ASYMPTOTIC EQUIVALENCE OF ESTIMATORS OF AVERAGE DERIVATIVES

By Wei Li[1]

Fuqua School of Business
Duke University
Durham, NC 27708
E-mail: Wei.Li@duke.edu

Economic Letter, 241–45, (November 1996).

This paper establishes the asymptotic equivalence among four semiparametric sample analog estimators of average derivatives of a regression function and their corresponding slope or ratio-of-moments estimators.

Keywords: Average derivative estimators, Semiparametric, Asymptotic equivalance.

JEL classification: C14; C20

[1]Thanks to Margarida Genius, Lung-Fei Lee and seminar participants at the University of Michigan for helpful comments.

1 Introduction

The recent pioneering work by Stoker (1986, 1990) and Härdle-Stoker (1989) on the semiparametric estimation of average derivatives of regression functions has provided a powerful set of tools in econometrics. To introduce the ideas, consider the following empirical problem. Let $(y, \mathbf{x})$ denote a $k+1$ random vector, where y is the response variable and $\mathbf{x}$ is a vector of k continuous random variables. Let $f^*(y, \mathbf{x})$ denote the joint density function of $(y, \mathbf{x})$, and $f(\mathbf{x})$ denote the marginal density of $\mathbf{x}$. If the mean regression function of y on $\mathbf{x}$ is given by $g(\mathbf{x}) \equiv \mathrm{E}[y|\mathbf{x}] = G(\mathbf{x})/f(\mathbf{x})$ where $G(\mathbf{x}) = \int y f^*(y, \mathbf{x}) dy$, then the vector of average derivatives is defined as $\beta \equiv \mathrm{E}[g'(\mathbf{x})]$ where $g'(\mathbf{x}) \equiv (\partial g/\partial x^1, \cdots, \partial g/\partial x^k)^T$ is a k-vector of partial derivatives of $g(\mathbf{x})$, and the expectation is taken with respect to the marginal distribution of $\mathbf{x}$. The empirical problem is to estimate β based on an observed random sample $(y_i, \mathbf{x}_i)$ for $i = 1, \cdots, N$.

A natural way to estimate β semiparametrically is to 1) nonparametrically estimate $\hat{g}'(\mathbf{x}_i)$ for $i = 1, \cdots, N$, and 2) form the sample average: $\hat{\beta} = N^{-1} \sum \hat{g}'(\mathbf{x}_i)$. Using this two-step procedure, it is possible to construct four sample analog estimators because β has four equivalent population representations if $f(\mathbf{x})$ vanishes on the boundary of the support of $\mathbf{x}$:

$$\begin{aligned} \beta &= E\left[-g(\mathbf{x})\frac{f'(\mathbf{x})}{f(\mathbf{x})}\right] = \mathrm{E}[g(\mathbf{x})l(\mathbf{x})] \quad \text{(Theorem 1 of Stoker (1990))} && (1) \\ &= \mathrm{E}[yl(\mathbf{x})] \qquad \text{(Law of iterated expectations)} && (2) \\ &= \mathrm{E}[g'(\mathbf{x})] = \mathrm{E}\left[\frac{G'(\mathbf{x})}{f(\mathbf{x})} - g(\mathbf{x})\frac{f'(\mathbf{x})}{f(\mathbf{x})}\right] = \mathrm{E}[r(\mathbf{x}) + g(\mathbf{x})l(\mathbf{x})] && (3) \\ &= \mathrm{E}[r(\mathbf{x}) + yl(\mathbf{x})] \qquad \text{(Law of iterated expectations)} && (4) \end{aligned}$$

