ECONOMICS BULLETIN

Simple Edgeworth approximations for semiparametric averaged derivatives

S. C. Goh
University of Toronto

Abstract

This note proposes a computationally simple empirical Edgeworth expansion for the limiting distribution of a Studentized estimator of a semiparametric single index model. The estimator in question is the density−weighted averaged derivative estimator implemented according to the method of Powell, Stock and Stoker (1989). The coefficients of the expansion are derived from the cumulants of a bootstrap estimate of the distribution of the Studentized estimator. Monte Carlo evidence indicates finite−sample performance comparable to that of the empirical Edgeworth expansions proposed by Nishiyama and Robinson (2000).

This note was inspired by a suggestion of Tom Rothenberg. I am indebted to him as well as to Jim Powell for encouragement and support. The research presented in this note was supported by the Department of Economics and the Graduate Division of the University of California, Berkeley and by the University of Toronto Connaught Fund. The usual disclaimer applies.
Citation: Goh, S. C., (2005) "Simple Edgeworth approximations for semiparametric averaged derivatives." *Economics Bulletin*, Vol. 3, No. 50 pp. 1−8
Submitted: November 30, 2005. **Accepted:** November 30, 2005.
URL: http://www.economicsbulletin.com/2005/volume3/EB−05C10020A.pdf

1 Introduction

Much of the literature over the past two decades dealing with the first-order asymptotic theory of semiparametric estimators has focused on estimators that nest an estimated nonparametric component. The slow rate of convergence of the embedded nonparametric estimators would suggest that the finite sample behaviour of the overall semiparametric procedure can be quite different from the attractively quasi-parametric behaviour suggested by first-order asymptotic approximations. Higher-order asymptotic theory, in contrast to finite-sample theory, appears to be both mathematically tractable and capable of shedding light on the finite-sample behaviour of semiparametric estimators.

In this connection Nishiyama and Robinson (2000) proposed Edgeworth expansions for the distribution of density-weighted averaged derivative estimators of a semiparametric single-index model (cf. e.g., Powell et al., 1989) with a view to developing a practical method of improving the accuracy of any resulting inferences.

This note proposes a computationally simpler alternative to the Edgeworth correction proposed by Nishiyama and Robinson (2000) for the Studentized version of the density-weighted averaged derivative estimator. The alternative proposed here is based on cumulants of a bootstrapped estimate of the distribution of the Studentized estimator and is also shown in simulations to furnish excellent smooth approximations to the tail behaviour of the Studentized statistic. The next section of this note provides a brief overview of the density-weighted averaged derivative estimator. The method of Edgeworth correction proposed here is also described. Monte Carlo evidence presented in Section 3 indicates the ability of the simple Edgeworth correction to deliver distributional approximations that equal or exceed in quality those furnished by the first-order standard normal approximation.

2 Edgeworth corrections for the Studentized averaged derivative estimator

2.1 The model

Consider a sequence $(Y_i, X_i^T)^T$, $i = 1, \ldots, n$, where Y_i and X_i are scalar and $k \times 1$ random variates, respectively. Assume that the regression function $r(X) \equiv E[Y|X]$ has the single index form $r(X) = R(X^T\beta)$ for some function $R : \Re \to \Re$

and some k-variate column vector β. The single-index form nests a number of models that have proven useful in empirical practice, include probit, logit and Tobit, as well as various transformation models.

In each of these cases maximum likelihood yields $\sqrt{n}$-consistent, efficient and asymptotically normal estimates of β if the researcher is willing to assume a parametric form for the regression function $R(\cdot)$. Misspecification of $R(\cdot)$ leads in general to inconsistency. On the other hand, regarding $R(\cdot)$ as nonparametric allows only for identification of β up to scale, since

$$\bar{\mu} \equiv -E[r'(X)f(X)] = c\beta, \tag{1}$$

where $c = -E[R'(X^T\beta)f(X)]$ and f denotes the density of X. From integration by parts we have that

$$\bar{\mu} = 2E[Yf'(X)]. \tag{2}$$

Given the sample $(Y_i, X_i^T)^T$, $i = 1, \ldots, n$, the quantity given by (2) is estimated by the density-weighted averaged derivative

$$U = \binom{n}{2}^{-1} \sum_{i<j} U_{ij}, \tag{3}$$

where

$$U_{ij} = (Y_i - Y_j)K'_{ij}, \tag{4}$$

$$K'_{ij} = \frac{1}{h^{k+1}} K' \left(\frac{X_i - X_j}{h} \right), \tag{5}$$

and where $K : \Re^k \to \Re$ is an even and differentiable kernel function with $\int K(u)du = 1$. Here $h > 0$ is a bandwidth that converges to zero as $n \to \infty$.

