques*(3rd ed.), Chichester: John Wiley.

Dorfman, A. H. (1992), Nonparametric Regression for Estimating Totals in Finite Populations, *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 622-625.

Dorfman, A. H. (1994), Open Questions in the Application of Smoothing Methods to Finite Population Inference, *Computationally Intensive Statistical Methods, Proceedings of the 26th Symposium on the Interface*, 201-204

Dorfman, A. H. and Hall, P. (1992) Estimators of the finite population distribution function using nonparametric regression, *Annals of Statistics*, 21, 1452-1475.

Fan (1992), Design-adaptive Nonparmatric Regression, *Journal of the American Statistical Association*, 87, 998-1004.

Fan and Gibjels (1996), *Local Polynomial Modelling and Its Applications*, London: Chapman & Hall.

Hardle, W. (1991), *Smoothing Techniques*, London: Springer-Verlag.

Hardle, W., Hall, P. and Marron, J. S. (1992), Regression Smoothing Parameters that are not far from their Minimum, *J. Am Statist. Assoc.* 87, 227-233.

Korn, E.L. and Graubard, B.I. (1998), Scatterplots with Survey Data, *The American Statistician*, 52, 58-69

Korn, E.L. and Graubard, B.I. (1999), *Analysis of Health Surveys*, New York: Wiley

Kuk, A. (1993) A Kernel Method for Estimating Finite Population Distribution Functions using Auxiliary Information, *Biometrika*, 80, 385-392.

Kuo, L. (1988), Classical and Prediction Approaches to Estimating Distribution Functions from Survey Data, *Proceedings of the Section on Survey Research Methods*, Amercian Statistical Association, 280-285.

Nadaraya, E. A. (1964), On estimating regression, *Theory of Prob. and Applic.* 9, 141-142.

Royall, R. M., and Cumberland, W. G. (1981), An empirical study of the ratio estimator and estimators of its variance, *J. Am Statist. Assoc.* 76, 66-77.

Royall, R. M. and Herson, J. (1973) Robust Estimation in Finite Populations I, *J. Am Statist. Assoc.* 68, 880-893.

Ruppert, D. and Wand, M. P. (1993) Multivariate Locally Weighted Least Squares Regression, Preprint.

Sarndal, C-E, Swenson, B., and Wretman, J. (1992), *Model Assisted Survey Sampling*, New York: Springer-Verlag

Scott, D.W. and Whittaker, G. (1996), Multivariate Applications of the ASH in Regression, *Communications in Statistics – Theory and Methods*, 25, 2521-2530

Smith, T. M. F. and Njenga, E. (1992), Robust Model-based Methods for Analytic Surveys, *Survey Methodology* 18, 187-208.

Valliant, R., Dorfman, A.H., and Royall, R.M. (2000), *Finite Population Sampling and Inference, A Prediction Approach*, New York: John Wiley

Wand, M.P. and Jones, M.C. (1995), *Kernel Smoothing*, London: Chapman & Hall.

Watson, G. S. (1964), Smooth regression analysis, *Sankhya,* Ser. A, 26, 359-372.

that	*b*		*Bias*		*std deviation*		*rmse*
$x^{.75} + x^{1.5}$ $(x^{1.5})$			-183027		4748211	(*i*)	4749364
x $(x^{1.5})$			-4117643		4569491	(*v*)	6149337
$x + x^2$ $(x^{2.25})$ [GREG]			252562		4929784	(*viii*)	4933788
x $(x^{2.25})$[GREG]			-7436504		4498170		8689930
$x + x^2$ $(x^{2.25})$			751825		5275263	(*ix*)	5325957
x $(x^{2.25})$			-63837		5307424		5305154
non-parametric	0.03	(*vii-a*)	734920	(*iv-a*)	5776354	(*iii*)	5820052
"	0.06		861868		5549170		5612959
"	0.09		1282969		5378162		5526456
"	0.12		1942576		5241891		5587803
"	0.18		3883900		5155032		6452322
"	0.24		6490121		5227771		8332102
"	0.36		13656985		5861154		14860417
np cal'n x $(x^{1.5})$	0.03	(*vii-b*)	358482	(*iv-b*)	5736777	(*vi*)	5745103
"	0.06		129230		5498354		5497123
"	0.09		-66546		5303162		5300928
"	0.12		-249492		5126170		5129677
"	0.18		-498893		4935485		4958180
"	0.24		-664681		4854144		4897035
"	0.36		-759459		4884844		4941115
"	0.42		-854932		4918558		4989883
"	0.48		-1122390		4924266		5048158
"	0.54		-1607828		4898834		5153609
"	0.60		-2343974		4841147		5376568
"	0.66		-3344623		4750129		5807552
"	0.72		-4550190		4650414		6504533
"	0.80		-6340820		4547456		7801582
"	0.88		-8169475		4498480		9325042
np cal'n $x^{.75} + x^{1.5}$ $(x^{1.5})$	0.03	(*vii-c*)	-74928	(*iv-c*)	5685966		5683616
"	0.06		-181656		5472476		5472755
"	0.09		-220582		5296469		5298414
"	0.12		-238532		5142265		5145225
"	0.18		-236860		4975430		4978580
"	0.24		-257958		4886428		4890792
"	0.36		-230507		4810480		4813596
"	0.42		-191383		4780368		4781809
"	0.48		-152914		4746429	(*ii*)	4746519
"	0.54		-100115		4717857		4716561
"	0.60		-50589		4695083		4693008
"	0.66		-21832		4670770		4668485
"	0.72		-15570		4647513		4645215
"	0.80		-35979		4628345		4626171
"	0.88		-78734		4624092		4622450