where $l(\mathbf{x}) = -f'(\mathbf{x})/f(\mathbf{x})$ and $r(\mathbf{x}) = G'(\mathbf{x})/f(\mathbf{x})$. The Härdle-Stoker's estimator ($\hat{\beta}_2$) is sample analogous to $\mathrm{E}[yl(\mathbf{x})]$ (2), while Stoker's estimator ($\hat{\beta}_3$) is sample analogous to $\mathrm{E}[g'(\mathbf{x})]$ (3). Härdle-Stoker and Stoker show, respectively, that $\sqrt{N}(\hat{\beta}_2 - \beta)$ and $\sqrt{N}(\hat{\beta}_3 - \beta)$ converge to a common limiting normal distribution with mean 0 and variance $\Sigma = \mathrm{Var}[r(\mathbf{x}) + yl(\mathbf{x})]$. In addition, Stoker (1990) proposes two "slope" estimators, denoted by $\hat{\beta}_{2IV}$ and $\hat{\beta}_{3IV}$, which estimate β as the linear slope coefficients β of y regressed on $\mathbf{x}$ using instrumental variables $\mathbf{w}_2$ and $\mathbf{w}_3$, derived from $\hat{\beta}_2$ and $\hat{\beta}_3$. Stoker shows that $\hat{\beta}_{2IV}$ and $\hat{\beta}_{3IV}$ are first-order asymptotically equivalent to $\hat{\beta}_2$ and $\hat{\beta}_3$.

We propose four more semiparametric estimators of average derivatives. In Section 2, we define two sample analog estimators, $\hat{\beta}_1$ and $\hat{\beta}_4$, based on

$E[g(\mathbf{x})l(\mathbf{x})]$ (1) and $\mathrm{E}[r(\mathbf{x})+yl(\mathbf{x})]$ (4), and two corresponding slope estimators, $\hat{\beta}_{1IV}$ and $\hat{\beta}_{4IV}$. In Section 3, we establishes asymptotical equivalence among all eight average derivative estimators. Some Cocluding remarks are given in Section 4.

2 Sample Analog Estimators and Slope Estimators

Let $\hat{f}(\mathbf{x}) = N^{-1}h^{-k}\sum_{j=1}^{N} K(\frac{\mathbf{x}-\mathbf{x}_j}{h})$ be a kernel estimator of the marginal density $f(\mathbf{x})$ and $\hat{g}(\mathbf{x}) = N^{-1}h^{-k}\sum_{j=1}^{N} y_j K(\frac{\mathbf{x}-\mathbf{x}_j}{h})/\hat{f}(\mathbf{x}) = \hat{G}(\mathbf{x})/\hat{f}(\mathbf{x})$ be a kernel estimator of $g(\mathbf{x})$, where $K : \mathcal{R}^k \rightarrow \mathcal{R}$ is a kernel function, h is the bandwidth parameter, and $h \rightarrow 0$ as $N \rightarrow \infty$. Then the kernel estimator of the derivative $g'(\mathbf{x})$ is defined as $\hat{g}'(\mathbf{x}) = \hat{r}(\mathbf{x}) + \hat{g}(\mathbf{x})\hat{l}(\mathbf{x})$ where $\hat{r}(\mathbf{x}) = \hat{G}'(\mathbf{x})/\hat{f}(\mathbf{x})$ is a kernel estimator of $r(\mathbf{x})$ and $\hat{l}(\mathbf{x}) = -\hat{f}'(\mathbf{x})/\hat{f}(\mathbf{x})$ is a kernel estimator of $l(\mathbf{x})$.

Härdle-Stoker (1989) and Stoker (1990) define $\hat{\beta}_2$ and $\hat{\beta}_3$ as the trimmed sample average of $y_i\hat{l}(\mathbf{x})$ and $\hat{g}'(\mathbf{x}_i)$, respectively, or $\hat{\beta}_2 = N^{-1}\sum_{i=1}^{N} y_i\hat{l}(\mathbf{x}_i)\hat{I}_i$ and $\hat{\beta}_3 = N^{-1}\sum_{i=1}^{N}\{\hat{r}(\mathbf{x}_i) + \hat{g}(\mathbf{x}_i)\hat{l}(\mathbf{x}_i)\}\hat{I}_i$, where $\hat{I}_i = I[\hat{f}(\mathbf{x}_i) > b]$ is an indicator function which equals 1 if $\hat{f}(\mathbf{x}_i) > b$, and 0 otherwise. The trimming bound b is chosen such that $b \rightarrow 0$ as $N \rightarrow \infty$.

