THE ASYMPTOTIC VARIANCE OF SEMIPARAMETRIC ESTIMATORS

Whitney K. Newey
Princeton University and BellCore

Econometric Research Program
Research Memorandum No. 346

August 1989

Helpful comments were provided by L. Hansen, J. Hausman, R. Klein, J. Powell, and T. Stoker. Financial support was provided by NSF and the Sloan Foundation.

Econometric Research Program
Princeton University
207 Dickinson Hall
Princeton, NJ 08544, USA

Abstract

Knowledge of the asymptotic variance of an estimator is important for large sample inference, efficiency, and as a guide to the specification of regularity conditions. The purpose of this paper is the presentation of general formulae for the asymptotic variance of semiparametric estimators. A particularly important feature of these formulae is a way of accounting for the presence of nonparametric estimates of nuisance functions. The general form of an adjustment factor for nonparametric estimates is derived and analyzed.

The paper illustrates how these results are useful as a guide to asymptotic inference, efficiency, and the specification of regularity conditions. The asymptotic efficiency of several types of estimators for a heteroskedastic partially linear model is evaluated. General regularity conditions for asymptotic normality are formulated and applied to nonparametric consumer surplus and policy analysis examples.

Keywords: semiparametric estimation, asymptotic variance, density estimation, nonparametric regression, kernel estimation.

1. Introduction

Knowledge of the asymptotic variance of an estimator is important for large sample inference, efficiency, and as a guide to the specification of regularity conditions. There are some well known formulas for the limiting distribution of parametric estimators. A classical result dating to the work of R.A. Fisher is the inverse information form of the asymptotic variance of the maximum likelihood estimator. A more general formula for the class of m-estimators was given by Huber (1967).

Estimators for semiparametric models have been of increasing interest in statistics and econometrics, and it would be useful to have analogous formulas for such estimators. A particularly important feature of such formulas would be accounting for the presence of nonparametric estimates of abstract (i.e. infinite dimensional) parameters. To date, there appears to be no general result. There are important results for some types of models and estimators of the abstract parameters, including those of Von Mises (1947) for estimators that are functionals of an empirical cumulative distribution function. Formulae for specific estimators have been obtained by Bickel (1982), Schick (1986), Powell, Stock, and Stoker (1989), Robinson (1988), Ichimura (1987), and others. Bickel, Klaassen, Ritov, and Wellner (1989) and Severini and Wong (1987) have obtained some model independent results when the abstract parameter is estimated by approximate or exact nonparametric maximum likelihood. Also, Andrews (1989a,b) gives some general regularity conditions for the case where estimation of the abstract parameters does not affect the asymptotic variance of parameters of interest. However, each of these results is specific to an estimation method and/or type of model. Furthermore, the asymptotic variance calculation for many of the specific models apparently involves tedious details of specific nonparametric estimation methods, such as kernel nonparametric regression. It would be useful to have general formulae that account for the presence of nonparametric estimates but did not require becoming embroiled in such details. Such results could be used as they are for parametric estimators, as a guide to asymptotic inference, efficiency, and the specification of regularity conditions.

The purpose of this paper is the presentation of general formulae for the asymptotic variance of semiparametric estimators that account for the presence of nonparametric estimates. These formulae are consequences of a simple result concerning estimators of a pathwise differentiable functional, where "functional" refers to a mapping from the distribution of the data to real vectors. As formulated by Koshevnik and Levit (1976), Pfanzagl (1982), Bickel, Klaassen, Ritov, and Wellner (1989), and Van der Vaart (1988), such functionals are those where the derivative of the functional along any finite dimensional, sufficiently regular, family of distributions can be represented as the expected outer product of the score for the family and the *derivative,* a random vector that is invariant with respect to the parametric subfamily. The simple result referred to above, which was pointed out in Newey (1989a), is that *when the distribution of the data is unrestricted*, the asymptotic variance of any estimator which is asymptotically equivalent to a sample average and is sufficiently regular is equal to the variance of the derivative. This result is analogous to those of Von Mises (1947) and Boos and Serfling (1980) for the derivative of functionals of an unrestricted distribution function, but applies to a much wider class of functionals.

The key step in applying this result to general semiparametric estimators is a conceptual one; the derivative must be calculated for the functional that is the limit of the estimator for an unrestricted family of distributions, not necessarily satisfying the assumptions of the semiparametric model. That is, the derivative must be calculated for the functional that specifies how the estimator behaves under general misspecification. It then follows from the previously mentioned result that the asymptotic variance of the estimator can be calculated from the pathwise derivative of the functional that allows for misspecification. This calculation is a nonparametric analog of Huber's (1967) formula for the limiting distribution of parametric m-estimators, which also allows for general misspecification. Of course, as in the parametric case, the formula may simplify when misspecification is not present.

Section 2 of the paper discusses pathwise differentiability and presents the simple result alluded to above. Also, the relationship of pathwise differentiability to other differentiability notions is briefly considered. Section 3 discusses calculation of the derivative, including a well understood example that is discussed for expository purposes. Section 4 considers estimators that depend on sample averages of nonparametric estimates. The form of a "correction term" for the presence of nonparametric estimates is derived. Specific formula are obtained for nonparametric estimates of derivatives of densities and conditional expectations. These formula give new results as well as reproduce recent results from the literature. A general formula for the asymptotic variance of semiparametric m-estimators is also given.

Section 5 illustrates one of the important uses of the semiparametric asymptotic variance formula, which is efficiency evaluation. In this Section the efficiency of several types of estimators for a heteroskedastic, partially linear model are compared, without detailed consideration of particular nonparametric estimation methods. Section 6 considers more primitive regularity conditions for asymptotic normality. Several types of estimators are discussed, including sample averages of nonparametric estimates and semiparametric m-estimators. A condition analogous to the Frechet differentiability condition of Boos and Serfling (1980), when combined with Andrews (1989a,b) stochastic equicontinuity condition, gives conditions for asymptotic normality that apply when the presence of nonparametric estimators affects the limiting distribution. Two examples are considered in detail; nonparametric consumer surplus estimation and the nonparametric policy analysis problem of Stock (1989).

2. Regular Estimators of Pathwise Differentiable Functionals

This section draws on the literature on semiparametric efficiency bounds; see Stein (1956), Koshevnik and Levit (1976), Pfanzagl (1982), Begun, et. al. (1983), and Bickel, et. al. (1989). Therefore, it is necessary to briefly review some ideas from this literature.

Let F index the distribution of a single observation z_i of a stationary stochastic process $z_1, z_2, \dots,$ with true value F_0. A functional of F is a mapping

$$\mu(F) : \mathcal{F} \longrightarrow \mathbb{R}^k,$$

where $\mathcal{F}$ is a family of possible distributions of z_i that includes F_0. A regular parametric subfamily of $\mathcal{F}$ is a subset of $\mathcal{F}$ that is i) parameterized by a finite dimensional parameter θ and satisfies certain regularity conditions; ii) is equal to the truth F_0 for some θ_0. Regularity conditions include absolute continuity with respect to a carrier measure for F_0 and smoothness of the square root of the density with respect to the parameters in the mean-square sense; see Appendix B. The regularity conditions may also include additional smoothness restrictions on $\mathcal{F}$ that are necessary for $\mu(F)$ to be well defined, such as existence of density functions.

For a parametric subfamily, the parameter μ will be a function $\mu(\theta)$ of θ, i.e. $\mu(\theta) = \mu(F_\theta)$ where F_θ indexes the distribution corresponding to θ. Let $S_\theta = S_\theta(z)$ denote the score for a parametric subfamily at θ_0, where z is a z_i argument that will often be suppressed for notational convenience. The score can typically be thought of as the derivative of the log-likelihood for z; it is precisely defined in Appendix B. Let $E[\circ]$ denote the expectation at $F = F_0$. A *pathwise differentiable functional* is one where there exists a $k \times 1$ *derivative* vector $d(z)$ such that $E[d'd] < \infty$ and for all regular parametric subfamilies $\mu(\theta)$ is differentiable at θ_0 and

$$(2.1) \qquad \partial\mu(\theta_0)/\partial\theta = E[dS_\theta'].$$

In general the derivative d could depend on F_0, $\mathcal{F}$, and the class of parametric subfamilies of $\mathcal{F}$, but not on the particular parametric subfamily of $\mathcal{F}$. It is not unique, but in an important special case discussed below it is unique up to a constant.

For a simple example, consider $\mu(F) = E_F[h(z)]$, where $h(z)$ is some known function of z and $E_F[\circ]$ denotes the expectation for the distribution F. Note that $\mu(\theta) = E_\theta[h(z)]$, where $E_\theta[\circ] = E_{F_\theta}[\circ]$. Under appropriate regularity conditions (e.g. see Lemma 7.2 of Ibragimov and Hasminskii, 1981), differentiation under the integral yields

$$(2.2) \qquad \partial\mu(\theta_0)/\partial\theta = \int h(z)[\partial f(z|\theta_0)/\partial\theta]dz = E[h(z)S_\theta'],$$

where $f(z|\theta)$ is the (marginal) density of an observation (with respect to some carrier measure). Thus, by comparison with equation (2.1), $d = h(z)$.

The class of estimators to be considered here are those that are asymptotically equivalent to a sample average, referred to as *asymptotically linear*. An estimator $\hat{\mu} \equiv \mu_n(z_1,\ldots,z_n)$ of $\mu_0 \equiv \mu(F_0)$ is asymptotically linear with *influence function* $\psi(z)$ (see Hampel, 1974) if

$$(2.3) \qquad \sqrt{n}[\hat{\mu} - \mu_0] = \textstyle\sum_{i=1}^{n}\psi(z_i)/\sqrt{n} + o_p(1), \quad E[\psi] = 0, \quad E[\psi'\psi] < \infty.$$

Under specific dependence restrictions on $\psi(z_i)$ (e.g. see White, 1984) it will follow from equation (2.3) that

$$\sqrt{n}[\hat{\mu} - \mu_0] \xrightarrow{d} N(0,V), \quad V = E[\psi\psi'] + \textstyle\sum_{j=1}^{\infty}E[\psi(z_i)\psi(z_{i+j})'+\psi(z_{i+j})\psi(z_i)'].$$

The purpose of this Section is to give a result that allows one to calculate ψ, and hence the asymptotic variance V of $\hat{\mu}$, from the derivative.

The results of this Section impose independent observations, so that the parameters θ of the marginal distribution of z characterize the data generating process. A local data generating process (LDGP) has for each n, z_i, $1 \le i \le n$, distributed as F_{θ_n}, where $\sqrt{n}(\theta_n-\theta_0)$ is bounded. An estimator $\hat{\mu}$ is locally regular for the parametric subfamily if $\sqrt{n}[\hat{\mu}-\mu(\theta_n)]$ has a limiting distribution that does not depend on the sequence $\{\theta_n\}$. An estimator is *regular* if it is locally regular for all parametric subfamilies and the limiting distribution does not depend on the subfamily.

Regularity has an important consequence. It rules out estimators that use more information than that contained in the statement $F \in \mathcal{F}$. Such estimators would have a local bias term that depends on the direction of approach of θ_n to θ_0 for some parametric subfamily. For an extreme example, note that the estimator $\hat{\mu} = \mu_0$ is not regular when $\mu(\theta)$ is differentiable with non zero derivative, since by a Taylor expansion, $\sqrt{n}(\hat{\mu}-\mu(\theta_n)) = -[\partial\mu(\theta_0)/\partial\theta]\sqrt{n}(\theta_n-\theta_0) + o(1)$ has a limit that depends on the direction of approach of θ_n to θ_0.

The main result of this Section is based on the following fundamental property of regular, asymptotically linear estimators. It is a simplified version of a well known result for parametric models (e.g. see Bickel et. al., 1989, Ch. 2), that appears as Theorem 2.2 of Newey (1989a).

Lemma 2.1: For an asymptotically linear estimator $\hat{\mu}$, *suppose that* i) $z_1, \ldots, z_n$ *are independent;* ii) $E[\psi\psi']$ *is nonsingular;* iii) *for all regular parametric subfamilies* $\mu(\theta)$ *is differentiable and* $E_\theta[\psi'\psi]$ *is continuous at* θ_0. *Then* $\hat{\mu}$ *is regular if and only if for all regular parametric subfamilies,*

(2.4) $$\partial\mu(\theta_0)/\partial\theta = E[\psi S_\theta'].$$

It is possible to use this result to derive a derivative formula for the influence function when $\mathcal{F}$ and the associated regular parametric subfamilies are sufficiently rich. For a matrix B let $\|B\| = [\text{trace}(B'B)]^{1/2}$, and for a constant matrix A with q rows let $\mathcal{S}$ be the mean-square closure of the union of all random vectors $AS_\theta(z)$ over all A and regular parametric subfamilies, i.e.

$$\mathcal{S} = \{\Delta \in \mathbb{R}^k : E[\Delta'\Delta] < \infty, \quad \exists\, A_j,\ S_{\theta j} \quad \text{s.t.} \quad E[\|\Delta - A_j S_{\theta j}\|^2] = o(1)\}.$$

Also, let $\tilde{\mathcal{S}} = \{\Delta : E[\Delta] = 0,\ E[\Delta'\Delta] < \infty\}$. The following result was noted in Newey (1989a), and is formally proved here.

Theorem 2.1: Suppose that $\hat{\mu}$ *is asymptotically linear and satisfies the hypotheses of Lemma 2.1 for a class of parametric subfamilies such that* $\mathcal{S} = \tilde{\mathcal{S}}$. *Then* d exists and $\psi = d - E[d]$.

The hypotheses state that the set of linear combinations of scores for all regular parametric subfamilies is rich enough to be able to approximate any mean zero vector, which should be interpreted as a condition that the scores associated with $\mathcal{F}$ and its parametric subfamilies are unrestricted, except for regularity conditions. The conclusion means that there is only one possible influence function, given by d-E[d]; thus, all asymptotically linear, regular estimators of $\hat{\mu}$ are asymptotically equivalent. Such a result is natural when $\mathcal{F}$ is not restricted, since $\mu(F)$ is exactly identified. Henceforth, d-E[d] will be referred to as the influence function of a functional satisfying the hypotheses of Theorem 2.1.

To apply this result to derive the variance of $\hat{\mu}$ when $\mathcal{F}$ is restricted, it is necessary to work with an extension of the functional $\mu(F)$. Suppose that $\mathcal{F}$ does not satisfy the hypotheses of Theorem 2.1, but that there is $\tilde{\mathcal{F}} \supset \mathcal{F}$ and an extension $\tilde{\mu}(F)$ of $\mu(F)$ to all of $\tilde{\mathcal{F}}$ (i.e. $\tilde{\mu}(F) = \mu(F)$ for $F \in \mathcal{F}$) that does. Then if $\tilde{\mu}(F)$ is pathwise differentiable with derivative $\tilde{d}$ and $\hat{\mu}$ is a regular estimator of $\tilde{\mu}(F)$ for a class of parametric subfamilies satisfying the hypotheses of Theorem 2.1, it follows that the influence function of $\hat{\mu}$ is $\tilde{d}$-E[$\tilde{d}$]. The regularity condition on $\hat{\mu}$ as an estimator of $\tilde{\mu}(F)$ means that local bias must be absent as F approaches F_0 along any path allowed in $\tilde{\mathcal{F}}$, so that $\tilde{\mu}(F)$ should be interpreted as the limit of $\hat{\mu}$ under any distribution in $\tilde{\mathcal{F}}$. That is, for a general semiparametric model the influence function of $\hat{\mu}$ will be the influence function for the functional that is the limit of the estimator under misspecification, i.e. when $F \notin \mathcal{F}$. Of course, when $F \in \mathcal{F}$, the resulting formula may simplify.

Although the hypotheses of this theorem include independence of the observations, $\psi = d - E[d]$ should still hold under dependence, where d is the derivative of an extended functional for the *marginal* distribution of z;

see Section 6 for results that allow for dependence.

For Von Mises (1947) functionals, which are those defined for all distribution functions, there is a Gateaux derivative formula for the influence function of an estimator: see Boos and Serfling (1980) and Serfling (1980). The Gateaux derivative is related to the pathwise derivative, and this relationship will be helpful for relating the results here to those of Section 6. The functional $\mu(F)$ is Gateaux differentiable at F_0 in the direction F_1 if for $\mu_0 = \mu(F_0)$,

$$(2.5)\qquad \Delta\mu(F_0,F_1-F_0) = \lim_{\lambda\to 0+}[\mu((1-\lambda)F_0 + \lambda F_1) - \mu_0]/\lambda$$

exists. Also, Δ is linear if there exists $\delta(z)$ such that for all F_1,

$$(2.6)\qquad \Delta\mu(F_0,F_1-F_0) = \int\delta(z)d(F_1-F_0)(z) = E_{F_1}[\delta] - E[\delta],$$

and $\delta(z)$ is referred to as the Gateaux derivative.

