ESTIMATING COVARIANCE MATRICES USING ESTIMATING FUNCTIONS IN NONPARAMETRIC AND SEMIPARAMETRIC REGRESSION

R. J. Carroll and S. J. Iturria
Department of Statistics
Texas A&M University
College Station, TX 77843–3143

R. G. Gutierrez
Department of Statistical Science
Southern Methodist University
Dallas, TX 75275–0332

ABSTRACT

We use ideas from estimating function theory to derive new, simply computed consistent covariance matrix estimates in nonparametric regression and in a class of semiparametric problems. Unlike other estimates in the literature, ours do not require auxiliary or additional nonparametric regressions.

Key Words: Estimating Equations; Kernel regression; Nonparametric regression; Plug–in Semiparametrics; Smoothing.

1 Introduction

Estimating functions form a powerful methodology for parametric analyses. Their use in nonparametric and semiparametric problems is less developed. Here we use estimating equations to derive standard error estimates in these contexts.

The first problem is ordinary nonparametric local polynomial regression. It has not been generally appreciated that these estimates are in fact solutions to estimating equations, a point which was first noticed by Carroll, Ruppert & Welsh (1996). We show how their looking at this problem via estimating equations leads to a new sandwich–type covariance matrix estimate.

The second problem is semiparametric regression, of a type we call "plug–in" (defined later in the paper). In semiparametric problems, estimation of a parameter is often of most interest. One way to obtain a covariance matrix for the estimated parameter involves a two–step process: (a) derive an asymptotic expression, usually involving a suite of densities and additional nonparametric regressions; and (b) estimate each term in turn. We show how Gutierrez & Carroll (1996) use estimating equations in a one–step process, leading to consistent covariance matrix estimates under minimal assumptions, and without the need for additional nonparametric regressions.

2 Ordinary Nonparametric Regression

Ordinary nonparametric regression is ideally suited to development of estimating functions. For example, consider local polynomial regression of order p, in regression Y on X. Based on a sample of size n, local polynomial estimates of $\theta(x_0) = E(Y|X = x_0)$ are formed by minimizing

$$\sum_{i=1}^{n} w(X_i, x_0)\{Y_i - \sum_{j=0}^{p} \beta_j (X_i - x_0)^j\}^2,$$

where $w(x, x_0)$ is a weight function, e.g., kernel weights or loess weights. The estimated function is $\widehat{\theta}(x_0) = \widehat{\beta}_0$. Defining $G_p(x) = (1, x, x^2, ..., x^p)^t$ and differentiating, local polynomial regression solves

$$0 = \sum_{i=1}^{n} w(X_i, x_0)\{Y_i - \sum_{j=0}^{p} \beta_j (X_i - x_0)^j\} G_p(X_i - x_0). \tag{2.1}$$

Carroll, et al. (1996) noted that (2.1) is an estimating equation, and they use this fact to develop a general theory of nonparametric regression which includes both much of the current literature as well as many new ideas. The estimating function is not unbiased in the usual sense, because the true mean $\theta(x_0)$ has been replaced by its local polynomial approximation $\sum_{j=0}^{p} \beta_j (X - x_0)^j$. However, asymptotically, as the weights become more concentrated at x_0, the estimating function becomes unbiased.

Routine application of Godambe's estimating function theory suggests that $\widehat{\theta}(x_0) - \theta(x_0)$ is asymptotically normally distributed with mean zero and variance

$$(1, 0, ..., 0) A_n^{-1}(x_0) B_n(x_0) A_n^{-1}(x_0) (1, 0, ..., 0)^t, \tag{2.2}$$

where

$$A_n(x_0) = E\{\sum_{i=1}^{n} w(X_i, x_0) G_p(X_i - x_0) G_p^t(X_i - x_0)\} \tag{2.3}$$

$$B_n(x_0) = E[\sum_{i=1}^{n} w^2(X_i, x_0) G_p(X_i - x_0) G_p^t(X_i - x_0)\{Y_i - \sum_{j=0}^{p} \beta_j (X_i - x_0)^j\}^2]. \tag{2.4}$$

As in the usual sandwich methodology, consistent estimation of $A_n(x_0)$ and $B_n(x_0)$ is accomplished by replacing the expectations in (2.3)–(2.4) by sums over the data.

This routine use of well–known parametric theory in nonparametric regression problems appears to be new, and Carroll, et al. (1996) develop this idea into contexts not previously considered in the literature. Ordinarily, researchers either (i) assume homogeneity of variance and replace $\{Y_i - \sum_{j=0}^{p} \beta_j (X_i - x_0)^j\}^2$ in (2.4) by the constant global variance; or (ii) work out all the details of the asymptotics and then estimate all of the terms. This use of parametric estimating equation theory provides a powerful way of forming estimated variances without having to go through the second alternative.