Nishiyama and Robinson (2000, Theorem 4) proposed a feasible empirical Edgeworth expansion for an arbitrary linear combination of the Studentized vector $\hat{\Sigma}^{-\frac{1}{2}}U$, where $\hat{\Sigma}$ is a jackknife estimate

$$\hat{\Sigma} \equiv \frac{4}{(n-1)(n-2)^2} \sum_{i=1}^{n} \left\{ \sum_{j \neq i}^{n} (U_{ij} - U) \right\} \left\{ \sum_{l \neq i}^{n} (U_{il} - U)^T \right\} \tag{6}$$

of the asymptotic covariance matrix of $\sqrt{n}(U - \bar{\mu})$. In particular, Nishiyama and Robinson (2000, Theorem 4) proposed an expansion of the distribution function of the statistic

$$\hat{Z} \equiv \sqrt{n}(v^T \hat{\Sigma} v)^{-\frac{1}{2}} v^T (U - \bar{\mu}), \tag{7}$$

where v is an arbitrary k-vector.[1] The expansion is to order q_n, where q_n is a sum of $n^{-\frac{1}{2}}$ and some other terms, the order of which depends on the bandwidth h. The remainder term is shown to be of order $o(q_n)$. Monte Carlo evidence indicates that the expansion of Nishiyama and Robinson (2000, Theorem 4) has the potential to perform quite well in approximating the distribution of $\hat{Z}$, subject to good choices of bandwidth and kernel order. The derivation and computation of the correction terms in their expansion is however somewhat complex. The next section of this note proposes a much simpler Edgeworth correction for the distribution of $\hat{Z}$.

2.2 A simple Edgeworth correction for the Studentized statistic

It is proposed that the distribution $F_n(z)$ of the Studentized statistic $\hat{Z}$ given in (7) be approximated by the formula

$$F_n(z) \approx \Phi(z) - \phi(z) \left[\frac{\rho_{3\hat{Z}}(z^2-1)}{6} + \frac{3\rho_{4\hat{Z}}(z^3-3z) + 2\rho_{3\hat{Z}}^2(z^5-10z^3+15z)}{72} \right]. \tag{8}$$

Here $\Phi(\cdot)$ and $\phi(\cdot)$ denote the standard normal distribution and density functions, respectively, and $\rho_{j\hat{Z}}$ ($j \in \{3,4\}$) denotes the jth *standardized* cumulant of $\hat{Z}$. In particular, for $j = 3, 4$,

$$\rho_{j\hat{Z}} = \frac{\kappa_{j\hat{Z}}}{\sigma_{\hat{Z}}^j}, \tag{9}$$

where $\kappa_{j\hat{Z}}$ and $\sigma_{\hat{Z}}$ denote the jth cumulant and standard error, respectively, of $\hat{Z}$.

It is clear that $\rho_{j\hat{Z}}$ for $j = 3, 4$ are unknown; in order to make the approximation given in (8) operational it is proposed that the corresponding standardized cumulants of the bootstrapped empirical distribution of $\hat{Z}$ appear in the corresponding places in (8).

In other words, the researcher would resample her data $(Y_i, X_i^T)^T$, $i = 1, \ldots, n$ randomly with replacement B times, each time computing the bootstrapped version

$$\hat{Z}_b^* \equiv \sqrt{n}(v^T\hat{\Sigma}_b^* v)^{-\frac{1}{2}} v^T(U_b^* - U)$$

of $\hat{Z}$ using the simulated data contained in the bth resample ($b = 1, \ldots, B$). The standardized population cumulants $\rho_{3\hat{Z}}$ and $\rho_{4\hat{Z}}$ would then be estimated using the third and fourth standardized cumulants, respectively, of the uniform discrete distribution with support points $\hat{Z}_b^*$, $b = 1, \ldots, B$.

[1] It should be noted that the validity of a full multivariate expansion for $\hat{Z}$ has not been established and that a Cramér-Wold device would not be applicable for this problem.