To relate the Gateaux and pathwise derivatives, note that the right hand side of (2.5) is the right derivative of the functional for a parameter $\theta = \lambda$, for a parametric subfamily consists of a convex combination of F_1 and F_0. If such a subfamily is regular (see Bickel, 1982, for conditions), then for densities $f_0(z)$ and $f_1(z)$ with respect to a carrier measure for F_0, the score for θ is

$$(2.7)\qquad S_\theta = \partial\ln[(1-\theta)f_0(z) + \theta f_1(z)]/\partial\theta|_{\theta=0} = \{[f_1(z)/f_0(z)] - 1\}.$$

In this case equation (2.1) for the pathwise derivative is

$$(2.8)\qquad \partial\mu(\theta)/\partial\theta|_{\theta=0} = E[d\{[f_1(z)/f_0(z)] - 1\}] = E_{F_1}[d] - E[d].$$

Thus, the Gateaux derivative coincides with the pathwise derivative for convex combination parametric subfamilies; compare equation (2.6). Note though that the definition of the pathwise derivative does not require that the subfamilies take this form. In fact, other types of parametric subfamilies may be more convenient for calculation of the pathwise derivative.

If the pathwise derivative calculation fails because $\tilde{\mu}(F)$ is not pathwise differentiable, then no $\sqrt{n}$-consistent, regular estimator exists; see Chamberlain (1986a) and Van der Vaart (1988). However, there are examples of pathwise differentiable functionals that have no $\sqrt{n}$-consistent estimator; see Ritov and Bickel (1987). Apparently, smoothness conditions in addition to those for pathwise differentiability are often required for $\sqrt{n}$-consistency. Such smoothness conditions will be discussed in Section 6.

The hypotheses of Theorem 2.1 are not primitive, but do result in a general formula for the influence function of an asymptotically linear semiparametric estimator, even when it depends on nonparametric estimates of functions. The point of Theorem 2.1 is to formalize the statement that "under sufficient regularity conditions" the influence function of a semiparametric estimator is the pathwise derivative, minus its expectation, of the functional that is the limit of the estimator under general misspecification. Of course, the usefulness of this formula depends on the simplicity of the pathwise derivative calculation. Some interesting examples where it is quite easy to find the derivative will be discussed below.

3. Calculating the Derivative

3.1 Specifying the Form of Parametric Subfamilies

One method of calculating d is based on parametric subfamilies of some specific form. By the conclusion of Theorem 2.1 d is invariant to the form of the subfamily such that $\hat{\mu}$ is regular and the richness hypothesis is satisfied. One example of a useful class of parametric subfamilies, that has been considered by Chamberlain (1986a) and others, has densities of the form

$$(3.1.1) \quad f(z|\theta) = f_0(z)[1 + \theta'\{s(z)-E[s(z)]\}], \quad \theta_0 = 0,$$

where $s(z)$ is bounded. Integration of $f(z|\theta)$ to one follows by the mean normalization of $s(z)$ and nonnegativity for θ close to θ_0 by boundedness of $s(z)$. It is easy to show that this subfamily is mean square differentiable and that the richness hypothesis is satisfied, even under rather strong smoothness restrictions on $s(z)$. Let q denote the dimension of z and let $C_c^\infty(\mathbb{R}^m)$ denote the set of functions with domain $\mathbb{R}^m$ that have compact support and continuous partial derivatives of all orders.

Lemma 3.1: If $s(z)$ *is bounded then* $f(z|\theta)^{1/2}$ *is smooth with score*

$$(3.1.2) \quad S_\theta = 1(f_0(z) > 0)\{s(z)-E[s(z)]\}.$$

In addition if $E[\psi'\psi] < \infty$, *then* $E_\theta[\psi'\psi]$ *is continuous at* $\theta_0 = 0$, *and* $\tilde{\mathcal{S}} = \mathcal{S}$, *even if each* $s(z)$ *is restricted to be in* $C_c^\infty(\mathbb{R}^q)$.

This result also states another convenient feature of this family, that the continuity condition for $E_\theta[\psi'\psi]$ is automatically satisfied.

Other types of parametric subfamilies, such as exponential families, may also be useful for calculation of the derivative in particular models. Also, it will often be possible to calculate the derivative by using general properties of scores, without imposing particular functional forms or specific regularity conditions. Indeed, regularity conditions for pathwise differentiability are somewhat unessential when the formula is only used as a guide to further results. Here, regularity conditions will be spelled out in some, but not all, of the pathwise derivative calculations.

3.2 An Example

The pathwise derivative formula for the influence function can be illustrated by considering a well understood but important example, U and V statistics. Although Serfling (1980) has given a Gateaux derivative interpretation of the influence function, a brief discussion of the pathwise derivative is useful for expository purposes.

Suppose that the observations are independent. For a bivariate function $a(z,\tilde{z})$ of $z \in \mathbb{R}^q$, $\tilde{z} \in \mathbb{R}^q$ consider

$$(3.2.1) \qquad \hat{\mu}_U = \Sigma_{i=1}^n \Sigma_{j=i+1}^n [a(z_i,z_j) + a(z_i,z_j)]/n(n-1),$$

$$\hat{\mu}_V = \Sigma_{i=1}^n \Sigma_{j=1}^n a(z_i,z_j)/n^2.$$

The statistic $\hat{\mu}_U$ is a U-statistic, introduced by Hoeffding (1948), and $\hat{\mu}_V$ is a corresponding V-statistic. $\hat{\mu}_U$ is an unbiased estimator of

$$(3.2.2) \qquad \mu(F) = \int\int a(z,\tilde{z})dF(z)dF(\tilde{z}),$$

and $E_F[|\hat{\mu}_U-\hat{\mu}_V|] \le (1/n)\{E_F[|\hat{\mu}_U|] + E[|a(z_1,z_2)|]\} = o(1/\sqrt{n})$, so that $\hat{\mu}_V$ is asymptotically equivalent to $\hat{\mu}_U$. Thus, it is natural to take the functional estimated by both $\hat{\mu}_U$ and $\hat{\mu}_V$ to be $\mu(F)$. Consider a parametric subfamily, for which the functional will be $\mu(\theta) = \tilde{\mu}(\theta,\theta)$, where $\tilde{\mu}(\theta,\tilde{\theta}) = \int\int a(z,\tilde{z})dF_\theta(z)dF_{\tilde{\theta}}(\tilde{z})$. By iterated integration and differentiation under the integral, $\partial\tilde{\mu}(\theta,\theta_0)/\partial\theta|_{\theta=\theta_0} = \partial E_\theta[\int a(z,\tilde{z})dF_0(\tilde{z})]/\partial\theta|_{\theta=\theta_0} = E[\int a(z,\tilde{z})dF_0(\tilde{z})\}S_\theta(z)'] = E[E[a(z_1,z_2)|z_1]S_\theta(z_1)']$, and similarly $\partial\tilde{\mu}(\theta_0,\theta)/\partial\theta|_{\theta=\theta_0} = E[E[a(z_2,z_1)|z_1]S_\theta(z_1)]$. Then the chain rule gives

(3.2.3) $$\partial\mu(\theta_0)/\partial\theta = \partial\tilde{\mu}(\theta,\theta_0)/\partial\theta|_{\theta=\theta_0} + \partial\tilde{\mu}(\theta_0,\theta)/\partial\theta|_{\theta=\theta_0} = E[dS_\theta'],$$

$$d(z_1) = E[a(z_1,z_2)|z_1] + E[a(z_2,z_1)|z_1],$$

The formula $\psi(z) = d(z) - E[d(z)]$ reproduces the well known "projection" form for the influence function of U and V statistics.

This formula for the influence function may simplify when F_0 takes on certain values. For instance, suppose that $z = (\varepsilon,x)$, where ε and x are independent, and consider $a(z,\tilde{z}) = a(\varepsilon,\tilde{x})$, so that $\hat{\mu}_V = \sum_{i=1}^n\sum_{j=1}^n a(\varepsilon_i,x_j)/n^2$. This example is important for bootstrap m-estimation of nonlinear simultaneous equations models; see Newey (1989b). By ε and x independent, the influence function reduces to

(3.2.4) $$\psi(z) = E[a(\varepsilon,x)|x] + E[a(\varepsilon,x)|\varepsilon] - 2E[a(\varepsilon,x)].$$

Note that by independence of x and ε, $\hat{\mu}_V$ is also an estimator of $E[a(\varepsilon,x)]$. However, the influence function for this functional is $a(\varepsilon,x)-E[a]$, which is not equal to $\psi(z)$. Thus, this example illustrates the importance of working with the functional that gives the limit of the estimator under general misspecification in the distribution of z. By the V-statistic structure of $\hat{\mu}_V$, the right functional is that of (3.2.2).

4. Expectation Functionals and Semiparametric M-Estimators

An interesting class of examples are those where $\hat{\mu}$ is a sample average that depends on a nonparametric estimate of a function, say

$$(4.1) \qquad \hat{\mu} = \sum_{i=1}^{n} a(z_i, \hat{h}(x_i))/n,$$

where $a(z,h)$ is a known function, x is a subvector of z, and $\hat{h}(x)$ is a nonparametric estimate of $h_0(x)$. Such estimators are of interest in their own right and play a key role in asymptotic distribution theory for semiparametric m-estimators.

To calculate the influence function of $\hat{\mu}$ via the pathwise derivative, note that if the limit of $\hat{h}(x)$ is $h(x,F)$ then $\hat{\mu}$ is a sample analog of

$$(4.2) \qquad \mu(F) = E_F[a(z,h(x,F))],$$

For a parametric subfamily, $\mu(\theta) = E_\theta[a(z,h(x,\theta))]$, where $h(x,\theta) = h(x,F_\theta)$. By differentiation under the integral and the chain rule,

$$(4.3) \qquad \partial\mu(\theta_0)/\partial\theta = \{\partial E_\theta[a(z,h_0(x))]/\partial\theta + \partial E[a(z,h(x,\theta))]/\partial\theta\}|_{\theta=\theta_o}$$

$$= E[a_0 S_\theta] + E[\partial a(z,h_0(x))/\partial h\{\partial h(x,\theta_0)/\partial\theta\}]$$

$$= E[a_0 S_\theta] + E[H(x)\partial h(x,\theta_0)/\partial\theta], \quad H(x) = E[\partial a(z,h_0(x))/\partial h|x],$$

where $a_0(z) = a(z,h_0(x))$. The pathwise derivative will exist if there exists $\alpha(z)$ such that for all parametric subfamilies,

$$(4.4) \qquad E[H(x)\partial h(x,\theta_0)/\partial\theta] = E[\alpha S_\theta].$$

Then $d(z) = a_0(z) + \alpha(z)$, and the influence function is

(4.5) $\psi(z) = a_0(z)-E[a_0] + \alpha(z)-E[\alpha(z)]$.

This influence function has an interesting structure. The leading term $a_0(z)-E[a_0]$ is the influence function for a sample average, which would be correct if $h_0(x)$ were used in place of $\hat{h}(x)$ in $\hat{\mu}$. Thus, the second term is an adjustment term for the estimation of $h_0(x)$, a nonparametric analog of adjustments that are familiar for two-step parametric estimators. It can also be interpreted as the pathwise derivative of the functional $E[a(z,h(x,F))]$.

When F_0 has a specific structure, the influence function function may simplify. In particular, if $H(x) = 0$, then $\alpha(z) = 0$, and no adjustment for estimation of h is present. This condition can be interpret as meaning that small deviations away from $h_0(x)$ do not affect $E[a(z,h(x))|x]$; the first term in a Taylor expansion of $E[a(z,h(x,\theta))|x]$ around θ_0 is $H(x)\partial h(x,\theta_0)/\partial\theta$. For instance, suppose that $a(z,h) = x\varepsilon/h$, with $E[\varepsilon|x] = 0$, as in Carroll (1982) and Robinson (1987), where x are regressors, ε is a disturbance, and h indexes the conditional variance of ε given x. Then $H(x) = E[x\varepsilon\circ\partial[1/h_0(x)]/\partial h|x] = -\{x/h_0(x)^2\}E[\varepsilon|x] = 0$.

To calculate the adjustment term $\alpha(z)$ when $H(x) \neq 0$, a solution to equation (4.4) must be found. The form of this solution will depend on the nature of $h(x)$. Here two examples will be considered, one where $h(x)$ is a derivative of a density function $f(x)$ for x and the other where $h(x)$ is the derivative of a conditional expectation of some variable w given x. Combinations of these cases will also be discussed.

The following notation will be used for derivatives. For $u \in \mathbb{R}^p$, a function $v(u)$, and a vector $\lambda = (\lambda_1,\ldots,\lambda_r)'$ of nonnegative integers, let $|\lambda| = \sum_{j=1}^{r}\lambda_j$ and denote a partial derivative by

$$D^{\lambda}v(u) = \partial^{|\lambda|}v(u)/\partial u_1^{\lambda_1}\circ\circ\circ\partial u_r^{\lambda_r}.$$

Consider first the case where $h(x) = D^{\lambda}f(x)$ for the density function $f(x)$ of x. Note that $h(x) = f(x)$ is included as a special case where $\lambda = 0$. It is well known that under appropriate regularity conditions the score can be decomposed as a sum of marginal and conditional scores, $S_{\theta}(z) = S_{\theta}(x) + S_{\theta}(\tilde{z}|x)$, where $\tilde{z}$ are the components of z other than x, and that $E[S_{\theta}(\tilde{z}|x)|x] = 0$. Assuming that $H(x)$ is continuously differentiable to order $|\lambda|$, and $f(x|\theta)$ and its derivatives are zero on the boundary of the support of x, by differentiation under the integral and integration by parts

$$(4.6) \qquad E[H(x)\partial\{D^{\lambda}f(x|\theta_0)\}/\partial\theta] = E[H(x)D^{\lambda}\{\partial f(x|\theta_0)/\partial\theta\}]$$

$$= \int H(x)f_0(x)D^{\lambda}\{\partial f(x|\theta_0)/\partial\theta\}dx = (-1)^{|\lambda|}\int D^{\lambda}\{H(x)f_0(x)\}[\partial f(x|\theta_0)/\partial\theta]dx$$

$$= E[(-1)^{|\lambda|}D^{\lambda}\{H(x)f_0(x)\}S_{\theta}(x)] = E[\alpha S_{\theta}], \quad \alpha(z) \equiv (-1)^{|\lambda|}D^{\lambda}\{H(x)f_0(x)\}.$$

For instance, if $\lambda = 0$ and $a(z,f) = f$, then $H(x) = 1$, $\alpha(z) = (-1)^0D^0f_0(x) = f_0(x)$, and $\psi(z) = 2\{f_0(x)-E[f_0(x)]\}$, a familiar result for the functional $E_F[f(x)] = \int f(x)^2dx$.

Consider next $h(x) = D^{\lambda}g(x)$, where $g(x,F) = E_F[w|x]$ for some variable w. Differentiating the conditional likelihood $f(w|x,\theta)$ under the integral,

$$\partial g(x,\theta_0)/\partial\theta = \partial\int wf(w|x,\theta_0)dw/\partial\theta = \int w\partial f(w|x,\theta_0)/\partial\theta dw = E[wS_{\theta}(w|x)|x]$$

$$= E[\{w-g_0(x)\}S_{\theta}(w|x)|x].$$

Then by an argument like that for equation (4.6),

$$(4.7) \qquad E[H(x)\partial\{D^{\lambda}g(x,\theta_0)\}/\partial\theta] = E[H(x)D^{\lambda}\{\partial g(x,\theta_0)/\partial\theta\}]$$

$= \int H(x)f_0(x)\mathcal{D}^\lambda\{\partial g(x,\theta_0)/\partial\theta\}dx = (-1)^{|\lambda|}\int\mathcal{D}^\lambda\{H(x)f_0(x)\}\partial g(x,\theta_0)/\partial\theta dx$

$= E[(-1)^{|\lambda|}\mathcal{D}^\lambda\{H(x)f_0(x)\}E[\{w-g_0(x)\}S_\theta(w,x)|x]/f_0(x)] = E[\alpha S_\theta],$

$\alpha(z) \equiv (-1)^{|\lambda|}\mathcal{D}^\lambda\{H(x)f_0(x)\}\{w-g_0(x)\}/f_0(x).$

For instance, if $a(z,h) = \omega(x)\partial g(x)/\partial x_1$ for a fixed weight function $\omega(x)$, so that $\mu(F)$ is a weighted average derivative functional with known weight, then $H(x) = \omega(x)$, $\alpha(z) = (-1)[\mathcal{D}^{(1,\dots 0)'}\{\omega(x)f(x)\}]\{w-g_0(x)\}/f(x)$, and

(4.8) $\psi(z) = \omega(x)\partial g(x)/\partial x_1-\mu_0 - \{\omega(x)\partial f(x)/\partial x_1/f(x) + \partial\omega(x)/\partial x_1\}\{w-g_0(x)\},$

reproducing a result of Newey and Stoker (1989).