Here we sketch the argument of Carroll, et al. (1996) in this special case, showing that at least for kernels the estimating equation–based standard errors are asymptotically correct. The only caveat concerns bias. Since (2.1) is not an unbiased estimating function, we cannot claim that $\widehat{\theta}(x_0)$ is consistent for $\theta(x_0)$ without accounting for bias. In fact, estimating this bias even in this simple context has been and remains a problem of considerable interest in the kernel literature (Ruppert, 1997). It is not clear whether, or how, one can use estimating equation methodology to estimate this bias.

Here is a sketch of the argument of Carroll, et al. (1996) showing the consistency of (2.2) for local linear regression ($p = 1$). Let $\sigma^2(x_0) = \mathrm{Var}(Y|X = x_0)$, which is assumed to be smooth. For kernel weights with bandwidth h, $w(X_i, x_0) = K_h(X - x_0) = h^{-1}K\{(X - x_0)/h\}$, and it is well known (Ruppert and Wand, 1994) that the asymptotic variance of local linear regression is

$$\{nhf_X(x_0)\}^{-1} k_2 \sigma^2(x_0), \tag{2.5}$$

where $f_X(\cdot)$ is the density of X, $k_1 = \int x^2 K(x)dx$, $k_2 = \int K^2(x)dx$, and $k_3 = \int x^2 K^2(x)dx$.

It is easily seen that (2.2) is unchanged if we replace $(X - x_0)$ by $(X - x_0)/h$ and adjust the definition of β_1 accordingly, in which case it can be shown that

$$n^{-1}A_n(x_0) \xrightarrow{p} f_X(x_0) \begin{pmatrix} 1 & 0 \\ 0 & k_1 \end{pmatrix};$$

$$(h/n)B_n(x_0) \xrightarrow{p} f_X(x_0)\sigma^2(x_0) \begin{pmatrix} k_2 & 0 \\ 0 & k_3 \end{pmatrix}.$$

Plugging these asymptotic expressions into (2.2), we obtain (2.5) as desired. For polynomials of order $p \neq 1$, similar arguments apply.

3 Plug–in Semiparametrics

Estimating equation methodology can also be used in what we call semiparametric plug–in problems to derive easily computed consistent covariance matrix estimates for parameters. These problems are derived as follows. Suppose that an estimating equation for a parameter α depends on vector–valued data $\widetilde{Y}$ along with a scalar–valued function $\theta(X)$, where

X is a subcomponent of $\widetilde{Y}$. In this case we can write the estimating function for α as $\Psi\{\widetilde{Y}, \alpha, \theta(X)\}$. By definition, a plug–in problem works as follows: $\theta(\cdot)$ can be estimated without reference to α by a local estimating equation based on (Δ, X) and an estimating function $\chi(\cdot)$, where Δ is another component of $\widetilde{Y}$, by solving

$$0 = \sum_{i=1}^{n} w(X_i, x_0) G_p(X_i - x_0) \mathcal{X}\{\Delta_i, \sum_{j=0}^{p} \beta_j (X_i - x_0)^j\},$$

with $\widehat{\theta}(x_0) = \widehat{\beta}_0$; note the similarity with (2.1). We now "plug–in" the estimated function $\widehat{\theta}(\cdot)$, and solve the following equation to form an estimate $\widehat{\alpha}$ for the parameter α:

$$0 = \sum_{i=1}^{n} \Psi\left\{\widetilde{Y}, \alpha, \widehat{\theta}(X_i)\right\}.$$

In what follows, we will ignore issues of bias, which are considered in detail by Gutierrez and Carroll (1996) and by Carroll, Fan, Gijbels and Wand (1997).

Gutierrez and Carroll (1996) derive the asymptotic distribution of $\widehat{\alpha}$ in this and more general situations. The asymptotic covariance matrix depends as expected on the density of the X's as well as various further nonparametric regressions. They show that the following is a consistent estimate of the asymptotic covariance matrix of $\widehat{\alpha}$ (the argument appears after the definitions). Remember that α may be vector–valued but that $\theta(\cdot)$ is scalar. Define