In the same vein, we define two sample analog estimators of average derivatives based on (1) and (4) as $\hat{\beta}_1 = N^{-1}\sum_{i=1}^{N}\hat{g}(\mathbf{x}_i)\hat{l}(\mathbf{x}_i)\hat{I}_i$ and $\hat{\beta}_4 = N^{-1}\sum_{i=1}^{N}\{\hat{r}(\mathbf{x}_i) + y_i\hat{l}(\mathbf{x}_i)\}\hat{I}_i$. Comparing the definitions of the four sample analog estimators yields

$$\hat{\beta}_3 - \hat{\beta}_1 \equiv \hat{\beta}_4 - \hat{\beta}_2 = \frac{1}{N}\sum_{i=1}^{N}\hat{r}(\mathbf{x}_i)\hat{I}_i \equiv \hat{\delta} \tag{5}$$

where $\hat{\delta} = N^{-1}\sum_{i=1}^{N}\hat{r}(\mathbf{x}_i)\hat{I}_i$ is a trimmed sample analog estimator of $\delta \equiv \mathrm{E}[r(\mathbf{x})] = 0$, if $f(\mathbf{x})$ vanishes on the boundary of the support of $\mathbf{x}$. Given that the asymptotic properties of $\hat{\beta}_2$ and $\hat{\beta}_3$ are already known, the asymptotic properties of $\hat{\beta}_1$ and $\hat{\beta}_4$ depend only on those of $\hat{\delta}$.

Rewrite the slope estimators as $\hat{\beta}_j = N^{-1}\sum_{i=1}^{N}\mathbf{w}_j(\mathbf{x}_i)y_i$ for $j = 1\ldots 4$. Let $\mathbf{X} = [\mathbf{x}_1 - \bar{\mathbf{x}}, \cdots, \mathbf{x}_N - \bar{\mathbf{x}}]^T$ denote the $N \times k$ data matrix, $\mathbf{W}_j(\mathbf{X}) = [\mathbf{w}_j(\mathbf{x}_1), \cdots, \mathbf{w}_j(\mathbf{x}_N)]^T$, and $\mathbf{y} = [y_1 - \bar{y}, \cdots, y_N - \bar{y}]^T$. Then corresponding to each slope estimator $\hat{\beta}_j$, we have a slope estimator defined as $\hat{\beta}_{jIV} = (\mathbf{W}_j^T\mathbf{X})^{-1}\mathbf{W}_j^T\mathbf{y}$. Clearly, each slope estimator is motivated by the use of an appropriate instrumental variable $\mathbf{w}_j$ for the estimation of β in $y_i - \bar{y} =$

$(\mathbf{x}_i - \bar{\mathbf{x}})^T \beta + v_i - \bar{v}$; see Stoker (1986). Slope estimators $\hat{\beta}_{2IV}$ and $\hat{\beta}_{3IV}$ are proposed by Stoker (1990), but $\hat{\beta}_{1IV}$ and $\hat{\beta}_{4IV}$ are new.

3 Asymptotic Equivalence

Härdle-Stoker (1989) and Stoker (1990) show that given Assumptions 1–7 listed in Stoker (1990), $\sqrt{N}(\hat{\beta}_2 - \beta)$ and $\sqrt{N}(\hat{\beta}_3 - \beta)$ have a common limiting normal distribution with mean 0 and variance Σ, where $\Sigma = \mathrm{Var}[g'(\mathbf{x}) + (y - g(\mathbf{x}))l(\mathbf{x})]$. In addition, Stoker (1990) shows that given Assumptions 1–9 listed in Stoker (1990), $\sqrt{N}(\hat{\beta}_{2IV} - \beta)$ and $\sqrt{N}(\hat{\beta}_{3IV} - \beta)$ also have a common limiting normal distribution with mean 0 and variance Σ. In this section, we show that the four newly proposed estimators of average derivatives have the same asymptotical properties.

We first study the asymptotical properties of $\hat{\delta}$ because the asymptotic properties of $\hat{\beta}_1$ and $\hat{\beta}_4$ will depend on those of $\hat{\delta}$, given the properties of $\hat{\beta}_2$ and $\hat{\beta}_3$; see (5).

LEMMA 1 *Given Assumptions 1–7 stated in Stoker (1990), if (1)* $N \to \infty$, $h \to 0$, $b \to 0$, *and* $h/b \to 0$*; (2) for some* $\epsilon > 0$, $b^4 N^{1-\epsilon} h^{2k+2} \to \infty$*; and (3)* $N h^{2p-2} \to 0$, *where* $p \geq k+2$, *then* $\sqrt{N}\hat{\delta} \xrightarrow{p} 0$.