Regularity conditions for pathwise differentiability in these examples include smoothness conditions such as the following:

Assumption 4.1: $a(z,h)$ is continuously differentiable in h, x has a density $f_0(x)$ with respect to Lebesgue measure, and $f_0(x)$ and $H(x) = E[\partial a(z,h_0(x))/\partial h|x]$ are continuously differentiable to order $|\lambda|$.

The following Assumption is a useful regularity condition in the density derivative case.

Assumption 4.2: For any $\tilde{s}(x) \in C_c^\infty(\mathbb{R}^r)$ there exists a function $A(z)$ and a neighborhood $\mathcal{N}$ of zero such that for $\delta(x,\theta) = \mathcal{D}^\lambda f_0(x) + \theta\mathcal{D}^\lambda[\tilde{s}(x)f_0(x)]$, $\sup_{\theta\in\mathcal{N}}[|a(z,\delta(x,\theta))|^2 + |\partial a(z,\delta(x,\theta))/\partial h|] \le A(z)$ and $E[A(z)] < \infty$;

Theorem 4.1: If Assumptions 4.1 and 4.2 are satisfied then there exists a class of parametric subfamilies such that the hypotheses of Theorem 2.2 are satisfied with influence function for $\mu(F) = E_F[a(z,f(x,F))]$ *given by* $\psi(z) = a(z,f_0(x))-E[a] + \alpha(z)-E[\alpha]$, *where* $\alpha(z) = (-1)^{|\lambda|}\mathcal{D}^\lambda\{H(x)f(x)\}$.

The following assumption is useful in the regression derivative case.

Assumption 4.3: w has a conditional density $f_0(w|x)$, with respect to some measure, that is continuously differentiable to order $|\lambda|$, and for all $\tilde{\lambda}$ with $|\tilde{\lambda}| \le |\lambda|$, and all x there exists a neighborhood $\mathcal{N}$ of zero such that $\int(1+|w|)\sup_{\alpha\in\mathcal{N}}|D_0^{\tilde{\lambda}}f(w|x+\alpha)|dw = 0$. There exists $C > 0$ and $A(z)$ such that $\sup_{|\delta|\le C}[|a(z,D^\lambda g_0(x)+\delta)|^2+|\partial a(z,D^\lambda g_0(x)+\delta)/\partial h|] \le A(z)$ and $E[A(z)] < \infty$;

Theorem 4.2: If Assumptions 4.1 and 4.3 are satisfied then there exists a class of parametric subfamilies such that the hypotheses of Theorem 2.2 are satisfied with influence function for $\mu(F) = E_F[a(z,g(x,F))]$ *given by* $\psi(z) = a(z,f_0(x))-E[a] + \alpha(z)-E[\alpha]$, *where* $\alpha(z) = (-1)^{|\lambda|}D^\lambda\{H(x)f_0(x)\}(w-g_0(x))/f_0(x)$.

These formulas can easily be generalized to the case where nonparametric estimates of several functions are present. Suppose that a depends on s functions $h_j(x_j)$, $(j=1,\ldots,s)$, of subvectors x_j of z, as $a(z,h_1(x_1),\ldots,h_s(x_s))$. Then equation (4.3) generalizes to

$$\partial\mu(\theta_0)/\partial\theta = E[a_0S_\theta] + \textstyle\sum_{j=1}^{s}E[H_j(x)\partial h_j(x_j,\theta_0)/\partial\theta], \tag{4.9}$$

$$H_j(x) = E[\partial a(z,h_{10}(x_1),\ldots,h_{s0}(x_s))/\partial h_j|x].$$

The derivative will have an outer product form if for some $\alpha_j(z)$,

$$E[H_j(x)\partial h_j(x_j,\theta_0)/\partial\theta] = E[\alpha_jS_\theta], \quad (j=1,\ldots,s), \tag{4.10}$$

in which case $d(z) = a_0(z) + \sum_{j=1}^{s}\alpha_j(z)$ and

(4.11) $\psi(z) = a_0(z)-E[a_0] + \sum_{j=1}^{s}\{\alpha_j(z)-E[\alpha_j]\}.$

This influence function has a structure analogous that previously discussed, only there are multiple adjustment factors, one for each function that is estimated. When $h_j(x)$ is a partial derivative of a density or a regression function, then the corresponding $\alpha_j(z)$ can be calculated from equation (4.6) or (4.7) respectively, with x, $H(x)$, $f_0(x)$, $g_0(x)$, and w indexed by j. For the weighted average derivative functional, if the weight is $\omega(x) = f_0(x)$ and $f_0(x)$ is estimated nonparametrically, then an additional correction term $\alpha_\omega(z) = (-1)^0 D^0\{E[\partial[\omega(x)\partial g_0(x)/\partial x_1]/\partial\omega|x]f_0(x)\} = \omega(x)\partial g(x)/\partial x_1$ is present, and $\psi_\omega(z) = \omega(x)\partial g_0(x)/\partial x_1 - \mu_0$ must be added to the right-hand side of equation (4.8) to obtain the correct influence function, reproducing a result of Powell, Stock, and Stoker (1989). Note that because $\psi_\omega(z)$ is perfectly correlated with the first term in equation (4.8) and the first term is uncorrelated with the second, estimating the weight $\omega(x) = f_0(x)$ rather than using the true density as a weight increases the asymptotic variance.

A particular $\alpha_j(z)$ will be zero if $H_j(x) = 0$. For instance, if $a(z,h_1,h_2) = [w_1-g_1(x)][w_2-g_2(x)]$ for $E[w_j|x] = g_{0j}(x)$, $(j=1,2)$, then $H_1(x) = -E[w_2-g_{02}(x)|x] = 0 = -E[w_1-g_{01}(x)|x] = H_2(x)$. Thus, no adjustment for nonparametric estimation is indicated for the sample cross product of two nonparametric residuals from regression on the same set of variables x, a result for kernel estimators due to Robinson (1988). When $w_1 = w_2$, where $g_0(x)$ solves $\min_{g(x)}E[(w-g(x))^2]$, this result follows by applying the envelope theorem; $\partial\min_{g(x)}E_\theta[(w-g(x))^2]/\partial\theta = \partial E_\theta[(w-g_0(x))^2]/\partial\theta = E[(w-g_0(x))^2S_\theta]$.

For brevity, regularity conditions for the vector h case will not be given. It would be straightforward, but notationally cumbersome, to combine

Assumptions 4.1 - 4.3 to formulate regularity conditions.

These results should prove useful in many new examples. For instance, as suggested by Pagan (1988) the functional $\mu(F) = E[s(x)^2]$, where x is a scalar and $s(x) = [\partial f(x)/\partial x]/f(x)$, which is the location information for a density F, may be of interest for assessing the efficiency gain from using adaptive estimation. If x is a parametric residual, then $\mu(F)$ will be a measure of the asymptotic efficiency of an adaptive estimator. Calculating as above for $h_1(x) = f(x)$, $h_2(x) = \partial f(x)/\partial x$, $H_1(x) = -2s(x)^2/f_0(x)$, and $H_2(x) = 2s(x)/f_0(x)$, and noting that by the usual information equality $E[\partial s(x)/\partial x] = -E[s(x)^2]$, the influence function for $\mu(F)$ is

$$(4.12) \qquad \psi(z) = -s(x)^2 - 2\partial s(x)/\partial x - \mu_0.$$

Another important class of examples are semiparametric m-estimators that depend on nonparametric estimates of some function. A general formulation of such estimators, which is like that of Ritov (1987) or Andrews (1988a), is the following. Let $\hat{h}(z,\mu)$ be a nonparametric estimate of some function $h(z,\mu)$ that can depend on μ, and let $m(z,\mu,h(z))$ be a fixed function of z, μ, and h. A semiparametric m-estimator is one which solves an asymptotic moment equation

$$(4.13) \qquad \sum_{i=1}^{n} m(z_i,\hat{\mu},\hat{h}(x_i,\hat{\mu}))/\sqrt{n} = o_p(1).$$

The general idea here is that $\hat{\mu}$ is obtained by a procedure that first "concentrates out" the nonparametric function $h(z,\mu)$. An early and important example is the Buckley and James (1979) estimator for censored regression; see also Horowitz (1986) and Ritov (1987).

By the usual method of moments reasoning, $\hat{\mu}$ is a sample analog of the functional $\mu(F)$ that is implicitly defined by

$$(4.14) \qquad E_F[m(z,\mu(F),h(x,\mu(F),F))] = 0,$$

where $h(z,\mu,F)$ is the limit of $\hat{h}(z,\mu)$ under a general distribution F. For a parametric subfamily $E_\theta[m(z,\mu(\theta),h(x,\mu(\theta),\theta))] = 0$ corresponds to (4.14). Assuming that $E_\theta[m(z,\mu,h(x,\mu,\theta))]$ is differentiable in μ and θ, that $M \equiv \partial E[m(z,\mu,h(x,\mu))]/\partial\mu|_{\mu=\mu_0}$ is nonsingular, differentiation under the integral and the implicit function theorem give

$$(4.15) \qquad \partial\mu(\theta_0)/\partial\theta = -M^{-1}\{E[m_0 S_\theta'] + E[\{\partial m_0(z,h_0(x))/\partial h\}\partial h_0(x,\theta_0)/\partial\theta]\}$$

$$= -M^{-1}\{E[m_0 S_\theta'] + E[H_m(x)\partial h_0(x,\theta_0)/\partial\theta]\}, \quad H_m(x) = E[\partial m_0(z,h_0(x))/\partial h|x],$$

where $h_0(x,\theta) = h(x,\mu_0,\theta_0)$, $h_0(x) = h_0(x,\theta_0)$, and $m_0(z,h_0(x)) = m(z,\mu_0,h_0(x))$. Assuming that there exists $\alpha_m(z)$ such that

$$(4.16) \qquad E[H_m(x)\partial h_0(x,\theta_0)/\partial\theta] = E[\alpha_m S_\theta'],$$

the pathwise derivative is $d(z) = -M^{-1}[m_0(z,h_0(x)) + \alpha_m(z)]$. Noting that $E[m_0(z,h_0(x))] = 0$ holds by equation (4.14), the influence function is

$$(4.17) \qquad \psi(z) = -M^{-1}\{m_0(z,h_0(x)) + \alpha_m(z)-E[\alpha_m]\}.$$

This influence function is quite similar to that of equation (4.5). This similarity can be explained by the usual Taylor expansion argument for m-estimators. Assuming $m(z,\mu,\hat{h}(x,\mu))$ is differentiable in μ, let $\hat{M}(\mu) = \sum_{i=1}^n \partial m(z_i,\mu,\hat{h}(x_i,\mu))/\partial\mu/n$. Expanding around μ_0 and solving gives

$$(4.18) \qquad \sqrt{n}(\hat{\mu}-\mu_0) = -\hat{M}(\bar{\mu})^{-1}\sqrt{n}[\textstyle\sum_{i=1}^n m_0(z_i,\hat{h}_0(x_i))/n] + -\hat{M}(\bar{\mu})^{-1}o_p(1),$$

$$= -M^{-1}\sqrt{n}[\textstyle\sum_{i=1}^n m_0(z_i,\hat{h}_0(x_i))/n] + o_p(1),$$

where $\bar{\mu}$ is the intermediate value and the final equality will follow from consistency of $\hat{\mu}$ and uniform convergence of $\hat{M}(\mu)$. Thus, we expect the influence function for $\hat{\mu}$ to be $-M^{-1}$ times the influence function for $\sum_{i=1}^{n} m_0(z_i, \hat{h}_0(x_i))/n$, which is an estimate of the functional $\mu_m(F) = E_F[m_0(z, h_0(x,F))]$. This is exactly the structure of equation (4.17). A correction term $\alpha_m(z) - E[\alpha_m]$ for the estimation of $h_0(x,F)$ is included.

This correction term can be interpreted as the pathwise derivative of the functional $E[m_0(z,h(x,F))]$. All the previous results apply to its calculation. If $H_m(x) = 0$, then $\alpha_m(x) = 0$. If $h_0(x,F)$ consists of density and regression derivatives, then $\alpha_m(z)$ will be the sum of components that can be obtained from equations (4.6) and (4.7) as appropriate.

Finding the influence function for a semiparametric estimator also requires calculation of M. Although finding M may be difficult when $h(x,\mu)$ varies with μ, the following result from the semiparametric efficiency literature is useful. As noted in Newey (1989b), $M = -E[\{m_0 + \alpha_m - E(\alpha_m)\}S']$, where S is the *efficient score* for μ in a semiparametric model with independent observations and likelihood $f(z|\mu,\tilde{h})$, and $\tilde{h}$ us an unknown function; see e.g. Newey(1989a) for exposition. Thus, when α_m and the efficient score exist in closed form, M can be calculated from this outer product formula.

5. Efficiency Comparisons for the Heteroskedastic Partially Linear Model.

The pathwise derivative can be used to compare asymptotic efficiencies of different estimators, without detailed consideration of particular nonparametric estimation methods. Such comparisons can be used to suggest a relatively efficient estimator, for which specific regularity conditions might be worked out. To illustrate this procedure, and provide further examples, consider the model

$$(5.1) \qquad y = r'\mu_0 + h_0(x) + \varepsilon, \quad E[\varepsilon|X] = 0, \quad E[\varepsilon^2|X] = \sigma^2(X), \quad X = (r',x')',$$

where $h_0(x)$ and $\sigma^2(X)$ are unknown. This is a heteroskedastic version of the partially linear model introduced by Engle et. al. (1986).

Estimators of μ_0 have been considered by N. Heckman (1986), Rice (1986), Schick (1986), Robinson (1988), Chamberlain (1986b), and Andrews (1988). An estimator analogous to that of Robinson (1988), can be formed as the solution $\hat{\mu}_1$ of

$$(5.2) \qquad \textstyle\sum_{i=1}^{n} m(z_i,\hat{\mu}_1,\hat{g}_r(x_i),\hat{g}_y(x_i))/n = 0,$$

$$m(z,\mu,g_r,g_y) = (r-g_r)(y-g_y-\{r-g_r\}'\mu),$$

where $\hat{g}_r(x)$ and $\hat{g}_y(x)$ are nonparametric estimators of the conditional expectations $g_r(x,F) = E_F[r|x]$ and $g_y(x,F) = E_F[y|x]$ of y and r given x, respectively; Robinson (1988) considered kernel estimators. Evidently, such an estimator is a semiparametric m-estimator, so that its influence function can be calculated as in Section 4. Since the nonparametric estimates do not depend on μ, $M = E[\partial m(z,\mu_0,E[r|x],E[y|x])/\partial\mu] = -E[\mathrm{Var}(r|x)]$. Also, since $E_F[m(z,\mu_0,g_r(x,F),g_y(x,F))]$ is a linear combination of covariances between nonparametric residuals, it follows from the discussion in Section 4 that $\alpha_m(z) = 0$. Thus, the influence function of $\hat{\mu}_1$ is $\psi(z) = -M^{-1}m_0 = \{E[\mathrm{Var}(r|x)]\}^{-1}(r-E[r|x])\varepsilon$, so that the asymptotic variance of $\hat{\mu}_1$ is

$$(5.3) \qquad \Omega_1 = E[\psi\psi'] = M^{-1}E[\sigma^2(X)(r-E[r|x])(r-E[r|x])']M^{-1}.$$

A lower bound for the asymptotic variance of estimators of μ_0 was derived by Chamberlain (1987). To describe the bound, consider $\omega(X) > 0$ with $E[\omega(X)] < \infty$, and let $E_F^\omega[\circ] = E_F[\omega(X)(\circ)]/E_F[\omega(X)]$. For any event $\mathcal{A}$ that is measurable with respect to x, $E_F^\omega[1(\mathcal{A})E_F[\omega(X)(\circ)|x]/E_F[\omega(X)|x]] = E_F[\omega(X)E_F[1(\mathcal{A})\omega(X)(\circ)|x]/E_F[\omega(X)|x]]/E_F[\omega(X)] = E_F[1(\mathcal{A})\omega(X)(\circ)]/E_F[\omega(X)] = E^\omega[1(\mathcal{A})(\circ)]$. Thus, $E_F^\omega[\circ|x]$ has the explicit representation

$$(5.4)\qquad E_F^\omega[\circ|x] = E_F[\omega(X)(\circ)|x]/E_F[\omega(X)|x].$$

Chamberlain's (1987) formula for the bound can then be written as

$$(5.5)\qquad \Omega^* = \{E[\sigma^{-2}(X)\{r-E^{\sigma^{-2}}[r|x]\}\{r-E^{\sigma^{-2}}[r|x]\}']\}^{-1}.$$

In general, $\Omega_1 \neq \Omega^*$, although equality occurs if $\sigma^2(X)$ is constant. This result is analogous to the inefficiency of ordinary least squares under heteroskedasticity.