3 Numerical Evidence

3.1 Normal Tobit

I first consider the same basic Monte Carlo design as was presented in Nishiyama and Robinson (2000). Simulated observations $(Y_i, X_i^T)^T$, $i = 1, \ldots, n$ were generated using the Tobit model $Y_i = (X_i^T\beta + \epsilon_i) \cdot 1(X_i^T\beta + \epsilon_i \geq 0)$, where $X_i \in \Re^2$ and (X_i^T, ϵ_i) are iid $N(0, I_3)$. With this model, $r(X) = (X^T\beta) \cdot \left[1 - \Phi(-X^T\beta)\right] + \phi(-X^T\beta)$ and $\bar{\mu} = \frac{-1}{8\pi}\beta$. The Monte Carlo design yields comparisons of the finite-sample performance of approximations to the distribution of $\hat{Z}$ given in (7). I set v to be $(1, 0)^T$ and $\beta = (1, 1)^T$. The bootstrapped empirical distribution of $\hat{Z}$ was computed by generating 200 resamples of size n with replacement from each simulated dataset.

The Monte Carlo experiments presented below use a second-order bivariate kernel function derived from the product of two standard normal density functions, i.e., $K(u_1, u_2) = \phi(u_1)\phi(u_2)$. This is in contrast to the case presented in Nishiyama and Robinson (2000), where kernel functions of order $L \in \{4, 8, 10\}$ were used. The second-order kernel avoids any problems arising out of negative values at certain values of its argument and is assumed to correspond to what is more popular in empirical practice.

The experiments presented here consider sample sizes $n \in \{100, 400\}$. Only results for $n = 100$ are presented, as the results for the larger sample size were essentially identical. The number of Monte Carlo replications used was set to 500.

Table 1 allows for an assessment the performance of approximate confidence intervals for $v^T\bar{\mu} = -\frac{1}{8\pi} \approx -.0398$ implied by the $N(0, 1)$ and simple Edgeworth-A approximations proposed in this note. Approximate 80% confidence intervals are compared to the "true" confidence intervals based on the empirical distribution of $\hat{Z}$. Each interval estimate presented in Table 1 was computed using the second-order kernel mentioned above over a range of different bandwidths. The endpoints of the intervals represent averages over 500 Monte Carlo replications of the endpoints of the interval estimates based on the empirical distribution of $\hat{Z}$ and the $N(0, 1)$ and Cornish-Fisher approximations. In particular, note that the interval implied by the $N(0, 1)$ approximation may be corrected by inverting the Edgeworth approximation proposed in Section 2.2, thus yielding a Cornish-Fisher expansion for a given quantile of the sampling distribution of $\hat{Z}$. As such, if w_γ denotes the γ-quantile of the distribution of $\hat{Z}$, its Cornish-Fisher approximation is given by

$$w_\gamma \approx z_\gamma + p_{11}(z_\gamma) + p_{21}(z_\gamma), \tag{10}$$

where z_γ denotes the γ-quantile of a $N(0,1)$ random variate, and where the formulae for $p_{11}(\cdot)$ and $p_{21}(\cdot)$ are as given in Hall (1992, p.88–89). Let w_γ^* denote the estimate of w_γ in (10) with the corresponding standardized cumulants of the bootstrapped empirical distribution appearing in place of $\rho_{3\hat{Z}}$, $\rho_{4\hat{Z}}$ in the expressions for p_{11}, p_{21}. The corresponding confidence interval is given by

$$\left(\hat{Z} - n^{-\frac{1}{2}}\left(v^T\hat{\Sigma}v\right)w^*_{1-\frac{\alpha}{2}},\ \hat{Z} - n^{-\frac{1}{2}}\left(v^T\hat{\Sigma}v\right)w^*_{\frac{\alpha}{2}}\right). \tag{11}$$

The effect of bandwidth choice on the bias and width of the interval estimates of $\bar{\mu}$ is evident from a glance at Table 1. In particular, the larger bandwidths of $h = 0.6, 0.4$ lead to "true" confidence intervals that are narrow but centred to the left of the target value of $\bar{\mu} = -0.0398$, while the smaller bandwidths of $h = 0.3, 0.2$, as expected, lead to true confidence intervals that are approximately centred about $\bar{\mu}$, but are rather wider than is the case when a greater degree of smoothing is applied in the construction of $\hat{Z}$.