This lemma can be proved using a four step procedure similar to that in Härdle-Stoker (1989) and Stoker (1990). First, $\hat{\delta}$ is approximated by $\bar{\delta}$ whose density trimming is based on true values of the density function. Second, $\bar{\delta}$ is approximated by $\tilde{\delta}$ through linearization by appealing to the uniform properties of the kernel estimators of $f(\mathbf{x})$ and $G'(\mathbf{x})$. Third, $\tilde{\delta}$ is approximated by the asymptotically normal sum of U-statistics with kernels that vary with sample size N. Fourth, $\tilde{\delta}$ is shown to be asymptotically unbiased. However, given the space limitation here, the proof is omitted but is available upon request.

The assumptions in Lemma 1 are technically complicated. But in general, they are weak regularity conditions on the joint distribution of $(y, \mathbf{x})$, and weak smoothness conditions on $f(\mathbf{x})$, $r(\mathbf{x})$ and $G(\mathbf{x})$; see Stoker (1990).

Combining Lemma 1, the asymptotic properties of $\hat{\beta}_2$ and $\hat{\beta}_3$ are already established, and Equation (5), we have the following results.

THEOREM 2 *Given Assumptions 1–7 stated in Stoker (1990) and Conditions (1), (2) and (3) stated in Lemma 1,* $\sqrt{N}(\hat{\beta}_1 - \beta)$ *and* $\sqrt{N}(\hat{\beta}_4 - \beta)$ *have a common limiting normal distribution with mean 0 and variance* Σ*, where* Σ *is the covariance matrix of* $r(\mathbf{x}) + yl(\mathbf{x})$.

From this theorem, we establish the asymptotic properties of $\hat{\beta}_{1IV}$ and $\hat{\beta}_{4IV}$.

COROLLARY 3 *Given Assumptions 1–9 stated in Stoker (1990), the following results hold:*

1. $N^{-1/2}\sum_{i=1}^{N}\mathbf{w}_j(\mathbf{x}_i) = o_p(1)$ *for* $j = 1, 4$.

2. $\sqrt{N}\{N^{-1}\mathbf{W}_j^T\mathbf{X} - \mathbf{I}_k\} = o_p(1)$ *for* $j = 1, 4$, *where* $\mathbf{I}_k$ *is a* $k \times k$ *identity matrix.*

3. $N^{-1/2}\mathbf{W}_j^T(\mathbf{y} - \mathbf{X}\beta) \xrightarrow{d} R \sim N(0, \Sigma)$ *for* $j = 1, 4$.

4. $\sqrt{N}(\hat{\beta}_{jIV} - \beta) \xrightarrow{d} R$ *for* $j = 1, 4$.

In the Appendix, we give the proof for Corollary 3.

Combining Theorem 3.1 of Härdle-Stoker (1989) and Theorem 1 of Stoker (1990) with Theorem 2 and Corollary 3 in this paper, we now have shown that for all $j = 1, \cdots, 4$, $\sqrt{N}(\hat{\beta}_j - \beta)$ and $\sqrt{N}(\hat{\beta}_{jIV} - \beta)$ converge in distribution to a common random variable R, where $R \sim N(0, \Sigma)$. Consequently, the four sample analog estimators and the four slope estimators of β are first-order equivalent to each other.

COROLLARY 4 *Given Assumptions 1–9 stated in Stoker (1990) and Conditions (1), (2) and (3) stated in Lemma 1,* $\sqrt{N}(\hat{\beta}_i - \hat{\beta}_j) = o_p(1)$ *and* $\sqrt{N}(\hat{\beta}_{iIV} - \hat{\beta}_i) = o_p(1)$, *for all* $1 \le i \le 4$, $1 \le j \le 4$.

By Theorem 3.1 of Newey and Stoker (1989), $\hat{\beta}_i$ and $\hat{\beta}_{iIV}$ $(1 \le i \le 4)$ are asymptotically efficient in the sense that Σ reaches the semiparametric efficiency bound.

4 Conclusion

This paper demonstrates that for each of the four possible population representations of the average derivatives of a regression function, there is a semiparametric sample analog estimator of the average derivative; and corresponding to each of the four sample analog estimators, there is a "ratio of moments" estimator or an OLS-IV estimator. Two of the four sample analog estimators and their corresponding slope estimators have been shown by Härdle-Stoker (1989) and Stoker (1990) to be $\sqrt{N}$-consistent and asymptotically normal, and by Stoker to be asymptotically equivalent. This paper establishes asymptotic equivalence among all four sample analog estimators and their corresponding OLS-IV estimators by showing that the new sample analog estimators are asymptotic equivalent to the Härdle-Stoker and the Stoker estimators.