One possible correction is to weight by $\sigma^{-2}(X_i)$. Consider $m^\sigma(z,\mu,g_r,g_y) = \sigma^{-2}(X)(r-g_r)(y-g_y-\{r-g_r\}'\mu)$, and $\hat{\mu}_2$ obtained as the solution to equation (5.2) with m^σ replacing m. To calculate the influence function of $\hat{\mu}_2$, note that $E_F[m(z,\mu_0,g_r(x,F),g_y(x,F))]$ is no longer a covariance between nonparametric residuals, and a correction term term for the estimation of g_r and g_y is required. Calculating expected derivatives as in Section 4,

$$(5.6)\qquad H_{my}(x) \equiv E[\partial m(z,\mu_0,E[r|x],E[y|x])/\partial g_y|x] = -E[\sigma^{-2}(X)\{r-E[r|x]\}|x],$$

$$H_{mr}(x) \equiv E[\partial m(z,\mu_0,E[r|x],E[y|x])/\partial g_r|x] = -H_{my}(x)\mu_0'.$$

Applying equation (4.7) to each of these terms gives $\alpha_m(z) = H_{my}(x)\{y-E[y|x]-\{r-E[r|x]\}'\mu_0\} = H_{my}(x)\varepsilon$. Noting that $M = E[\partial m(z,\mu_0,E[r|x],E[y|x])/\partial\mu] = E[\sigma^{-2}(X)\{r-E[r|x]\}\{r-E[r|x]\}']$ and $E[\alpha_m] = 0$, it follows that the influence function is $\psi = M^{-1}(m_0+\alpha_m) =$

$M^{-1}\{\omega(X)(r-E[r|x])+H_{my}(x)\}\varepsilon$. The corresponding asymptotic variance is

$$(5.7)\quad \Omega_2 = M^{-1}E[\sigma^2(X)\{\sigma^{-2}(X)(r-E[r|x])+H_{my}(x)\}\{\sigma^{-2}(X)(r-E[r|x])+H_{my}(x)\}']M^{-1}.$$

In general, $\Omega_2 \neq \Omega^*$, although equality occurs if $\sigma^{-2}(X)$ is a function only of x; then $E^{\sigma^{-2}}[r|x] = E[\sigma^{-2}(x)r|x]/E[\sigma^{-2}(x)|x] = E[r|x]$, and $H_{my}(x) = \sigma^{-2}(x)E[r-E[r|x]|x] = 0$.

In practice, $\sigma^2(X)$ would have to be estimated, leading to the presence of an additional term that corrects for its estimation. Recall that to apply Theorem 2.1 it is necessary to specify the limit of the estimator under misspecification. Consider weighting by a nonparametric estimator of the conditional variance $\Sigma_F^2(X)$ of y given X, say $\hat{\Sigma}^2(X) = \hat{g}_{y2}(X)-[\hat{g}_{y1}(X)]^2$, where $\hat{g}_{yj}(X)$ is a nonparametric regression estimator of $E_F[y^j|X]$, which will be consistent for $\sigma^2(X)$ when the model is correct. Then a version of $\hat{\mu}_2$ with an estimated weight could be obtained from equation (5.2) with m^{σ} replacing m and $\hat{\Sigma}^2(X)^2$ replacing $\sigma^2(X)$. The influence function adjustment for estimation of $\sigma^2(X)$ depends on

$$(5.8)\qquad H_{\Sigma}(X) = E[\partial m_0(z,E[r|x],E[y|x],\sigma^2(X))/\partial\Sigma^2|X]$$

$$= -E[(r-E[r|x])\varepsilon/(\sigma^2(X))^2|X] = 0,$$

by $E[\varepsilon|X] = 0$. Thus, just as in the parametric linear model, no adjustment for estimation of $\Sigma^2(X)$ is required.

The form of an efficient estimator is suggested by the presence of $E^{\sigma^{-2}}[r|x]$ in the bound. In addition to weighting by $\hat{\Sigma}^{-2}(X)$, consider replacing the nonparametric estimators of $E[r|x]$ and $E[y|x]$ in equation (5.2) by estimators of $E^{\sigma^{-2}}[r|x]$ and $E^{\sigma^{-2}}[y|x]$. By equation (5.4), $\hat{\Sigma}^2(X)$ can be used to form estimators $\hat{g}_y^{\Sigma}(x)$ and $\hat{g}_r^{\Sigma}(x)$ of $E_F^{\Sigma^{-2}}[y|x]$ and $E_F^{\Sigma^{-2}}[r|x]$ respectively, as ratios of nonparametric regressions of $\hat{\Sigma}^2(X)y$ on x and of $\hat{\Sigma}^2(X)r$ on x to a nonparametric regression of $\hat{\Sigma}^2(X)$ on x, respectively. An estimate of μ can then be obtained from the solution to an equation analogous to (5.2),

$$(5.9) \qquad \sum_{i=1}^{n} m(z_i, \hat{\mu}^*, \hat{g}_r^{\Sigma}(x_i), \hat{g}_y^{\Sigma}(x_i), \hat{\Sigma}^2(X_i))/n = 0,$$

$$m(z,\mu,g_r^{\Sigma},g_y^{\Sigma},\Sigma^2) = (r-g_r^{\Sigma})(y-g_y^{\Sigma}-\{r-g_r^{\Sigma}\}'\mu)/\Sigma^2.$$

This estimator can be motivated by the weighted least problem

$$(5.10) \qquad \min_{\mu,h(x)} E[\sigma^{-2}(X)\{y-r'\mu-h(x)\}^2] = C\min_{\mu,h(x)} E^{\sigma^{-2}}[\{y-r'\mu-h(x)\}^2],$$

where $C = E[\sigma^{-2}(X)]$. By analogy with the parametric case, one would expect that the optimal estimator of μ solves a sample analog of this equation. For a given μ the solution for h is $h(x,\mu) = E^{\sigma^{-2}}[(y-r'\mu)|x] = E^{\sigma^{-2}}[y|x] - E^{\sigma^{-2}}[r|x]'\mu$. Thus, the first order conditions for μ are

$$E[\sigma^{-2}(X)(r-E^{\sigma^{-2}}[r|x])\{y-E^{\sigma^{-2}}[y|x] - (r-E^{\sigma^{-2}}[r|x])'\mu\}] = 0.$$

Equation (5.9) is a sample analog of this equation. The presence of $E^{\sigma^{-2}}[y|x]$ and $E^{\sigma^{-2}}[r|x]$ provides an optimal way of "concentrating out" h(x).

To calculate the influence function of $\hat{\mu}^*$, note first that $M = E[\partial m(z,\mu_0,E^{\sigma^{-2}}[r|x],E^{\sigma^{-2}}[y|x],\sigma^2(X))/\partial\mu] = -(\Omega^*)^{-1}$. Also, $E^{\sigma^{-2}}[\varepsilon|x] = 0$, so that $y-E^{\sigma^{-2}}[y|x] - \{r-E^{\sigma^{-2}}[r|x]\}'\mu_0 = \varepsilon$, and

$$H_{mr}(x) \equiv E[\partial m(z,\mu_0,E^{\sigma^{-2}}[r|x],E^{\sigma^{-2}}[y|x],\sigma^2(X))/\partial g_r^{\Sigma}|x]$$

$$= \{-E[\varepsilon/\sigma^2(X)|x] + E[\sigma^{-2}(X)\{r-E^{\sigma^{-2}}[r|x]\}|x]\}\mu_0'$$

$$= \{-E[E[\varepsilon/\sigma^2(X)|X]|x] + E[\sigma^{-2}(X)|x]E^{\sigma^{-2}}[\{r-E^{\sigma^{-2}}[r|x]\}|x]\}\mu_0' = 0,$$

$$H_{my}(x) \equiv E[\partial m(z,\mu_0,E^{\sigma^{-2}}[r|x],E^{\sigma^{-2}}[y|x],\sigma^2(X))/\partial g_y^{\Sigma}|x]$$

$$= E[\sigma^{-2}(X)\{r-E^{\sigma^{-2}}[r|x]\}|x] = 0,$$

$$H_{m\Sigma}(X) \equiv E[\partial m(z,\mu_0,E^{\sigma^{-2}}[r|x],E^{\sigma^{-2}}[y|x],\sigma^2(X))/\partial\Sigma^2|X]$$

$$= -E[\{r-E^{\sigma^{-2}}[r|x]\}\varepsilon/(\sigma^2(X))^2|X] = 0.$$

Thus, no adjustment is required for the estimation of the nonparametric components. Therefore, the influence function of this estimator is $\psi = \Omega^*\sigma^{-2}(X)\{r-E^{\sigma^{-2}}[r|x]\}\varepsilon$, with corresponding asymptotic variance matrix equal to Ω^*, the bound.

In summary, the pathwise derivative formulas suggests that if the sample moments are weighted by an estimator of the conditional variance of y, and estimators of corresponding weighted conditional expectations are used in the moments, then the resulting estimator will be efficient. It should be possible to work out regularity conditions for such an estimator, although this task is beyond the scope of this paper.

6. Formulating Regularity Conditions for Asymptotic Normality

The pathwise derivative provides a guide to the specification of regularity conditions for asymptotic normality. At the very least, the derivative formula for $\psi(z)$ can be used as a starting point for the verification of equation (2.2). Furthermore, by imposing a nonparametric interpretation of the derivative, it is possible to formulate general regularity conditions for asymptotic normality, for several different types of estimators.

One type of estimator for which general regularity conditions can be formulated takes the form $\hat{\mu} = \tilde{\mu}(\hat{F})$ for some functional $\tilde{\mu}$ and estimator $\hat{F}$ of the distribution of F. An example of such an estimator is

$$(6.1) \qquad \hat{\mu} = \int_a^b \hat{g}(x)dx, \quad \hat{g}(x) = \sum_{i=1}^n w_i K((x-x_i)/\sigma)/\sum_{i=1}^n K((x-x_i)/\sigma),$$

where x is a scalar and K(u) a kernel satisfying $\int K(u)du = 1$, so that $\hat{g}(x)$ is a kernel estimator of $g_0(x) = E[w|x]$. If w were demand and x price, then $-\hat{\mu}$ would be a nonparametric estimator of the expected change in consumer surplus for a price movement from a to b. For $1_a^b = 1(a \le x \le b)$ let

$$(6.2) \qquad \tilde{\mu}(F) = \int_a^b E_F[w|x]dx = E_F[1_a^b w/f(x)] = \int_a^b [\int wf(w,x)dw)/\int f(w,x)dw]dx,$$

where f(w,x) is the joint density of w, x and f(x) the marginal density of x. It is easy to check that $\hat{\mu} = \mu(\hat{F})$ for a kernel density estimator $\hat{f}(w,x) = \sum_{i=1}^n \mathcal{K}((w-w_i)/\sigma, (x-x_i)/\sigma)/n\sigma^2$ such that $\int \mathcal{K}(v,u)dv = K(u)$ and $\int v\mathcal{K}(v,u)dv = 0$.

The following Assumption will be sufficient for asymptotic linearity of estimators of this type.

Assumption 6.1: There exists a set of measures $\mathcal{F}$ and a functional $\tilde{\mu}(F)$ defined on $\mathcal{F}$ such that i) $F_0 \in \mathcal{F}$, $\mu_0 = \tilde{\mu}(F_0)$, and there exists $\hat{F}$ such that with probability approaching one $\hat{F} \in \mathcal{F}$ and $\hat{\mu} = \tilde{\mu}(\hat{F})$; ii) For $F \in \mathcal{F}$, $|\tilde{\mu}(F)-\mu_0 - \int d(z)d(F-F_0)| \le R(F,F_0)$; iii) $n^{1/2}R(\hat{F},F_0) = o_p(1)$; iv) $\sqrt{n}\{\int d(z)d\hat{F} - \sum_{i=1}^n d(z_i)/n\} = o_p(1)$.

If $\mathcal{F}$ is a linear space with norm $\|\circ\|$ and $R(F,F_0) = o(\|F-F_0\|)$, then condition ii) implies that d is the Frechet derivative of $\tilde{\mu}(F)$, although iii) may impose a strong condition on the remainder $R(F,F_0)$. Sufficient conditions for iii) are $R(F,F_0) = O(\|F-F_0\|^2)$, corresponding to a Lipschitz derivative in finite dimensional spaces, and $n^{1/4}\|\hat{F}-F_0\|$ bounded in probability, which is a familiar convergence rate for semiparametric problems; e.g. see Schick (1986). In general one expects that $d(z)$ satisfying iii) will be a pathwise derivative of the functional $\tilde{\mu}(F)$; a Frechet derivative is also a Gateaux derivative and, as discussed in Section 2, the Gateaux derivative is the pathwise derivative for parametric subfamilies consisting of convex combinations of densities.

Condition v) says that the the integral of $d(z)$ over the estimated measure $\hat{F}$ must approach the integral of $d(z)$ over the empirical measure at a $\sqrt{n}$ rate. The validity of this condition will depend on the exact nature of $\hat{F}$. If $\hat{F}$ is a smoothed empirical measure, such as a kernel density function, then condition iv) specifies a convergence rate for smoothing bias. To see this suppose that $\hat{F}$ has density $\hat{f}(z) = \sum_{i=1}^{n}\mathcal{K}((z-z_i)/\sigma)/n\sigma^q$ with $\int\mathcal{K}(u)du = 1$. Then by a change of variables $u = (z-z_i)/\sigma$,

$$\int\alpha(z)d\hat{F} - \sum_{i=1}^{n}\alpha(z_i)/n = \sum_{i=1}^{n}\int[\alpha(z_i-u\sigma)-\alpha(z_i)]K(u)du/n, \tag{6.3}$$

a bias term that is familiar from work on kernel-based semiparametric estimators; e.g. see Robinson (1988) or Powell, Stock, and Stoker (1989). This term will be $o_p(1/\sqrt{n})$ under certain smoothness conditions and a sufficiently fast rate of convergence of σ to zero; e.g. see Lemma B.5.

Assumption 6.1 delivers the following result:

Theorem 6.1: If Assumption 6.1 is satisfied then

$$\sqrt{n}(\hat{\mu}-\mu_0) = \sum_{i=1}^{n}\{d(z_i)-E[d]\}/\sqrt{n} + o_p(1).$$

It is interesting to note that none of the conditions of Assumption 6.1 require independent observations, so that this result allows for dependence. Each of the conditions i) - iii) pertain to the marginal distribution of z

and its estimator $\hat{F}$. Some dependence restrictions may be useful for verifying iv).

Asymptotic normality of $\hat{\mu}$ will follow from asymptotic linearity in the way discussed in Section 2. To carry out asymptotic inference it is helpful to have a consistent estimator of the asymptotic variance, which can be constructed from a sufficiently well behaved estimate of the influence function. For example, if the observations are independent and $\hat{\psi}_i$ is an estimator of $\psi(z_i)$ such that $\sum_{i=1}^{n}\|\hat{\psi}_i-\psi(z_i)\|^2/n = o_p(1)$, then $\sum_{i=1}^{n}\hat{\psi}_i\hat{\psi}_i'/n$ will be consistent; e.g. see Powell, Stock, and Stoker (1989). For dependent observations, one could use a weighted autocovariance estimator, as in Newey and West (1987). Further details and primitive regularity conditions for consistency of asymptotic variance estimators are beyond the scope of this paper.

It is straightforward to use Assumption 6.1 to specify conditions for asymptotic linearity of $\hat{\mu} = \int_a^b \hat{g}(x)dx$ from equation (6.1). To do so it is useful to impose some restrictions on the kernel. Note that the estimator depends only on $z = (w,x)$. For a vector $u \in \mathbb{R}^q$ let λ denote a vector of nonnegative integers as in Section 4 and $u^\lambda = \prod_{j=1}^{p} u_j^{\lambda_j}$.

