Department of Economics
University of Southampton
Southampton SO17 1BJ
UK

# Discussion Papers in Economics and Econometrics

## A METHOD OF ESTIMATING THE AVERAGE DERIVATIVE, THE MULTIVARIATE CASE

Anurag N Banerjee

No. 0215

**This paper is available on our website**
**http://www/soton.ac.uk/~econweb/dp/dp02.html**

# A Method of Estimating the Average Derivative, the multivariate case.

Anurag N Banerjee*
Department of Economics
University of Southampton
Southampton, UK

**Abstract**

The paper uses local linear regression to estimate the "direct" Average Derivative $\delta = E\left(D\left[m\left(\mathbf{x}\right)\right]\right)$, where $m\left(\mathbf{x}\right)$ is the regression function. The estimate of $\delta$ is the weighted average of local slope estimates. We prove the asymptotic normality of the estimate under conditions which are different from the conditions used by Härdle-Stoker (H-S) (1989). Using Monte-Carlo simulation experiments we give some small sample results comparing our estimator with the H-S estimator under our conditions for asymptotic normality.

JEL codes: C13, C14, C15
Keywords: Semi-parametric estimation, Average Derivative, Linear regression

*I thank Prof Wolfgang Härdle for comments on a previous version of this paper.

# 1 Introduction

Let $(Y_t, \mathbf{X}_t)$, $t = 1, \ldots, T$ be a multivariate sample from an unknown distribution $F(y, \mathbf{x})$ which is generated from the following model

$$Y_t = m(\mathbf{X}_t) + u_t,\ t = 1, \ldots, T \tag{1}$$

where $m(\mathbf{x})$ is the unknown regression function, and the conditional expectation of $u_t$ given $X_t$ is zero, i.e. $E(u_t|\mathbf{X}_t) = 0$. We assume $\mathbf{x}$ is a $l - \textit{dimensional}$ vector.[1]

Further, we assume that the regression function $m(x)$ is differentiable. We define the average slope or the average derivative (A.D) of the regression function $m(\mathbf{x})$ as

$$\delta = \int D\left[m(\mathbf{x})\right] f(\mathbf{x})dx = E\left(D\left[m(\mathbf{x})\right]\right) \tag{2}$$

where $f(\mathbf{x})$ is the marginal density of $\mathbf{X}$'s and $D\left[m(\mathbf{x})\right]$ is the first derivative of $m$. We can argue that $\delta$ represents sensible "coefficients" of changes in $\mathbf{x}$. We can also show by integrating by parts.

$$\delta = E\left(\mathcal{L}\left(\mathbf{X}\right) Y\right)$$

where $\mathcal{L}(\mathbf{X}) = -D\left[f(\mathbf{X})\right]/f(\mathbf{X})$.

The primary interest for Average Derivative Estimation (A.D.E) comes from the General Index Model. where

$$m(\mathbf{x}) = G(\mathbf{x}'\beta)$$

then $E\left(D\left[m\right]\right) = E\left(dG/d(\mathbf{x}'\beta)\right)\beta$ is proportional to $\beta\ \forall\ \mathbf{x}$. So $\delta = E[dG/d(\mathbf{x}'\beta)]\beta = \gamma\beta$, for some $\gamma$, is proportional to $\beta$. We can equivalently replace $\beta$ by $\theta = \beta/\gamma$, by normalising as $m(\mathbf{x}) = G(\mathbf{x}'\theta)$ st. $E[dG/d(\mathbf{x}'\theta)] = 1$. Thus it can be interpreted as units of change in $y$ to changes in $\mathbf{x}$. Härdle and Stoker (1989) gives an application of it with a "Collision Data".

The use of A.D in the context of Partial Index Models is also useful. For this $\mathbf{x}$ is partitioned as $(\mathbf{x}_{(1)}, \mathbf{x}_{(2)})$ into a $l - \bar{l}$ vector of $\mathbf{x}_{(1)}$ and $\bar{l}$ of vector $\mathbf{x}_{(2)}$, and partition $\delta$ analogously as $(\delta_1, \delta_2)$. average derivatives will measure the true coefficient when the regressions obeys a Partial Index Structure ( Newey and Stoker (1989) ) if

$$m(x) = G(\mathbf{x}'_{(1)}\beta, \mathbf{x}_{(2)})$$

[1]Banerjee (1994) analyses the case of $l = 1$.

then $\delta_1$ equals $\beta_1$ up to a scale. With an estimator $\widehat{\delta}_1$ of $\delta_1$, we can extend the A.D.E method to fitting a $\bar{l}+1$ dimensional regression in the second stage, as $\hat{G}(\mathbf{x}'_{(1)}\widehat{\delta}_1, \mathbf{x}_{(2)})$. If the model is multiple index form as

$$m(\mathbf{x}) = G(\mathbf{x}'_{(1)}\beta_1, \mathbf{x}'_{(2)}\beta_2)$$

then $\delta_2$ is likewise proportional to $\beta_2$ (namely $E[dG/d(\mathbf{x}'_{(2)}\beta_2)]$). Again the A.D.E method is easily extended.