In contrast to some of the evidence presented in Nishiyama and Robinson (2000, Figures 24–27), the interval estimates based on the $N(0,1)$ and Cornish-Fisher approximations are virtually identical across each of the simulations that were conducted. This is to be expected, given the normal Tobit design used in this section. The next section presents results of a Monte Carlo experiment indicating the ability of the simple Edgeworth correction to improve upon the distributional approximation yielded by first-order theory.

3.2 A semi-logarithmic model

Here observations $(Y_i, X_i^T)^T$, $i = 1, \ldots, n$ are generated in accordance with the semi-logarithmic model $\log Y_i = X_i^T\beta + \epsilon_i$, where $X_i^T\beta = X_{i1}\beta_1 + X_{i2}\beta_2$. The semiparametric averaged derivative estimator is relevant when the researcher is unwilling to make any assumptions regarding the form of the transformation function $T(\cdot)$, where $T(Y_i) = X_i^T\beta + \epsilon_i$, other than that it is increasing in its argument. The performance of the Studentized statistic under the null hypothesis $H_0 : \beta_1 = 0$ is the focus of this experiment.

In contrast to the previous experiment, X_i is drawn from a standardized χ_3^2 distribution centred about its mean, i.e., $X_i \sim \frac{1}{\sqrt{2\cdot 3}}(\chi_3^2 - 3)$, while the error term is simulated as $\epsilon_i \sim W - \frac{1}{2}$, where W is uniformly distributed on the unit interval. Sample sizes of 100 and 400 were considered, although as the results for $n = 400$ are qualitatively similar, only results for $n = 100$ are presented. The estimator was constructed using the same second-order kernel function used in

Section 3.1, but this time with the bandwidth selected to minimize asymptotic mean squared error.[2] A total of 6500 Monte Carlo simulations were employed to approximate the sampling distribution of the test statistic $\hat{Z}$. The bootstrapped cumulant estimates were generated from 1600 bootstrap resamples of size n with the resampling conducted on the first of the 6500 Monte Carlo samples mentioned previously.

Figure 1 clearly illustrates the ability of the simple Edgeworth expansion to improve on the standard normal approximation with respect to the approximation of tail probabilities under the restriction imposed by the null hypothesis.

References

Hall, P. (1992) *The Bootstrap and Edgeworth Expansion* (New York: Springer-Verlag)

Nishiyama, Y., and P. M. Robinson (2000) 'Edgeworth expansions for semiparametric averaged derivatives.' *Econometrica* 68, 931–979

Powell, J. L., and T. M. Stoker (1996) 'Optimal bandwidth choice for density-weighted averages.' *Journal of Econometrics* 75, 291–316

Powell, J. L., J. H. Stock, and T. M. Stoker (1989) 'Semiparametric estimation of index coefficients.' *Econometrica* 57, 1403–1430

[2] In particular, the bandwidth used in this section is given by

$$h = \left(\frac{(d+2) \cdot 1.724}{L \cdot .397^2} \right)^{\frac{1}{2L+d+2}} n^{\frac{-2}{2L+d+2}},$$

where d denotes the dimension of X_i and L the kernel order. Obviously, $d = L = 2$ in this example. Here 1.724 and $.397^2$ denote estimates of leading constants in polynomial expansions of the variance and squared bias, respectively, of the averaged derivative estimator. Cf. Powell and Stoker (1996, Section 4.4).

Table 1: Confidence intervals for $\bar{\mu} = \frac{-1}{8\pi} \approx -.0398$, $n = 100$

80% Confidence Intervals	
$h = 0.6$	
True	$(-0.0514, -0.0316)$
$N(0,1)$	$(-0.0361, -0.0206)$
Cornish-Fisher	$(-0.0361, -0.0206)$
$h = 0.4$	
True	$(-0.0541, -0.0303)$
$N(0,1)$	$(-0.0461, -0.0223)$
Cornish-Fisher	$(-0.0462, -0.0223)$
$h = 0.3$	
True	$(-0.0568, -0.0246)$
$N(0,1)$	$(-0.0526, -0.0190)$
Cornish-Fisher	$(-0.0526, -0.0191)$
$h = 0.2$	
True	$(-0.0664, -0.0168)$
$N(0,1)$	$(-0.0708, -0.0052)$
Cornish-Fisher	$(-0.0708, -0.0052)$

Figure 1: Upper tail probabilities for $\hat{Z}$ under $H_0 : \beta_1 = 0$