Assumption K: $\mathcal{K}(z)$ is an even function of z and there are integers $k \geq 1$ and $\ell \geq 0$ such that $K(u) = \int\mathcal{K}(v,u)du$, $\int v\mathcal{K}(v,u)dv = 0$, $\int\mathcal{K}(u)^2du < \infty$, $\int\mathcal{K}(u)du = 1$, $\int u^\lambda\mathcal{K}(u)du = 0$ for all $1 \leq |\lambda| \leq k-1$, and for all $|\lambda| \leq \ell$ and $\xi_\lambda(t) \equiv \int e^{-it'u}\mathcal{D}^\lambda K(u)du$, $\int|\mathcal{D}^\lambda K(u)|du < \infty$, $\int|\xi_\lambda(t)|dt < \infty$.

The absolute integrability of the Fourier transform of the kernel and its derivatives is useful for obtaining uniform convergence rates for kernel regression derivatives analogous to those of Bierens (1987).

It is also useful to impose the following restriction on the distribution of the data.

Assumption X: z_i are i.i.d., $E[w^2] < \infty$, x is continuously distributed with density $f_0(x)$ that is continuously differentiable to order $\ell \geq 2$ with bounded partial derivatives, $g_0(x) \equiv E[w|x]$ is continuously differentiable to order $\ell \geq 2$ and $D^{\lambda}[f_0(x)g_0(x)]$ is bounded for all $|\lambda| \leq \ell$.

The independence assumption could be relaxed if sufficient other conditions for uniform convergence rates for kernel density and regression estimators were imposed.

Theorem 6.2: Suppose that Assumptions K and X are satisfied with $\ell = 0$ *and* $\ell = k = 3$, $f_0(x) > 0$ *for* $x \in [a,b]$, *and* $\sigma = \sigma(n)$ *such that* $\sigma(n)^6 n \longrightarrow 0$, $\sigma(n)^4 n \longrightarrow \infty$. *Then for* $\psi(z) = 1(a \leq x \leq b) f_0(x)^{-1}[w - g_0(x)]$ *and* $\hat{\psi}_i = 1(a \leq x_i \leq b)\hat{f}(x_i)^{-1}[w_i - \hat{g}(x_i)]$,

$$\sqrt{n}[\int_a^b \hat{g}(x)dx - \int_a^b g_0(x)dx] \xrightarrow{d} N(0,\ E[\psi^2]), \quad \sum_{i=1}^n \hat{\psi}_i^2/n \xrightarrow{p} E[\psi^2].$$

Another important type of estimator is of the form

$$(6.4) \qquad \hat{\mu} = \sum_{i=1}^n a(z_i, \hat{h}(x_i))/n,$$

which was discussed in Section 4. This estimator will often not be a special case of (6.1), because the sample average corresponds to the empirical measure, while $\hat{h}(x)$ often corresponds to some other estimate of the distribution of z, such as a kernel. Nevertheless, general regularity conditions can be formulated by combining Assumption 6.1 with Andrews (1989a,b) stochastic equicontinuity condition. Let

$$\Delta(h) \equiv \sum_{i=1}^n [a(z_i, h(x_i)) - a(z_i, h_0(x_i))]/n - \int [a(z,h(x)) - a(z,h_0(x))]dF_0(z)\}.$$

Assumption 6.2: There exists a set $\mathcal{H}$ of functions of x and a pseudo-metric $\rho_{\mathcal{H}}(h_1,h_2)$ on $\mathcal{H}$ such that i) $h_0(x) \in \mathcal{H}$ and $\hat{h}(x) \in \mathcal{H}$ with probability approaching one; ii) $\rho_{\mathcal{H}}(\hat{h},h_0) = o_p(1)$; iii) for all $\epsilon, \eta > 0$ there exists $\delta > 0$ such that $\overline{\lim}_{n\to\infty}\text{Prob}^*(\sup_{\{h\in\mathcal{H}:\rho_{\mathcal{H}}(h,h_0)\le\delta\}}\sqrt{n}|\Delta(h)| > \eta) < \epsilon$, where Prob^* denotes outer probability.

Part iii) is a stochastic equicontinuity condition, for which Andrews (1989b) gives a number of primitive conditions. They typically involve smoothness restrictions on $a(z,h(x))$, as well as existence of sufficient moments.

Theorem 6.3: If Assumption 6.2 is satisfied and Assumption 6.1 is satisfied for $\hat{\mu}^{\alpha} \equiv \int a(z,\hat{h}(x))dF_0(z)$, *then for the derivative* $\alpha(z)$ *of the functional* $\tilde{\mu}(F)$ *from Assumption 6.1,*

$$\sqrt{n}(\hat{\mu}-\mu_0) = \textstyle\sum_{i=1}^{n}\{a(z_i,h_0(x))-E[a_0] + \alpha(z_i)-E[\alpha]\}/\sqrt{n} + o_p(1).$$

To interpret this result, note that the influence function for $\hat{\mu}$ includes an adjustment factor $\alpha(z)-E[\alpha]$ for the estimation of $\hat{h}$. One expects this adjustment factor to correspond to that considered in Section 4; the natural choice for the functional of Assumption 6.1 is $\tilde{\mu}(F) \equiv \int a(z,h(x,F))dF_0(z) = E[a(z,h(x,F))]$, for which the pathwise derivative is the adjustment factor discussed in Section 4.

Unlike the results of Andrews (1989a,b), this result allows the nonparametric estimates to affect the limiting distribution of $\hat{\mu}$, and specifies the way in which they do. Andrews imposes the condition that

(6.5) $\qquad \alpha(z) = 0, \quad \sqrt{n}R(\hat{F},F) = o_p(1).$

Following earlier discussion, one expects that a necessary condition for

(6.5) would be that the pathwise derivative of $E[a(z,h(x,F))]$ is zero.

For an example of Theorem 6.3, consider the functional

$$\mu(F) = E_F[1(\mathcal{A})\{g(x^*,F)-g(x,F)\}], \quad g(x,F) = E_F[w|x], \tag{6.6}$$

where $\mathcal{A}$ is a bounded set of possible values for x and x^* is some function of x. A version of this functional without the $1(\mathcal{A})$ term was considered by Stock (1989), who interpreted it as the effect of a policy shift from x to x^* on the average value of the variable w. The $1(\mathcal{A})$ term is present for technical convenience, but also has the effect of excluding outlying values of x.

An estimator of $\mu(F)$ is given by

$$\hat{\mu} = \textstyle\sum_{i=1}^{n} 1(x_i \in \mathcal{A})[\hat{g}(x_i^*)-w_i]/n, \quad \hat{g}(x) = \textstyle\sum_{j=1}^{n} w_j K((x-x_j)/\sigma)/\sum_{j=1}^{n} K((x-x_j)/\sigma). \tag{6.7}$$

This is somewhat different than Stock's estimator, where w_i was replaced by $\hat{g}(x_i)$. Theorem 2.1 would lead us to expect that both estimators of $\mu(F)$ in equation (6.6).

To calculate the influence function for $\mu(F)$, note that $\mu(F) = E_F[1(\mathcal{A})\{g(x^*,F)-w\}]$, which has the form considered in Section 4. Thus, the pathwise derivative should be $d = 1(\mathcal{A})\{g_0(x^*)-w\} + \alpha(z)$, where $\alpha(z)$ is the pathwise derivative of $E[1(\mathcal{A})\{g(x^*,F)-w\}]$. Let $f_0^*(x)$ denote the density of x^* and $\gamma(x) \equiv f_0^*(x)/f_0(x)$. Note that $E[1(\mathcal{A})g(x^*,F)] = E[1(\mathcal{A})\gamma(x)g(x,F)]$. Thus, by equation (4.7) for $H(x) = \gamma(x)$, $\alpha(z) = 1(\mathcal{A})\gamma(x)[w-g_0(x)]$, and

$$\begin{aligned} \psi(z) &= 1(x\in\mathcal{A})\{g_0(x^*)-w - \gamma(x)[w-g_0(x)]\} - \mu_0 \\ &= 1(x\in\mathcal{A})\{g_0(x^*)-g_0(x) - [\gamma(x)-1][w-g_0(x)]\} - \mu_0 \end{aligned} \tag{6.8}$$

The following Theorem makes use of Assumptions 6.1 and 6.2 in the specification of primitive regularity conditions for this to be the influence function.

Theorem 6.4: Suppose that Assumptions K and X are satisfied with $\ell = k > 2r$ *and* $r \geq \ell > (r+1)/2$, $x^* = \tau(x)$, $f_0^*(x)$ *is continuous differentiable with bounded derivative, the boundary of* $\mathcal{A}$ *has Lebesgue measure zero, there is a bounded, convex, open set* $\mathcal{W}$ *containing the closure of* $\mathcal{A} \cup \tau(\mathcal{A})$ *and* $\epsilon > 0$ *such that* $f_0(x) > \epsilon$ *for* $x \in \mathcal{W}$, *and* $\sigma = \sigma(n)$ *with* $\sigma(n)^{4r}n \longrightarrow \infty$, $\sigma(n)^{2\ell}n \longrightarrow 0$. *Then for* $\hat{f}^*(x) = \sum_{i=1}^n K((x-x_i^*)/\sigma)/\sigma^r n$, $\hat{\gamma}(x) = \hat{f}^*(x)/\hat{f}(x)$, *and* $\hat{d}(z) = 1(x \in \mathcal{A})\{\hat{g}(x^*)-w-\hat{\gamma}(x)[w-\hat{g}(x)]\}$,

$$(6.9) \qquad \sqrt{n}(\hat{\mu}-\mu_0) \xrightarrow{d} N(0, E[\psi^2]), \quad \sum_{i=1}^n \hat{d}(z_i)^2/n - [\sum_{i=1}^n \hat{d}(z_i)/n]^2 \xrightarrow{p} E[\psi^2].$$

This result differs from Stock's (1989) in that the estimator is centered at the truth and the convergence in distribution is unconditional rather than conditional on the x_i observations. The unconditional nature of the result explains the presence of the term $\text{Var}(g_0(x^*)-g_0(x))$ in the asymptotic variance $E[\psi^2] = \text{Var}(g_0(x^*)-g_0(x)) + E[(\gamma(x)-1)^2\text{Var}(w|x)]$, which is not present in Stock's (1989) variance measure; here it accounts for the variability of $\sum_{i=1}^n [g_0(x_i^*)-g_0(x_i)]/n$ as an estimator of μ_0.

In Section 4 results for sample averages of nonparametric estimates were used to derive results for semiparametric m-estimators. Similar reasoning allows one to use Assumptions 6.1 and 6.2 in the specification of regularity conditions for semiparametric m-estimators.

Theorem 6.5: Suppose that i) $\hat{\mu} = \mu_0 + o_p(1)$; ii) $m(z,\mu,\hat{h}(x,\mu))$ *is continuously differentiable in* μ; iii) *for any* $\bar{\mu} = \mu_0 + o_p(1)$, $\hat{M}(\bar{\mu}) \equiv \sum_{i=1}^{n}\partial m(z_i,\bar{\mu},\hat{h}(x_i,\bar{\mu}))/\partial\mu/n = M + o_p(1)$ *and* M *is nonsingular;* iv) *the hypotheses of Theorem 6.2 are satisfied for* $a(z,h) = m(z,\mu_0,h)$, $\hat{h}(x) = \hat{h}(x,\mu_0)$, *and* $\alpha(z) = \alpha_m(z)$. *Then*

$$\sqrt{n}(\hat{\mu}-\mu_0) = -M^{-1}\sum_{i=1}^{n}\{m(z_i,\mu_0,h_0(x_i)) + \alpha_m(z_i)-E[\alpha_m]\}/\sqrt{n} + o_p(1). \tag{6.10}$$

In a particular model, Assumptions 6.1 and 6.2 may not provide the most convenient approach to specification of regularity conditions. Specific models and estimation methods can lead to simple arguments that do not require the generality of Assumptions 6.1 and 6.2. An example is Newey (1989c), which makes use of in sample properties of series nonparametric regression estimators. Nevertheless, the pathwise derivative formula should prove useful, even in such specific circumstances, since it gives the form of the influence function, a convenient starting point for showing that the estimator is asymptotically equivalent to a sample average.

7. Conclusion

This paper has considered a pathwise derivative formula for the influence function of a semiparametric estimator. For several examples of interest this result yields the form a correction factor for the presence of nonparametric estimates. Examples of the use of this formula for asymptotic efficiency comparisons and the specification of regularity conditions were given.

The pathwise derivative could be used to compute the asymptotic variance of other types of estimators than those considered here, such as those that maximize some objective function. Also, one could derive the form of adjustment factors for nonparametric estimators other than density and regression derivatives. An important class of estimators not considered here are those with partially observed data, such as censored regression. The semiparametric efficiency literature suggests that it may be somewhat harder to derive the pathwise derivative for in such models, although results such as those of Ritov and Wellner (1988) should prove useful.

The generality of the regularity conditions of Section 6 deserves further investigation. As currently formulated they seem best suited to kernel regression and density estimators, which have natural smoothed empirical measure interpretations. It would be useful to know if they could be applied or generalized in order to encompass other types of nonparametric estimators. This research is currently under way.

Appendix A: Proofs of Theorems

Throughout the appendix C will denote a generic constant that will be different in different uses.

Proof of Lemma 2.1: Consider an LDGP with parameter θ_n and let $E_n[\circ]$ denote the expectation taken at θ_n. Since LDGP's for regular parametric subfamilies are contiguous to the process with $\theta_n = \theta_0$, $\sqrt{n}(\hat{\mu}-\mu_0) = \sum_{i=1}^n \psi_i/\sqrt{n} + o_p(1)$, also holds under the LDGP, where $\psi_i = \psi(z_i)$. Then by addition of appropriate terms,

$$\text{(A.1)} \qquad \sqrt{n}(\hat{\mu}-\mu_n) = \sum_{i=1}^n (\psi_i - E_n[\psi])/\sqrt{n} + \sqrt{n}(\mu_0-\mu_n) + \sqrt{n}E_n[\psi] + o_p(1),$$

where $\mu_n = \mu(\theta_n)$. Let $f(\theta)$ denote the likelihood for a single observation, where the z argument is suppressed for notational convenience, and let $f_n \equiv f(\theta_n)$ and $f_0 \equiv f(\theta_0)$. By regularity $f_n \xrightarrow{a.s.} f_0$, so that for $K_n \to \infty$, note that $1(\|\psi\| \geq K_n)\|\psi\|^2 f_n \xrightarrow{a.s.} 0$. Also, $1(\|\psi\| \geq K_n)\|\psi\|^2 f_n \leq \|\psi\|^2 f_n \xrightarrow{a.s.} \|\psi\|^2 f_0$ and by the continuity hypothesis, $\int \|\psi\|^2 f_n dz$ converges to $\int \|\psi\|^2 f_0 dz$. Then by the dominated convergence theorem of Pitman (1978),

$$\text{(A.2)} \qquad \int 1(\|\psi\| \geq K_n)\|\psi\|^2 f_n dz \to 0.$$

It follows similarly that $Var_n(\psi) = E_n[\psi\psi'] - E_n[\psi]E_n[\psi'] \to E[\psi\psi']$, so that by eq. (A.2) the Lindbergh-Feller conditions are satisfied and

$$\text{(A.3)} \qquad \sum_{i=1}^n (\psi_i - E_n[\psi])/\sqrt{n} \xrightarrow{d} N(0, E[\psi\psi']).$$

By continuity, $\int \|\psi\|^2 \ell(\theta) dz$ is bounded on a neighborhood of θ_0, so that by Lemma 7.2 of Ibragimov and Hasminskii (1981), $\int \psi \ell(\theta) dz$ is differentiable with derivative at θ_0 equal to $E[\psi S_\theta']$. Thus,

$$\sqrt{n}E_n[\psi] = \sqrt{n}\{E[\psi] + E[\psi S_\theta'](\theta_n-\theta_0) + o(\|\theta_n-\theta_0\|)\} = E[\psi S_\theta']\sqrt{n}(\theta_n-\theta_0) + o(1).$$

Also, by $\mu(\theta)$ differentiable,

$$\text{(A.4)} \qquad \sqrt{n}(\mu-\mu_n) = \sqrt{n}\{[-\partial\mu(\theta_0)/\partial\theta](\theta_n-\theta_0) + o(\|\theta_n-\theta_0\|)\}$$

$$= [-\partial\mu(\theta_0)/\partial\theta]\sqrt{n}(\theta_n-\theta_0) + o(1).$$

Since $\sqrt{n}(\hat{\mu}-\mu_0) \xrightarrow{d} N(0,E[\psi\psi'])$ for $\theta_n = \theta_0$, it follows from (A.1), (A.3), and (A.4) that the limiting distribution of $\sqrt{n}(\hat{\mu}-\mu(\theta_n))$ exists and does not depend on the sequence $\{\theta_n\}$ if and only if

$$\text{(A.5)} \qquad \{\partial\mu(\theta_0)/\partial\theta - E[\psi S_\theta']\}\sqrt{n}(\theta_n-\theta_0) = o(1).$$

This equation holds for all sequences such that $\sqrt{n}(\theta_n-\theta_0)$ is bounded if and only if $\partial\mu(\theta_0)/\partial\theta - E[\psi S_\theta'] = 0$. ■

Proof of Theorem 2.1: By Lemma 2.1, the definition of differentiability, and $\tilde{\mathcal{S}} \supset \mathcal{S}$,

$$\text{(A.6)} \qquad E[(\psi-d+E[d])S_\theta'] = E[\psi S_\theta'] - E[dS_\theta'] = \partial\mu(\theta_0)/\partial\theta - \partial\mu(\theta_0)/\partial\theta = 0.$$

Note that $\psi-d+E[d] \in \mathcal{S}$. Consider sequences A_j, $S_{\theta j}$ such that $A_jS_{\theta j}$ converges in mean square to $\psi-d+E[d]$. It then follows that

$$\text{(A.7)} \qquad E[\|\psi-d+E[d]\|^2] = \lim_{j\to\infty}E[(\psi-d+E[d])'A_jS_{\theta j}] = 0. \qquad ■$$

Proof of Lemma 3.1: In the hypotheses of Lemma B.2, $\theta = \eta$ and $\Delta(z,\theta) = 1+\theta'\{s(z)-E[s(z)]\}$. Note that by $s(z)$ bounded and $f_0(z)$ independent of any parameters, the remainder of the hypotheses of Lemma B.2 are satisfied with $A(z) = C$. The first conclusion follows by the conclusion of Lemma B.2.