A.D.E's are also used in specific measurement problems in economics. A primary example by Härdle, Hilderbrand and Jerison (1991) is on measuring the positive definiteness of the aggregate income effects matrix for assessing the "Law of Demand".

The A.D.E is used to estimate the following matrix.

$$\delta_{jj'} = E\left(\frac{dE(Y_j Y_{j'}|x)}{dx}\right)$$

where $Y_j =$ demand for the $j^{th}$ good and $x =$ income level.

Further applications are suggested by the central role of derivatives in economic modelling in form of marginal reactions and elasticities. Examples like profit maximisation of firms can be given. In this problem the firm equate their marginal profit derived from a particular good to the price of that good. The average marginal reaction can be assessed by the A.D estimate of the marginal profit. One such example is given in Stoker (1992).

# 2 Method of Estimation.

Several methods have been suggested to estimate $\delta$, the Average Derivative. Härdle and Stoker (1989) proposed an "indirect" estimate, $\widehat{\delta}_{hs}$ which is the sample analog of $E\left[\mathcal{L}(\mathbf{X})Y\right]$. This method estimates the covariance between $L(\mathbf{X})$ and $Y$, using consistent non-parametric estimators of $f(\mathbf{X})$ and $D\left[f(\mathbf{X})\right]$. Stoker (1991b) defines the "direct" Average Derivative Estimate, the sample analog of $E(D\left[m(\mathbf{X})\right])$ as $\widehat{\delta}_d$ using the average of the consistant non-parametric estimator of $D\left[m(\mathbf{X})\right]$. Stoker also shows the asymptotic equivalence of the "direct" and "indirect" estimators. The consistent estimates used in the "direct" and "indirect" estimator are generally kernel estimators of the respective functions. The estimator we are going to define, uses local linear regression as a method.

We shall only assume some smoothness properties of the regression function and moment restrictions on the random variables which we state the next section. One important difference in this method from the other methods is that there are no smoothness assumptions on the marginal density of $\mathbf{X}$, i.e. $f(\mathbf{x})$. We do assume that, the support of $\mathbf{X}$ is the compact set $S$, without loss of generality it is assumed to be a subset of $[0,1]^l$. Unlike the Härdle and Stoker method, the Fisher's information $\mathcal{L}(\mathbf{X})$ may not exist. Therefore the "indirect" estimator will not exist as well. For example suppose $\mathbf{X}$ is distributed $U[0,1]$, then $\mathcal{L}(\mathbf{X})$ does not exist. This case will not be covered by the method proposed by Härdle and Stoker. On the other hand if $\mathbf{X}$ is distributed with a Normal density we cannot use our method since we assume the domain of $f(\mathbf{x})$ to be a compact interval. Though in this case we can use the Härdle and Stoker method. So comparisons of our two methods in terms of the asymptotics cannot be made and our methods complement the Härdle and Stoker method.

Let us motivate our method when $x$ is univariate. Without loss of generality, let $S$ be the interval $[0,1]$. This interval is then partitioned, in equal intervals. We denote the partition as $P$. Let the partition be $0 < t_1 < t_2 < \ldots < t_{k-1} < 1$, we denote $(t_r, t_{r+1}]$ as $H_r$ ($H_r$ is called a bin ). These bins are of equal size ($|H_r| = h$). In the bins, which have at least 3 observations we linearly regress $Y_t$ on $X_t$, st. $X_t \in H_r$. We denote the coefficient of the slope of the regression as $\hat{\beta}_r$. This is a least squares estimate of the tangent of the regression curve $m(\mathbf{x})$, in the interval $H_r$.

We then take the an weighted average of the slopes in each of the bin $H_r$. The weights are taken to be the average number of observations in the bin $H_r$.

Let us now generalise the idea when the dimension of $x$ is $l$.