$$\begin{aligned}
A_n(\alpha, \theta) &= -\sum_{i=1}^{n} \frac{\partial}{\partial \alpha^t} \Psi\left\{\widetilde{Y}_i, \alpha, \theta(X_i)\right\}; \\
B_n(x, \theta) &= -(1, 0, ..., 0) \sum_{i=1}^{n} w(X_i, x) G_p(X_i - x) G_p^t(X_i - x) \mathcal{X}_2\left\{\Delta_i, \theta(X_i)\right\}; \\
\mathcal{X}_2(\Delta, v) &= \frac{\partial}{\partial v} \mathcal{X}(\Delta, v); \\
C_n(x, \theta) &= \sum_{i=1}^{n} w(X_i, x) G_p(X_i - x) \mathcal{X}\left\{\Delta_i, \theta(x)\right\}; \\
\Lambda_n(\Delta, X, \widetilde{Y}, \alpha, \theta) &= \Psi\left\{\widetilde{Y}, \alpha, \theta(X)\right\} + \\
&\quad \sum_{i=1}^{n} w(X, X_i) \frac{\partial}{\partial \theta} \Psi\left\{\widetilde{Y}_i, \alpha, \theta(X_i)\right\} B_n^{-1}(X_i, \theta) G_p(X - X_i) \mathcal{X}\left\{\Delta, \theta(X_i)\right\}; \\
D_n(\alpha, \theta) &= \sum_{i=1}^{n} \Lambda_n(\Delta_i, X_i, \widetilde{Y}_i, \alpha, \theta) \Lambda_n^t(\Delta_i, X_i, \widetilde{Y}_i, \alpha, \theta).
\end{aligned}$$

A consistent covariance matrix estimate is

$$A_n^{-1}(\widehat{\alpha}, \widehat{\theta}) D_n(\widehat{\alpha}, \widehat{\theta}) A_n^{-t}(\widehat{\alpha}, \widehat{\theta}). \tag{3.1}$$

To justify (3.1), we provide the following sketch based on the arguments of Gutierrez & Carroll (1996). First note that by ordinary estimating equation theory, $\widehat{\theta}(x) - \theta(x) \approx$

$B_n^{-1}(x,\theta)C_n(x,\theta)$. Then with $\Psi_i = \Psi\{\widetilde{Y}_i, \alpha, \theta(X_i)\}$ and $\Psi_\theta = (\partial/\partial\theta)\Psi$,

$$\begin{aligned} \widehat{\alpha} - \alpha &\approx A_n^{-1}(\alpha,\theta)\sum_{i=1}^{n}\left[\Psi_i + \Psi_{\theta i}\left\{\widehat{\theta}(X_i) - \theta(X_i)\right\}\right] \\ &\approx A_n^{-1}(\alpha,\theta)\sum_{i=1}^{n}\left\{\Psi_i + \Psi_{\theta i}B_n^{-1}(X_i,\theta)C_n(X_i,\theta)\right\} \\ &= A_n^{-1}(\alpha,\theta)\sum_{i=1}^{n}[\Psi_i + \Psi_{\theta i}B_n^{-1}(X_i,\theta)\sum_{\ell=1}^{n} w(X_\ell, X_i)G_p(X_\ell - X_i)\mathcal{X}\{\Delta_\ell, \theta(X_i)\}]. \end{aligned}$$

Interchanging indices of summation, $\widehat{\alpha} - \alpha \approx A_n^{-1}(\alpha,\theta)\sum_{i=1}^{n}\Lambda_n(\Delta_i, X_i, \widetilde{Y}_i, \alpha, \theta)$, justifying (3.1).

While informal, all of these calculations are easily justified in kernel regression with bandwidth h. Generally though, in order that $n^{1/2}(\widehat{\alpha} - \alpha) = O_p(1)$, it is required that $nh^2 \to \infty$ and $nh^4 \to 0$. Certain problems weaken $nh^4 \to 0$ to $nh^6 \to 0$, see Gutierrez and Carroll (1996).

Implementation of (3.1) is easy, because all the terms involved are building blocks in the estimation process. We have found in other contexts (Simpson, et al., 1997) that inference is improved if it is based on percentiles of the t–distribution with $n - 2(p+q+1)$ degrees of freedom, and if (3.1) is multiplied by $n/\{n - 2(p+q+1)\}$, where q is the dimension of α and p is the size of the local polynomial.

Acknowledgment

This research was supported by a grant from the National Cancer Institute (CA–57030). Carroll's research was partially completed while visiting the Institut für Statistik und Ökonometrie, Sonderforschungsbereich 373, Humboldt Universität zu Berlin, with partial support from a senior Alexander von Humboldt Foundation research award.

References

Carroll, R. J., Fan, J., Gijbels, I. and Wand, M. P. (1997). Generalized partially linear single–index models. *J. Am. Statist. Assoc.*, to appear.

Carroll, R. J., Ruppert, D. & Welsh, A. (1996). Nonparametric calibration: nonparametric regression based on estimating equations. Preprint.

Gutierrez, R. G. and Carroll, R. J. (1996). Plug–in semiparametric estimating equations. Preprint.

Ruppert, D. (1997). Empirical Bias Bandwidth Selection. *J. Am. Statist. Assoc.*, to appear.

Ruppert, D. and Wand, M. P. (1994). Multivariate locally weighted least squares regression. *Ann. Statist.*, 22, 1346–1370.

Simpson, D. G., Carroll, R. J., Guth, D. & Zhou, H. (1997). Interval censoring and marginal analysis in ordinal regression. *J. Agric., Biol. Env. Statist.*, to appear.