The second conclusion is a trivial consequence of $E_\theta[\psi'\psi] = E[\psi'\psi] + E[\psi'\psi\{s(z)-E[s(z)]\}']\theta$, where the finiteness of the second expectation follows by $s(z)$ bounded. Let $|\circ|_2 = \{E[(\circ)^2]\}^{1/2}$. The last conclusion follows by Lemma A2 of Chamberlain (1986a): That result implies that for any > 0 there exists $s(z)$ in $C_c^\infty(\mathbb{R}^q)$ with $(E[\|\delta-s\|^2])^{1/2} < \epsilon/2$, so that by the triangle and Cauchy-Schwarz inequalities, $|\|\delta-\{s-E[s]\}\||_2 \le |\|\delta-s\||_2 + |\|E[\delta-s]\||_2 \le \epsilon + |\|\delta-s\||_2 \le \epsilon$. ■

Proof of Theorem 4.1: For $s_1(x) \in C_c^\infty(\mathbb{R}^r)$, $s_2(z) \in C_c^\infty(\mathbb{R}^q)$, let $\tilde{s}_1(x) = s_1(x)-E[s_1]$, $\tilde{s}_2(z) = s_2(z)-E[s_2|x]$, and consider

(A.8) $$f(z,\theta_1,\theta_2) = f_0(z)[1+\theta_2\tilde{s}_2(z)][1+\theta_1\tilde{s}_1(x)], \quad \theta_0 = 0.$$

Note that $\tilde{s}_1$ and $\tilde{s}_2$ are bounded, $E[\tilde{s}_1] = E[\tilde{s}_2] = 0$, and $E[\tilde{s}_1\tilde{s}_2] = E[\tilde{s}_1E[\tilde{s}_2|x]] = 0$. It then follows analogously to the proof of Theorem 3.1 that this parametric subfamily is smooth, with $S_\theta = (\tilde{s}_1,\tilde{s}_2)'$. Also, by Assumption 4.2 iii), $E[\psi'\psi] < \infty$, so that $E_\theta[\psi'\psi]$ is continuous. Consider any $\delta \in \mathcal{S}$, and let $\delta_x = E[\delta|x]$, $\delta_z = \delta-\delta_x$. Let $\epsilon > 0$. Since $E[\delta_x] = 0$, it follows by Lemma A2 of Chamberlain (1986a) that there exists s_1 such that $|\delta_x-\tilde{s}_1|_2 < \epsilon/3$, and that there exists s_2 such that $|\delta_z-s_2|_2 < \epsilon/3$. Then $|\delta_z-\tilde{s}_2|_2 = |\delta_z-s_2-E[\delta_z-s_2|x]|_2 \le \epsilon/3 + |E[\delta_z-s_2|x]|_2 \le 2\epsilon/3$. Thus, $|\delta-(\tilde{s}_1+\tilde{s}_2)|_2 \le |\delta_x-\tilde{s}_1|_2 + |\delta_z-\tilde{s}_2|_2 < \epsilon$, implying $\tilde{\mathcal{S}} = \mathcal{S}$ by ϵ arbitrary. It now suffices to show differentiability of $\mu(\theta)$ and to verify the formula for $\alpha(z)$.

Integrating with respect to the components $\tilde{z}$ of z other than x gives $f(x,\theta) = [1+\theta_1\tilde{s}_1(x)]f_0(x)\{1+\theta_2E[\tilde{s}_2|x]\} = f_0(x)[1+\theta_1\tilde{s}_1(x)]$. Then by the definition of $\tilde{s}_1$, $\mathcal{D}^\lambda f(x,\theta) = \mathcal{D}^\lambda f_0(x) + \theta_1\mathcal{D}^\lambda[f_0(x)\tilde{s}_1(x)]$, so that

(A.9) $$\mu(\theta) = E_\theta[a(z,\mathcal{D}^\lambda f_0(x) + \theta_1\mathcal{D}^\lambda[f_0(x)\tilde{s}_1(x)])].$$

It then follows by Assumption 4.2, continuity of $E_\theta[A(z)]$ (by $\tilde{s}_1$ and $\tilde{s}_2$ bounded), and Lemma B.1 that $\mu(\theta)$ is differentiable at 0 with derivative

$$\text{(A.10)} \qquad \partial\mu(0)/\partial\theta = E[aS_\theta] + (E[a_h \mathcal{D}^\lambda\{f_0(x)\tilde{s}_1(x)\}],0)'$$

$$= E[aS_\theta] + (E[H(x)\mathcal{D}^\lambda\{f_0(x)\tilde{s}_1(x)\}],0), \quad H(x) = E[a_h|x],$$

where $a_h = \partial a(z,h_0(x))/\partial h$. Note that $\tilde{s}_1(x)$ and all its derivatives are zero outside a compact set, which can be chosen to be a cube without loss of generality. It follows that $f_0(x)\tilde{s}_1(x)$ and each of all its derivatives to order $|\lambda|$ are zero outside a cube. Then repeated integration by parts gives

$$\text{(A.11)} \qquad E[H(x)\mathcal{D}^\lambda\{f_0(x)\tilde{s}_1(x)\}] = \int\{H(x)f_0(x)\}\mathcal{D}^\lambda\{f_0(x)\tilde{s}_1(x)\}dx$$

$$= \int(-1)^{|\lambda|}\mathcal{D}^\lambda\{H(x)f_0(x)\}f_0(x)\tilde{s}_1(x)dx = E[\alpha\tilde{s}_1].$$

Noting that $E[\alpha\tilde{s}_2] = 0$, $\partial\mu(0)/\partial\theta = E[(a+\alpha)S_\theta] = 0$. ∎

Proof of Theorem 4.2: For $s_1(x) \in C_c^\infty(\mathbb{R}^r)$, $s_2(w,x) \in C_c^\infty(\mathbb{R}^{r+1})$, $s_3(z) \in C_c^\infty(\mathbb{R}^q)$, let $\tilde{s}_1(x) = s_1(x)-E[s_1]$, $\tilde{s}_2(w,x) = s_2(w,x)-E[s_2|x]$, $\tilde{s}_3(z) = s_3(z)-E[s_3|w,x]$, and

$$\text{(A.12)} \qquad f(z,\theta_1,\theta_2,\theta_3) = f_0(z)][1+\theta_3\tilde{s}_3(z)][1+\theta_2\tilde{s}_2(w,x)][1+\theta_1\tilde{s}_1(x)], \quad \theta_0 = 0.$$

It follows as in the proof of Theorem 4.1 that this is a smooth parametric subfamily, with $S_\theta = (\tilde{s}_1,\tilde{s}_2,\tilde{s}_3)'$, and $\tilde{\mathcal{S}} = \mathcal{S}$. It now suffices to show differentiability of $\mu(\theta)$ and to verify the formula for $\alpha(z)$.

As in the proof of Theorem 4.1, the marginal densities of (w,x) and x respectively are $f_0(w,x)[1+\theta_2\tilde{s}_2(w,x)][1+\theta_1\tilde{s}_1(x)]$ and $f_0(x)[1+\theta_1\tilde{s}_1(x)]$. Thus, $g(x,\theta) = E_\theta[w|x] = E[w\{1+\theta_2\tilde{s}_2(w,x)\}|x] = g_0(x) + \theta_2\Delta(x)$, $\Delta(x) = E[w\tilde{s}_2|x]$. By Assumption 4.3 and repeated application of Corollary 5.8 of

Bartle (1966), it follows that $E[s_2(w,x)|x]$ is continuously differentiable to order $|\lambda|$ and is zero outside a compact set. It follows similarly that $\Delta(x)$ has the same properties. Thus,

$$\text{(A.13)} \qquad \mu(\theta) = E_\theta[a(z, \mathcal{D}^\lambda g(x) + \theta_1 \mathcal{D}^\lambda \Delta(x))].$$

By $\mathcal{D}^\lambda \Delta(x)$ bounded, for any $C > 0$ there is an open set containing zero such that $|\theta_1 \mathcal{D}^\lambda \Delta(x)| \le C$ for all θ_1 in this neighborhood. Then by Assumption 4.3 and Lemma B.1, $\mu(\theta)$ is differentiable at 0 with derivative

$$\text{(A.14)} \qquad \partial\mu(0)/\partial\theta = E[aS_\theta] + (E[a_h \mathcal{D}^\lambda \Delta(x)],\ 0)'$$

$$= E[aS_\theta] + (E[H(x)\mathcal{D}^\lambda \Delta(x)],\ 0)', \quad H(x) = E[a_h|x].$$

Note that $\Delta(x) = E[w\tilde{s}_2|x] = E[(w-g(x))\tilde{s}_2|x]$. It then follows from integration by parts as in the proof of Theorem 4.1 that

$$\text{(A.15)} \qquad E[H(x)\mathcal{D}^\lambda \Delta(x)] = \int\{H(x)f(x)\}\mathcal{D}^\lambda \Delta(x)dx = \int(-1)^{|\lambda|}\mathcal{D}^\lambda\{H(x)f(x)\}\Delta(x)dx$$

$$= E[(-1)^{|\lambda|}\mathcal{D}^\lambda\{H(x)f(x)\}\Delta(x)/f(x)] = E[\alpha\tilde{s}_2].$$

Also, by $E[\alpha(z)|x] = 0$, $E[\alpha\tilde{s}_1] = 0$, and by $\alpha(z)$ measurable with respect to w and x, $E[\alpha\tilde{s}_3] = 0$, so that $\partial\mu(0)/\partial\theta = E[(a+\alpha)S_\theta] = 0$. ∎

Proof of Theorem 6.1: With probability approaching one,

$$\text{(A.16)} \qquad \sqrt{n}|\hat{\mu}-\mu_0 - \textstyle\sum_{i=1}^n\{d(z_i)-E[d]\}/n|$$

$$\le \sqrt{n}|\tilde{\mu}(\hat{F})-\mu_0 - \textstyle\int d(z)d(\hat{F}-F_0)| + \sqrt{n}|\int d(z)d\hat{F} - \sum_{i=1}^n d(z_i)/n|$$

$$\le \sqrt{n}R(\hat{F},F_0) + o_p(1) = o_p(1). \qquad ∎$$

Proof of Theorem 6.2: The proof proceeds by verifying Assumption 6.1. Take

$\mathcal{A}$ = [a,b], note $\inf_{[a,b]}f_0(x) > 0$ by continuity and compactness, and take $\mathcal{F}$ as specified in Lemma B.3. Let $\hat{F}$ be absolutely continuous with density $\hat{f}(z) = \sum_{i=1}^{n}\mathcal{K}((z-z_i)/\sigma)/\sigma^2$, and note that the marginal density of x for this distribution is $\hat{f}(x)$. By Lemma B.4 with, $r = 1$, $\lambda = 0$ and by $1/\sqrt{n}\sigma = (n\sigma^2)^{-1/2} \le (n\sigma^4)^{-1/2} = o(1)$ and $\sigma = o(1)$, $\hat{f}(x)$ converges uniformly to $f_0(x)$, implying that $\hat{F} \in \mathcal{F}$ with probability approaching one. Since for $\hat{F} \in \mathcal{F}$, $1(a\le x\le b)E_{\hat{F}}[w|x] = 1(a\le x\le b)\hat{g}(x)$, condition i) follows. Condition ii) follows by Lemma B.3 with $\omega(x) = 1/f_0(x)$, $\mathcal{A} = [a,b]$, and $R(F,F_0)$ in equation (B.1). Also, note that $n^{1/4}/[\sqrt{n}\sigma] = 1/[n^{1/4}\sigma] = (n\sigma^4)^{-1/4} = o(1)$ and $n^{1/4}\sigma^3 = (n\sigma^{12})^{1/4} \le (n\sigma^6)^{1/4} = o(1)$. Then from Lemma B.4 with $\ell = 3$ it follows that

$$\text{(A.17)} \qquad n^{1/4}\sup_{x\in\mathcal{A}}|\hat{f}(x)-f_0(x)| = o_p(1) \qquad n^{1/4}\sup_{x\in\mathcal{A}}|\hat{g}(x)-g_0(x)| = o_p(1).$$

Then condition iii) follows by Lemma B.3. Also, by Lemma B.6 with $\omega(x) = 1/f_0(x)$ and $\mathcal{A} = [a,b]$ it follows by $\sqrt{n}\sigma^3 = (n\sigma^6)^{1/2} = o(1)$ that condition iv) is satisfied, implying $\sqrt{n}(\hat{\mu}-\mu_0) = \sum_{i=1}^{n}\psi(z_i)/\sqrt{n} + o_p(1)$ by Theorem 6.1. The first conclusion then follows by the Lindbergh Levy central limit theorem. For the second conclusion, note that by $f_0(x)$ is bounded below on $\mathcal{A}$ = [a,b], so that by eq. (A.17) and $E[w_i^2] < \infty$.

$$\sum_{i=1}^{n}|\hat{\psi}_i-\psi_i|^2/n \le C[\inf_{\mathcal{A}}\hat{f}(x)]^{-2}[\sup_{\mathcal{A}}|\hat{g}(x)-g_0(x)|^2$$
$$+ \sup_{\mathcal{A}}|\hat{f}(x)-f_0(x)|^2\sum_{i=1}^{n}|w_i-g_0(x_i)|^2/n] = o_p(1).$$

The conclusion then follows similarly to the consistency argument for the asymptotic variance estimator in Powell, Stock, and Stoker (1989). ∎

Proof of Theorem 6.3: With probability approaching one, (A.18) $\sqrt{n}|\hat{\mu}-\mu_0 - \sum_{i=1}^n\{a_0(z_i)-E[a_0]+\alpha(z_i)-E[\alpha]\}/n| \le T_1 + T_2$

$$T_1 \equiv \sqrt{n}|\sum_{i=1}^n\{a(z_i,\hat{h}(x_i))-a_0(z_i)-\int a(z,\hat{h}(x))dF_0+E[a_0]\}/n|,$$

$$T_2 \equiv \sqrt{n}|\int a(z,\hat{h}(x))dF_0-E[a_0] - \sum_{i=1}^n\{\alpha(z_i)-E[\alpha]\}/n|$$

$$= \sqrt{n}|\hat{\mu}^{\alpha}-\mu_0 - \sum_{i=1}^n\{\alpha(z_i)-E[\alpha]\}/n|.$$

It follows as in Andrews (1989a) that $T_1 = o_p(1)$, while $T_2 = o_p(1)$ follows from Theorem 6.1 (compare eq. (A.16)), giving the conclusion. ■

Proof of Theorem 6.4: The proof proceeds by verifying the hypotheses of Theorem 6.3 for $\hat{\mu} = \sum_{i=1}^n 1(x_i\in\mathcal{A})\hat{g}(x_i^*)/n$ and $\mu_0 = E[1(x\in\mathcal{A})g_0(x^*)]$. Consider $\hat{\mu}^{\alpha} = \int 1(x\in\mathcal{A})\gamma(x)\hat{g}(x)f_0(x)dx$. Let $\tilde{\mu}(F) = E[1(x\in\mathcal{A})\gamma(x)g_0(x)]$ and let $\mathcal{F}$ be as specified in Lemma B.3, noting that $\inf_{\mathcal{A}}f_0(x) > 0$ by $\mathcal{A} \subseteq \mathcal{W}$. Specifying $\hat{F}$ as in the proof of Theorem 6.2, it follows analogously to the arguments given there that by $1/\sqrt{n}\sigma^r = 1/\{n\sigma^{2r}\}^{1/2} = o(1)$ that $\sup_x|\hat{f}(x)-f_0(x)| = o_p(1)$, and Assumption 6.1 i) is satisfied. Also, it follows analogously to the proof of Theorem 6.2 that Assumption 6.1 ii) is satisfied by Lemma B.3, with $\omega(x) = \gamma(x)$ and $R(F,F_0)$ given in eq. (6.1). Also, for large n, $n^{1/4}/[\sqrt{n}\sigma^r] = (n\sigma^{4r})^{-1/4} = o(1)$ and $n^{1/4}\sigma^{\ell} = (n\sigma^{4\ell})^{1/4} \le (n\sigma^{2\ell})^{1/4} = o(1)$. Then eq. (A.17) holds by Lemma B.4, so that Assumption 6.1 iii) follows by Lemma B.3. Also, by Lemma B.6 with $\omega(x) = \gamma(x)$ and $\mathcal{A} = [a,b]$ it follows by $\sqrt{n}\sigma^{\ell} = (n\sigma^{2\ell})^{1/2} = o(1)$ that condition iv) is satisfied. Thus Assumption 6.1 is satisfied for $\hat{\mu}^{\alpha}$ and $\alpha(z) = 1(x\in\mathcal{A})\gamma(x)[w-g_0(x)]$.