Assume without loss of generality, the interval $[0,1]$ is the domain of the marginal density of $X_i$. We partition the domain, in equal intervals and denote the partition as $P_i$. Let the partition be $0 < t_{i1} < t_{i2} < \ldots < t_{i(k_i-1)} < 1$, we denote $(t_r^n, t_{r+1}^n]$ as $H_{r_i}$ ($H_{r_i}$ is called a bin in the $x_i^{th}$ dimension).These bins are of equal size ($|H_{r_i}|$). The partition for the whole of domain of $f(\mathbf{x})$ is then $\mathbf{P} = P_1 \times \ldots \times P_l$ where $H_r = H_{r_1} \times \ldots \times H_{r_l}$ is the bin to be considered in this $l-dimensional$ space. Notice the number of bins is now at most $k = k_1 \times \ldots \times k_l$. We shall only consider those bins such that $H_r \subset S$. Note that when $x$ is univariate then $H_r \subset S$ for all $r$. The same is true if $X'_{it}s$ are independent. The rest of the method is similar to the univariate case.

Suppose we have atleast $p \geq l+2$ points in $\mathbf{H}_r$, we linearly regress $y_t$ on $\mathbf{x}_t$ as

$$y_t = \alpha_r + \beta_r' \mathbf{x}_t \ , \ s.t \ \mathbf{X}_t \in \mathbf{H}_r.$$

We denote the estimate coefficient of the slope of the regression, $\beta_r$ as

$$\hat{\beta}_r = [\mathbf{S}_x^r]^{-1} \sum_{t=1}^{T} (\mathbf{x}_t - \bar{\mathbf{x}}_r) \, I\{\mathbf{x}_t \in \mathbf{H}_r\} \, y_t$$

$$\text{where } \mathbf{S}_x^r = \sum_{t=1}^{T} (\mathbf{x}_t - \bar{\mathbf{x}}_r)(\mathbf{x}_t - \bar{\mathbf{x}}_r)^T I\{\mathbf{x}_t \in \mathbf{H}_r\}$$

$$\text{and } \bar{\mathbf{x}}_r = \frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t I\{\mathbf{x}_t \in \mathbf{H}_r\} \, .$$

This is a least squares estimate of the tangent of the regression curve $m(\mathbf{x})$, within the interval $\mathbf{H}_r$.

We then take the an weighted average of the slopes in each of the bin $\mathbf{H}_r$. The weights are taken to be the average number of observations in the bin $H_r$, denoted by

$$w_r = \frac{1}{T} \sum_{t=1}^{T} I\{\mathbf{x}_t \in H_r\}.$$

where $I$ is the indicator function.

**Definition 1** *We define our Average Derivative estimator as*

$$\hat{\delta} = \sum_{r=1}^{k} w_r \widehat{\beta}_r I\{T_r \geq p\}$$

*where $T_r = w_r T$, the number of observations in the $r^{th}$ bin and $k$ is the number of bins.*

Note that in definition (1), we assume that if there are insufficient number of observations to regress, the observations in the bin contribute nothing to the Average Derivative Estimate.

We will show, under some assumptions made later that asymptotically

$$\sqrt{T}(\hat{\delta} - \delta) \simeq N\left(0, Var\{m'(\mathbf{X})\} + \sigma_u^2 \Sigma^{-1}\right),$$

where $\sigma_u^2$ is the variance of $u_i$'s and $\Sigma$ is the variance-covariance matrix of $\mathbf{X}$.

We also show that the large sample variance $Var\{m'(\mathbf{X})\} + \sigma_u^2 \Sigma^{-1}$ can be consistently estimated by the following estimator.

**Definition 2** *We define the estimated variance of $\hat{\delta}$ as*

$$\hat{V} = \sum_{r=1}^{k} w_r \hat{\beta}_r \hat{\beta}_r{}' I\{T_r \geq p\} - \hat{\delta}\hat{\delta}'$$

# 3 Distributional properties and comparison with Härdle-Stoker Estimator.

## 3.1 Large Sample Results

We shall now prove some large sample results under the following assumptions

A1 The support of $f(\mathbf{x})$ is the compact set $S \subset [0,1]^l$ and $f(\mathbf{x})$ is uniformly bounded above by a constant $C$, for some $C > 0$.

A2 The second derivative of $m(x)$,$D^2[m]$, exists and bounded.

A3 The variance of $u_t$ is $E(u_t^2|X_t) = \sigma_u^2$, exists and is bounded.

A4 As $T \longrightarrow \infty$,

$$\sqrt{T}h \longrightarrow 0 \text{ and } \frac{log(T)}{Th} \longrightarrow 0.$$

We will make some brief comments on the assumptions. The first assumption (A1) is not a popular assumption is the non-parametric econometrics literature. This assumption of $f(\mathbf{x}) > 0$ is necessary to ensure that there is atleast $p - observations$ in each bin to perform the required regression (in large sample). However we also want the density to be bounded above since we do not want to put too much weight on any particular $\hat{\beta}_r$. The smoothness assumption of the regression function (A2) is also necessary for the same reason. Assumption three (A3) is a standard assumption for linear models. Finally the last assumption (A4) ensures that the size of the bins shrinks at the rate of $\sqrt{T}$, but the size should not get too small too quickly ($log(T)/Th \longrightarrow 0$) otherwise there will be insufficient number of observations in the bin to do a regression.