References

[1] Härdle, W. and T. M. Stoker (1989): "Investigating Smooth Multiple Regression by the Method of Average Derivatives," *Journal of the American Statistical Association*, 84(408):986–995, December.

[2] Newey, W. K. and T. M. Stoker (1989): "Efficiency Properties of Average Derivative Estimators," manuscript.

[3] Stoker, T. M. (1986): "Consistent Estimation of Scaled Coefficients," *Econometrica*, 54:1461–1481.

[4] Stoker, T. M. (1990): "Equivalence of Direct, Indirect and Slope Estimators of Average Derivatives," in J. L. Powell W. A. Barnett and G. Tauchen, editors, *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, Cambridge: Cambridge University Press.

Appendix

In this appendix, we show the properties of $\hat{\beta}_{1IV}$ only since the properties of $\hat{\beta}_{4IV}$ can be established analogously.

1. Let $y_i = y = 1$ be a constant. Note that $g(\mathbf{x}) \equiv \mathrm{E}[y|\mathbf{x}] = 1$, $\beta \equiv \mathrm{E}[g'(\mathbf{x})] = 0$, and $q(y, \mathbf{x}) = g'(\mathbf{x}) + (y - g(\mathbf{x}))l(\mathbf{x}) = 0$. From Theorem 2, we have

$$\sqrt{N}(\hat{\beta}_1 - \beta) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \{q(y_i, \mathbf{x}_i) - \mathrm{E}q(y, \mathbf{x})\} + o_p(1) \qquad (6)$$

   Substituting $y_i = y = 1$ into the above equation yields $N^{-1/2} \sum_{i=1}^{N} \mathbf{w}_1(\mathbf{x}_i) = o_p(1)$.

2. Let x_ℓ denote the ℓth component oMf $\mathbf{x}$ and let $y = x_\ell$. Noting that $g(\mathbf{x}) = \mathrm{E}[y|\mathbf{x}] = x_\ell$ and $g'(\mathbf{x}) = e_\ell$, a k-vector with its ℓth component equal 1 and other components 0, we have $q(y, \mathbf{x}) = g'(\mathbf{x}) + [y - g(\mathbf{x})]l(\mathbf{x}) = e_\ell$. By substituting $y = x_\ell$ in Equation 6 for $\ell = 1, \cdots, k$, and stacking the k equations together, we get

$$\sqrt{N}\{N^{-1}\mathbf{W}_1^T\mathbf{X} + N^{-1} \sum_{i=1}^{N} \mathbf{w}_1(\mathbf{x}_i)\bar{\mathbf{x}} - \mathbf{I}_k\} = o_p(1) \qquad (7)$$

   It follows that $\sqrt{N}(N^{-1}\mathbf{W}_1^T\mathbf{X} - \mathbf{I}_k) = o_p(1)$, since $\bar{\mathbf{x}}$ is bounded by Chebychev's inequality, and $N^{-1/2} \sum \mathbf{w}_1(\mathbf{x}_i) \xrightarrow{p} 0$.

3. Because $\sqrt{N}(\hat{\beta}_1 - \beta) \xrightarrow{d} R$, $\bar{y}$ is bounded by Chebychev's inequality, and $N^{-1/2} \sum \mathbf{w}_1(\mathbf{x}_i)$ converges in probability to 0, we have

$$\frac{1}{\sqrt{N}} \mathbf{W}_1^T (\mathbf{y} - \mathbf{X}\beta) = \sqrt{N}\hat{\beta}_1 + \bar{y} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{w}_1(\mathbf{x}_i) - \sqrt{N}\beta + o_p(1) \xrightarrow{d} R$$

4. It follows immediately from 2 and 3 that

$$\sqrt{N}(\hat{\beta}_{1IV} - \beta) = \left(\frac{\mathbf{W}_1^T \mathbf{X}}{N} \right)^{-1} \frac{\mathbf{W}_1^T (\mathbf{y} - \mathbf{X}\beta)}{\sqrt{N}} \xrightarrow{d} R \qquad (8)$$