To check Assumption 6.2, let $\mathcal{H}$ be the set of functions with domain $\mathcal{W}$ such that for all λ with $|\lambda| \le \ell$, $\sup_{\mathcal{W}}|D^{\lambda}h(x)| \le \sup_{\mathcal{W}}|D^{\lambda}g_0(x)|+1$. Note that for each such λ, $g_0(x)$ is differentiable to order $|\lambda| + 1$, that for large n, $1/\sqrt{n}\sigma^{r+|\lambda|} = (n\sigma^{2r+2|\lambda|})^{-1/2} \leq (n\sigma^{4r})^{-1/2} = o(1)$ and $\sigma^{\ell-|\lambda|} = o(1)$, so that by Lemma B.4,

$$\sup_{\mathcal{W}}|\mathcal{D}^{\lambda}\hat{g}(x)-\mathcal{D}^{\lambda}g_0(x)| = o_p(1), \quad |\lambda| \leq \ell.$$

Thus, $\hat{g}(x) \in \mathcal{H}$ with probability approaching one, so that Assumption 6.2 i) is satisfied for $\hat{h}(x) \equiv \hat{g}(x)$ and $h_0(x) = g_0(x)$. Let $a(z,h) = 1(x\in\mathcal{A})h(x)$ For h_1, $h_2 \in \mathcal{H}$, let $\rho_{\mathcal{H}}(h_1,h_2) = \{\int_{\mathcal{W}}[a(x,h_1) - a(x,h_2)]^2dx\}^{1/2} \leq C\sup_{\mathcal{W}}|h_1(x)-h_2(x)|$. It follows from Lemma B.4 that $\rho_{\mathcal{H}}(\hat{h},h_0) \leq C\sup_{\mathcal{W}}|\hat{g}(x)-g_0(x)| = o_p(1)$, giving Assumption 6.2 ii). Assumption 6.2 iii) then follows by Theorem II.7 of Andrews (1989b). It now follows from Theorem 6.3 that for $1_i = 1(x_i\in\mathcal{A})$

$$\sum_{i=1}^{n}\{1_i\hat{g}(x_i^*) - E[1_i g_0(x_i^*)]\}/\sqrt{n}$$

$$= \sum_{i=1}^{n}\{1_i[g(x_i^*) + \gamma(x)\{w_i-g_0(x_i)\}] - E[1_i g_0(x_i^*)]\}/\sqrt{n} + o_p(1).$$

The first conclusion then follows by subtracting the term $\sum_{i=1}^{n}(1_i w_i-E[1_i w_i])/\sqrt{n} = \sum_{i=1}^{n}\{1_i[w_i-g_0(x_i)\} + g_0(x_i)-E[1_i w_i]\}/\sqrt{n}$ and by the Lindbergh-Levy central limit theorem. To show the second conclusion, note that $E[\psi^2] = Var(d)$ for $d(z) = 1(x\in\mathcal{A})\{g_0(x^*)-w-\gamma(x)[w-g_0(x)]\}$. Also, by Lemma B.4, $\sup_x|\hat{f}^*(x)-f^0(x)| = o_p(1)$, so that $\sup_{x\in\mathcal{A}}|\hat{\gamma}(x)-\gamma(x)| \leq [(\sup_x|\hat{f}^*(x)-f_0^*(x)|) + C(\sup_x|\hat{f}(x)-f_0(x)|)]/\inf_{x\in\mathcal{A}}|\hat{f}(x)| = o_p(1)$. Noting that $\gamma(x)$ is bounded on $\mathcal{A}$, it follows that

$$\sum_{i=1}^{n}|\hat{d}(z_i)-d(z_i)|^2/n \leq \sup_{x\in\tau(\mathcal{A})}|\hat{g}(x)-g_0(x)|^2$$

$$+ \sup_{x\in\mathcal{A}}|\hat{\gamma}(x)|^2\sup_{x\in\mathcal{A}}|\hat{g}(x)-g_0(x)|^2$$

$$+ \sup_{x\in\mathcal{A}}|\hat{\gamma}(x)-\gamma(x)|^2[\sum_{i=1}^{n}|w_i-g_0(x_i)|^2/n] = o_p(1).$$

The conclusion then follows similarly to the consistency argument for the asymptotic variance estimator in Powell, Stock, and Stoker (1989). ∎

Proof of Theorem 6.5: Follows the conclusion of Theorem 6.2 for $\hat{\mu} = \sum_{i=1}^{n} m(z_i, \mu_0, \hat{h}(x_i))/\sqrt{n}$ from a standard Taylor expansion argument; see equation (4.16). ∎

Appendix B: Useful Lemmas

It is helpful to define smoothness in the root-likelihood/mean-square sense. Following Ibragimov and Hasminskii (1981, Ch. 7), suppose that $\mathcal{P}_\theta = \{f(z|\theta) : \theta \in \Theta\}$ is a family of densities $f(z|\theta)$ with respect to some carrier measure, and let dz denote integration with respect to that measure.

Definition A.1: $\mathcal{P}_\theta$ is *smooth* if Θ is open and i) $f(z|\theta)$ is continuous on Θ a.s. z; ii) $f(z|\theta)^{1/2}$ is m.s. differentiable with respect to θ on Θ with derivative $D(z,\theta)$, i.e. $\int\|D(z,\theta)\|^2 dz$ is finite on Θ and for each θ and $\theta_i \to \theta$, $\int[f(z|\theta_i)^{1/2} - f(z|\theta)^{1/2} - D(z,\theta)'(\theta_i-\theta)]^2 dz/\|\theta_i-\theta\|^2 \to 0$; iii) $D(z,\theta)$ is m.s. continuous. Also, for smooth $\mathcal{P}_\theta$ the score is defined by $S_\theta \equiv 2\circ 1(f(z|\theta)>0)D(z,\theta)/f(z|\theta)^{1/2}$ and the information matrix by $\int S_\theta S_\theta' f(z|\theta)dz$. $\mathcal{P}_\theta$ is *regular* if it is smooth and the information matrix is nonsingular on Θ.

See, for example, Ibragimov and Hasminskii (1981) for further details.
The following two lemmas are useful for the proof of Theorems 4.1 and 4.2. They are proved in Newey (1989d).

Lemma B.1: *Suppose that* $\theta = (\theta_1', \theta_2')'$ *such that i)* $f(z|\theta)$ *is smooth; ii)* $a(z,\theta)$ *is continuous in* θ *at each* θ (a.s. z) *and differentiable in* θ_1 *with derivative that is continuous in* θ *at each* θ; *iii) there exists* $A(z)$ *such that* $\|a(z,\theta)\|^2 \le A(z)$, $\|\partial a(z,\theta)/\partial\theta_1\| \le A(z)$, *and* $E[A(z)|\theta]$ *is continuous in* θ. *Then* $E_\theta[a(z,\theta)]$ *is differentiable in* θ_1 *with continuous derivative* $G(\theta) = E_\theta[a(z,\theta)S_1(z,\theta)'] + E_\theta[\partial a(z,\theta)/\partial\theta_1]$, *where* $S_1(z,\theta)$ *is the score for* θ_1. *Also,* $\|G(\theta)\| \le 2E[b(z)|\theta] + tr[I(\theta)]$, *where* $I(\theta)$ *is the information matrix for* θ.

Lemma B.2: *Suppose* $f(z|\beta)$ *is smooth, with score* S_β *at* β_0. *For* $\theta \equiv (\beta', \eta')'$, *let* $\Delta(z,\theta)$ *be bounded, bounded away from zero, continuously differentiable in an open ball* Θ *containing* $\theta_0 \equiv (\beta_0', 0')'$, *with* $\|\partial\Delta(z,\theta)/\partial\theta\| \le A(z)$ *for* $\theta \in \Theta$, *such that* $\int A(z)^2 \ell(z|\beta)d\mu$ *exists and is continuous on* Θ, $\Delta(z,\beta,0) = 1$, *and* $\int \ell(z|\beta)\Delta(z,\theta)d\mu = 1$. *Then* $f(z|\theta) \equiv f(z|\beta)\Delta(z,\theta)$ *is smooth with score* $S_\theta = (S_\beta', \Delta_\eta')'$ *at* θ_0, *where* $\Delta_\eta = 1(\ell(z|\beta_0)>0)\partial\Delta(z,\theta_0)/\partial\eta$.

The next Lemma works out the form of the remainder in Assumption 6.1 ii) for a functional that includes as special cases the examples of Section 6.

Lemma B.3: *Suppose that for a set* $\mathcal{A}$, *Assumption X is satisfied,* $\omega(x)$ *is bounded on* $\mathcal{A}$, *and* $\inf_{\mathcal{A}} f_0(x) > 0$. *Let* $\mathcal{F}$ *be the set of distributions that are absolutely continuous with respect to* x *with density* $f(x) > \inf_{\mathcal{A}} f_0(x)/2$ *for all* $x \in \mathcal{A}$, $\tilde{\mu}(F) \equiv E[1_{\mathcal{A}}\omega(x)g(x,F)]$, *and* $d(z) = 1_{\mathcal{A}}\omega(x)[w-g_0(x)]$. *Then Assumption 6.1 iii) is satisfied with*

$$\text{(B.1)} \qquad R(F,F_0) = C[\sup_{\mathcal{A}}|g(x,F)-g_0(x)|^2 + \sup_{\mathcal{A}}|f(x)-f_0(x)|^2].$$

Proof: Since $E[d(z)] = 0$,

(B.2) $|\tilde{\mu}(F)-\tilde{\mu}(F_0) - E_F[d(z)]|$

$$= |E[1_{\mathcal{A}}\omega(x)\{g(x,F)-g_0(x)\}] - E_F[1_{\mathcal{A}}\omega(x)\{g(x,F)-g_0(x)\}]|$$

$$= |E[1_{\mathcal{A}}\omega(x)\{g(x,F)-g_0(x)\}\{1 - [f(x)/f_0(x)]\}]|$$

$$\le \sup_{\mathcal{A}}\{|\omega(x)||g(x,F)-g_0(x)||1 - [f(x)/f_0(x)]|$$

$$\le C[\sup_{\mathcal{A}}|g(x,F)-g_0(x)|^2 + \sup_{\mathcal{A}}|f(x)-f_0(x)|^2]. \quad \blacksquare$$

The following Lemma is a generalization of Theorem 2.3.1 of Bierens (1987) that gives uniform convergence rates for derivatives of regression and density functions. It is useful for checking Assumption 6.1 i) and iii) and the stochastic equicontinuity condition for the examples of Section 6. Let $\hat{g}(x)$ be as defined in the text and let $\hat{f}(x) = \sum_{i=1}^{n}K((x-x_i)/\sigma)/n\sigma^r$.

Lemma B.4: Suppose that for some λ *Assumption K and X are satisfied with* $\ell = |\lambda|$, $k = \ell-|\lambda| > 0$, *and* $\sigma = \sigma(n)$ *such that* $\sigma(n)^r\sqrt{n} \longrightarrow \infty$ and $\sigma(n) \longrightarrow 0$. *Then for any* $\epsilon > 0$,

(B.3) $$\sup_{\{x\in\mathbb{R}^r\}}|D^\lambda\hat{f}(x)-D^\lambda f_0(x)| = O_p(1/[\sqrt{n}\sigma^{r+|\lambda|}]) + O_p(\sigma^{\ell-|\lambda|}),$$

$$\sup_{\{x\in\mathbb{R}^r: f_0(x)\ge\epsilon\}}|D^\lambda\hat{g}(x)-D^\lambda g_0(x)| = = O_p(1/[\sqrt{n}\sigma^{r+|\lambda|}]) + O_p(\sigma^{\ell-|\lambda|}).$$

Proof: Let $\hat{h}(x) = \sum_{i=1}^{n}w_iK((x-x_i)/\sigma)/\sigma^r n$, $h_0(x) = g_0(x)f_0(x)$. Note that $D^\lambda\hat{h}(x) = \sum_{i=1}^{n}w_iD^\lambda K((x-x_i)/\sigma)/\sigma^{r+|\lambda|}n$. Following Bierens (1987, p. 114), the Fourier inversion formula gives

(B.4) $$D^\lambda\hat{h}(x) = \sum_{i=1}^{n}w_i\int\exp(-it'[x-x_i]/\sigma)\xi_\lambda(t)dt/[(2\pi)^r\sigma^{r+|\lambda|}n]$$

$$= \int\sum_{i=1}^{n}w_i\exp(it'x_i)\exp(-it'x)\xi_\lambda(\sigma t)dt/[(2\pi)^r\sigma^{|\lambda|}n],$$

so that

$$\text{(B.5)} \qquad E[\sup_x|D^\lambda\hat{h}(x)-E[D^\lambda\hat{h}(x)]|$$

$$\leq (2\pi)^{-r}\sigma^{-|\lambda|}\int E|\sum_{i=1}^n[w_i\exp(it'x_i)-E[w_i\exp(it'x_i)]]\}/n||\xi_\lambda(\sigma t)|dt$$

$$\leq C\sigma^{-|\lambda|}\int\{Var(\sum_{i=1}^n w_i\cos(t'x_i)/n) + Var(\sum_{i=1}^n w_i\sin(t'x_i)/n)\}^{1/2}|\xi_\lambda(\sigma t)|dt$$

$$\leq C\{E[w_i^2]\}^{1/2}/[\sqrt{n}\sigma^{|\lambda|}]\int|\xi_\lambda(\sigma t)|dt \leq C/[\sqrt{n}\sigma^{|\lambda|+r}]\int|\xi_\lambda(t)|dt$$

$$= O(1/[\sqrt{n}\sigma^{|\lambda|+r}]).$$

Also, note that $g_0(x)f_0(x)$ is continuously differentiable to order k, with bounded derivatives. Then by a change of variables, integration by parts, and a Taylor expansion in σ around zero

$$\text{(B.6)} \qquad |E[D^\lambda\hat{h}(x)]-D^\lambda h_0(x)| = |\int g_0(u)f_0(u)[D^\lambda K((x-u)/\sigma)/\sigma^{r+|\lambda|}]du - D^\lambda h_0(x)|$$

$$= |\sigma^{-|\lambda|}\int g_0(x+\sigma u)f_0(x+\sigma u)D^\lambda K(u)du - D^\lambda h_0(x)|$$

$$= |\int D^\lambda[g_0(x+\sigma u)f_0(x+\sigma u)]K(u)du - D^\lambda h_0(x)|$$

$$\leq |\sum_{|\tilde{\lambda}|<\ell-|\lambda|}\zeta(\lambda)D^{\tilde{\lambda}}[D^\lambda[g_0(x)f_0(x)]]\int u^{\tilde{\lambda}}K(u)du|$$

$$+ \sigma^{\ell-|\lambda|}\sum_{|\tilde{\lambda}|=\ell-|\lambda|}\zeta(\lambda)\int|D^{\tilde{\lambda}}[D^\lambda[g_0(x+\bar{\sigma}u)f_0(x+\bar{\sigma}u)]]||K(u)|du$$

$$\leq C\sigma^{\ell-|\lambda|},$$

where the third equality holds by $K(u)$ an even function (meaning that the $(-1)^{|\lambda|}$ integration by parts term can be ignored) and the last inequality by boundedness of $D^\lambda[g_0(x)f_0(x)]$ for $|\lambda| = \ell$. Combining equations (B.5) and (B.6), it follows by the triangle and Markov inequalities that

(B.7) $$\sup_{\{x\in\mathbb{R}^r\}}|\mathcal{D}^{\lambda}\hat{h}(x)-\mathcal{D}^{\lambda}h_0(x)| = O_p(1/[\sqrt{n}\sigma^{r+|\lambda|}]) + O_p(\sigma^{\ell-|\lambda|}).$$

Applying this equation with $w_i \equiv 1$ gives the first line of (B.3). To obtain the second line, note that it follows by varying the value of λ in eq. (B.7) and the first line of (B.3) that the same results hold when λ is of smaller order than that give in the statement of the Lemma. In particular, $\sup_x|\hat{f}(x)-f(x)| = O_p(1/[\sqrt{n}\sigma^r]) + O_p(\sigma^{\ell}) = o_p(1)$, implying $\{x : f(x) > \epsilon\} \subseteq \{x : \hat{f}(x) > \epsilon/2\}$ with probability approaching one. Then since $\mathcal{D}^{\lambda}\hat{g}(x) = \mathcal{D}^{\lambda}[\hat{h}(x)/\hat{f}(x)]$, and the denominator is bounded below by $\epsilon/2$ on the set $\{x : f(x) > \epsilon\}$ with probability approaching one, the second conclusion then follows from equation (B.7) applied to $\hat{h}(x)$ and $\hat{f}(x)$ for each partial derivative on the left-hand side of order up to λ. ∎

The next Lemma is useful in verifying iv) for kernel estimators when $d(z)$ need not be smooth but the density of the data is restricted to be smooth.

Lemma B.5: Suppose that Assumptions K and X are satisfied for $k = \ell$ *and for some* $\alpha_1(z)$, $\alpha_2(x)$, $h(x) \equiv E[\alpha_1|x]$ *is continuously differentiable to order* ℓ, $\alpha_2(x)$ *is continuous a.s. on* $\mathbb{R}^r$, *there is a neighborhood* $\mathcal{N}$ *of* 0 *such that* $E[E[\alpha_1^2|x]\int\sup_{\alpha\in\mathcal{N}}|\alpha_2(x+\sigma u)|^2K(u)^2du] < \infty$ *and* $\int\sup_{\sigma\in\mathcal{N}}|\alpha_2(x)\mathcal{D}^{\lambda}[g(x+\sigma u)f(x+\sigma u)]u^{\lambda}K(u)|dudx < \infty$ *for all* $|\lambda| = k$, *and for all* $|\lambda| < k$, $\int|\alpha_2(x)\mathcal{D}^{\lambda}[g(x)f(x)]|dx < \infty$. *Then for* $\sigma = \sigma(n) \longrightarrow 0$,

(B.8) $$\sum_{i=1}^{n}\alpha_1(z_i)\{\int\alpha_2(x)[K((x-x_i)/\sigma)/\sigma^r]dx - \alpha_2(x_i)\}/\sqrt{n} = o_p(1) + O(\sqrt{n}\sigma^{\ell}).$$

Proof: Let

(B.9) $$\delta(z_i,\sigma) \equiv \alpha_1(z_i)\{\int\alpha_2(x)[K((x-x_i)/\sigma)/\sigma^r]dx - \alpha_2(x_i)\}$$

$$= \alpha_1(z_i)\int[\alpha_2(x_i+\sigma u) - \alpha_2(x_i)]K(u)du,$$

where the second equality follows from the change of variables $u = (x-x_i)/\sigma$ and by $\int K(u)du = 1$. Note that $\alpha_2(x+\sigma u)$ is continuous at $\sigma = 0$ for all u with probability one. Also, $|\alpha_2(x+\sigma u)-\alpha_2(x)|^2 \le 4\sup_{\sigma\in\mathcal{N}}|\alpha_2(x+\sigma u)|^2$. Then the dominated convergence theorem gives

(B.10) $$E[\delta(z_i,\sigma)^2] \le CE[E[\alpha_1(z)^2|x]\int|\alpha_2(x+\sigma u)-\alpha_2(x)|^2K(u)^2du] \longrightarrow 0.$$

Note that $K(x)$ is a k^{th} order kernel by $\mathcal{K}(z)$ a k^{th} order kernel. By a Taylor expansion in σ around $\sigma = 0$, for $\sigma \longrightarrow 0$,

(B.11) $$|E[\delta(z_i,\sigma)]| = |\int g(x)\alpha_2(x+\sigma u)K(u)f(x)dudx - \int g(x)\alpha_2(x)f(x)dx|$$

$$= |\int\alpha_2(x)\{\int[g(x-\sigma u)f(x-\sigma u) - g(x)f(x)]K(u)du\}dx|$$

$$\le |\int\alpha_2(x)\{\int\sum_{1\le|\lambda|<k}\zeta(\lambda)D^\lambda[g(x)f(x)]\sigma^{|\lambda|}u^\lambda K(u)du\}dx|$$

$$+ |\int\alpha_2(x)\{\int\sum_{|\lambda|=k}\zeta(\lambda)D^\lambda[g(x+\bar{\sigma}(\lambda,x,u)u)f(x+\bar{\sigma}(\lambda,x,u)u)]\sigma^k u^\lambda K(u)du\}dx|$$

$$\le C\sigma^k\sum_{|\lambda|=k}\int\sup_{\sigma\in\mathcal{N}}|\alpha_2(x)D^\lambda[g(x+\sigma u)f(x+\sigma u)]u^\lambda K(u)|dudx \le C\sigma^k,$$

where $0 \le \bar{\sigma}(x,u) \le \sigma$ and $\zeta(\lambda)$ denote the Taylor expansion coefficients. Then by i) and the Chebyshev and Cauchy-Schwarz inequalities,

(B.12) $$|\sum_{i=1}^n\delta(z_i,\sigma)/\sqrt{n}| \le |\sum_{i=1}^n\{\delta(z_i,\sigma)-E[\delta(z_i,\sigma)]\}/\sqrt{n}| + \sqrt{n}|E[\delta(z_i,\sigma)]|$$

$$= O_p([Var(\delta(z_i,\sigma))]^{1/2}) + O(\sqrt{n}\sigma^k) = O_p(|\delta(z_i,\sigma)|_2) + O(\sqrt{n}\sigma^k)$$

$$= O_p(o(1)) + O(\sqrt{n}\sigma^k) = o_p(1) + O(\sqrt{n}\sigma^k). \quad \blacksquare$$

The following Lemma specializes the previous Lemma to the Section 6 examples.

Lemma B.6: Suppose that Assumptions K *and* X *are satisfied with* $k = \ell$ *and* $\omega(x)$ *is continuous on* $\mathcal{A}$ *that is bounded with boundary that has Lebesgue measure zero. Then for* $z = (w,x)$ *and* $\alpha(z) = 1(x\in\mathcal{A})\omega(x)[w-g(x)]$, *if* $\sqrt{n}\sigma^{\ell} = o(1)$,

$$\text{(B.13)} \qquad \sum_{i=1}^{n}\{\int\alpha(z)[\mathcal{K}((z-z_i)/\sigma)/\sigma^{r+1}]dz - \alpha(z_i)\}/\sqrt{n} = o_p(1).$$

Proof: The proof proceeds by checking the hypotheses of Lemma B.5. Consider first $\alpha_1(z) = w$, $\alpha_2(x) = 1(x\in\mathcal{A})\omega(x)$. Note that $\alpha_2(x)$ is continuous a.s. by the boundary of $\mathcal{A}$ having measure zero, and that it is bounded and zero outside a compact set. Thus, the first dominance condition of Lemma B5 is satisfied by $E[w^2] < \infty$. The second and third dominance conditions are satisfied by boundedness of $D^{\lambda}[f_0(x)g_0(x)]$. Then by $\int w[\mathcal{K}((z-z_i)/\sigma)/\sigma]dz = \int[\int(w_i+v\sigma)\mathcal{K}(v,(x-x_i)/\sigma)dv]dx = w_i\int K((x-x_i)/\sigma)dx$ and the conclusion of Lemma B5,

$$\sum_{i=1}^{n}\{\int 1(x\in\mathcal{A})\omega(x)w[\mathcal{K}((z-z_i)/\sigma)/\sigma^{r+1}]dz - 1(x_i\in\mathcal{A})\omega(x_i)w_i\}/\sqrt{n}$$

$$= \sum_{i=1}^{n}w_i\{\int 1(x\in\mathcal{A})\omega(x)[K((x-x_i)/\sigma)/\sigma^{r}]dx - 1(x_i\in\mathcal{A})\omega(x_i)\}/\sqrt{n} = o_p(1).$$

It follows similarly for $\alpha_1(z) = 1$ and $\alpha_2(x) = 1(x\in\mathcal{A})\omega(x)g(x)$ that

$$\sum_{i=1}^{n}\{\int 1(x\in\mathcal{A})\omega(x)g_0(x)[\mathcal{K}((z-z_i)/\sigma)/\sigma^{r+1}]dz - 1(x_i\in\mathcal{A})\omega(x_i)g_0(x_i)\}/\sqrt{n}$$

$$= o_p(1),$$

so that the conclusion follows by the triangle inequality. ∎

References

Andrews, D.W.K. (1988): "Asymptotic Normality of Series Estimators for Various Nonparametric and Semiparametric Models," mimeo, Cowles Foundation, Yale University.

Andrews, D.W.K. (1989a): "Asymptotics for Semiparametric Econometric Models: I. Estimation," mimeo, Cowles Foundation, Yale University.

Andrews, D.W.K. (1989b): "Asymptotics for Semiparametric Econometric Models: II. Stochastic Equicontinuity," mimeo, Cowles Foundation, Yale University.

Bartle, R. G. (1966): *The Elements of Integration*, New York: John Wiley and Sons.

Begun, J., W. Hall, W. Huang, and J. Wellner (1983): "Information and Asymptotic Efficiency in Parametric-Nonparametric Models," *Annals of Statistics*, 11, 432-452.

Bickel, P. (1982): "On Adaptive Estimation," *Annals of Statistics*, 10, 647-671.

Bickel P., C.A.J. Klaassen, Y. Ritov, and J.A. Wellner (1989): "Efficient and Adaptive Inference in Semiparametric Models" monograph, Johns Hopkins University Press, forthcoming.

Bierens, H.J. (1987): "Kernel Estimators of Regression Functions," Ch. 3 of Bewley, T.F., ed., *Advances in Econometrics: Fifth World Congress, Vol. I*, Cambridge, England: Cambridge University Press.

Boos, D.D. and R.J. Serfling (1980): "A Note on Differentials and the CLT and LIL for Statistical Functions, with Application to M-Estimates," *Annals of Statistics*, 8, 618-624.

Buckley, J. and I. James (1979): "Linear Regression with Censored Data," *Biometrika*, 66, 429-436.

Carroll, R. J. (1982): "Adapting for Heteroskedasticity in Linear Models," *Annals of Statistics*, 10, 1224-1233.

Chamberlain, G. (1986a): "Asymptotic Efficiency in Semiparametric Models with Censoring," *Journal of Econometrics* 32, 189-218.

Chamberlain, G. (1986b): "Notes on Semiparametric Regression," mimeo, Department of Economics, Harvard University.

Chamberlain, G. (1987): "Efficiency Bounds for Semiparametric Regression," mimeo, Department of Economics, Harvard University.

Engle, R.F., C.W.J. Granger, J. Rice, and A. Weiss (1986): "Semiparametric Estimates of the Relation Between Weather and Electricity Sales." *Journal of the American Statistical Association*, 81, 310-320.

Hampel, F.R. (1974): "The Influence Curve and Its Role in Robust Estimation," *Journal of the American Statistical Association*, 62, 1179-1186.

Heckman, N.E. (1986): "Spline Smoothing in a Partly Linear Model," *Journal of the Royal Statistical Society*, Series B, 48, 244-248.

Hoeffding, W. (1948): "A Class of Statistics with Asymptotically Normal Distribution," *Annals of Mathematical Statistics,* 19, 293-325.

Horowitz, J.L. (1986): "Semiparametric m-estimation of Censored Regression Models," Department of Economics Working Paper 86-14, University of Iowa.

Huber, P., (1967): "The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley: University of California Press.

Ibragimov, I.A. and R.Z. Has'minskii (1981): *Statistical Estimation: Asymptotic Theory*, New York: Springer-Verlag.

Ichimura, H. (1987): "Estimation of Single Index Models," Ph. D. dissertation, Massachusetts Institute of Technology.

Koshevnik and Levit (1976): "On a Non-parametric Analouge of the Information Matrix," *Theory of Probability and Applications*, 21, 738-753.

Luenberger, D.G. (1969): *Optimization by Vector Space Methods*, New York: John Wiley and Sons.

Newey, W.K. (1989a): "Semiparametric Efficiency Bounds," mimeo, Department of Economics, Princeton University.

Newey, W.K. (1989b): "Efficient Estimation of Semiparametric Models Via Moment Restrictions," mimeo, Department of Economics, Princeton University.

Newey, W.K. (1989c): "Series Estimation of Regression Functionals," mimeo, Department of Economics, Princeton University.

Newey, W.K. (1989d): "Efficient Estimation of Tobit Models Under Conditional Symmetry," paper presented at the Fifth International Symposium in Economic Theory and Econometrics, Duke University.

Newey, W.K. and T.M. Stoker (1989): "Efficiency Properties of Average Derivative Estimators," mimeo, Sloan School of Management, MIT.

Newey, W.K. and K.D. West (1987): "A Simple, Positive Semidefinite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, 55, 703-708.

Pfanzagl, J. (1982): *Contributions to a General Asymptotic Statistical Theory*, New York: Springer-Verlag.

Pitman, E.J.G. (1979): *Some Basic Theory for Statistical Inference*, London: Chapman and Hall.

Powell, J.L., J.H. Stock, and T.M. Stoker (1989): "Semiparametric Estimation of Index Coefficients," *Econometrica,* forthcoming.

Rice, J. (1986): "Convergence Rates for Partially Splined Estimates," *Statistics and Probability Letters*, 4, 203-208.

Ritov, Y. (1987): "Estimation in a Linear Regression Model with Censored Data," Technical Report # 114, Department of Statistics, UC Berkeley.

Ritov, Y. and P.J. Bickel (1987): "Achieving Information Bounds in Non and Semiparametric Models," Technical Report No. 116, Department of Statistics, University of California, Berkeley.

Ritov, Y. and J.A. Wellner (1988): "Censoring, Martingales, and the Cox Model," *Contemporary Mathematics*, 80, 191-219.

Robinson, P. (1987): "Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form," *Econometrica*, 55, 875-891.

Robinson, P. (1988): "Root-N-Consistent Semiparametric Regression," *Econometrica,* 56, 931-954.

Schick, A. (1986): "On Asymptotically Efficient Estimation in Semiparametric Models," *Annals of Statistics*, 14, 1139-1151.

Serfling (1980): "Approximation Theorems of Mathematical Statistics," New York: Wiley.

Severini, T.A. and W.H. Wong (1987): "Profile Likelihood and Semiparametric Models," manuscript, University of Chicago.

Stein, C. (1956): "Efficient Nonparametric Testing and Estimation," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, Berkeley: University of California Press.

Stock, J.H. (1989): "Nonparametric Policy Analysis," *Journal of the American Statistical Association,* 84, 567-575.

Stoker, T.M. (1988): "Equivalence of Direct and Indirect Estimators of Average Derivatives," paper presented at the Fifth International Symposium in Economic Theory and Econometrics, Duke University.

Van der Vaart, A. (1988): "On Differentiable Functionals," mimeo, Department of Statistics, University of Washington.

Von Mises (1947): "On the Asymptotic Distributions of Differentiable Statistical Functionals," *Annals of Mathematical Statistics*, 18, 309-348.

White, H. (1984): *Asymptotic Theory for Econometricians,* New York: Academic Press.