**Theorem 1** *Under the stated assumptions A(1) to A(4) we have the following*

$$\sqrt{T}\left(\hat{\delta} - \delta\right) \xrightarrow{D} N\left(0, Var\left(D\left[m\left(\mathbf{X}\right)\right]\right) + \sigma_u^2 \Sigma^{-1}\right)$$

*where $\widehat{\delta}$ is the A.D.E defined in Definition 1.*

The interesting thing to observe here is that if $m(\mathbf{x})$ is linear (i.e. $m(\mathbf{x}) = \alpha + \beta'\mathbf{x}$) then the asymptotic variance coincides with the asymptotic variance of the classical Least Squares estimator of $\beta$. Note that in case of $m(\mathbf{x})$ being linear, $\delta = \beta$. So in this particular case we get a standard classical result. This implies that in the case of linearity we will not lose efficiency when compared to the Least Square Estimation method.

**Theorem 2** *Under the assumptions A(1) to A(4) we have the following*

$$\widehat{\mathbf{V}} \xrightarrow{P} Var\left(D\left[m\left(\mathbf{X}\right)\right]\right) + \sigma_u^2 \Sigma^{-1}$$

*where $\widehat{\mathbf{V}}$ is the estimated variance of $\widehat{\delta}$ as defined in Definition 2.*

Theorem (2) facilitates the measurement of precision of $\hat{\delta}$ as well as the inference on hypotheses about $\delta$. For instance, getting interval estimates using the estimated covariance matrix of $\widehat{\delta}$, $\widehat{\mathbf{V}}$.

Moreover, consider testing restrictions of $H_o : \delta = \delta_0$. Tests of this hypothesis can be based on the Wald type $W$ statistic

$$W = \left(\widehat{\delta} - \delta_o\right)' \widehat{\mathbf{V}}^{-1} \left(\widehat{\delta} - \delta_o\right) \tag{3}$$

which will have a limiting $\chi^2$ distribution.

As a practical application, since we do not require the density $f\left(\mathbf{x}\right)$ to vanish, our method can be used to test for linearity or stability by dividing the data into different regions and calculating the ADE of each region and testing for equality like a Chow test using (3).

## 3.2 Small Sample Results

We will now study the small sample properties of our estimator and compare it with the Hardle Stoker Estimator. We do so by using Monte Carlo simulations on a model satisfying the assumptions listed before.

**Model**

We study a univariate model as described below,

$$\begin{aligned} m(x) &= 1 - x + x^2 \\ u &\sim N(0, \sigma_u^2) \\ X &\sim U[0, 1] \end{aligned}$$

Therefore, for this model:

$$\delta = 0$$

and

$$Var\{D[m(x)]\} + \sigma_u^2 \Sigma^{-1} = \frac{1}{3} + 12\sigma_u^2$$

Let us describe the algorithm for computing our estimate.

**Algorithim**

**Step(0)** Generate $\{(X_t, Y_t)\}_{t=1}^{T}$ from the model.

**Step(1)** Choose the size of the bin such that it satisfies A(4).

**Step(2)** Divide the domain into $k$ parts as described before.

**Step(3)** Compute the Least Square Estimate, $\hat{\beta}_r$ with at least 3 observations in each bin, $\mathbf{H}_r$. Compute the ratio $\#0\{\mathbf{X}_t \in \mathbf{H}_r\}/T = w_r$. Multiply and get $\hat{\beta}_r w_r$.

**Step(4)** Add $\hat{\beta}_r w_r$ over all bins, $\mathbf{H}_r$ and get the estimate $\delta$.

**Choice of Bin Size.**

As observed before the size of the bin is inversely proportional to the number of partitions. We describe here an adhoc method of choosing $h$ from the data size $(T)$. We will use A(4) and the "definition of limit" to choose our bin width. We have,

$$\sqrt{T}h \longrightarrow 0 \text{ and } \frac{\log(T)}{Th} \longrightarrow 0 \text{ as } T \longrightarrow \infty.$$

implies given $\epsilon > 0$ , $\exists\ \overline{T}$ st. for $T \geq \overline{T}$ $\sqrt{\phantom{x}}$ $\frac{}{Th}$ □ $\epsilon$, and from these two inequalities we have

$$\frac{\log(T)}{T\epsilon} \,\square\, h \,\square\, \frac{\epsilon}{\sqrt{T}}$$

From this we fix $h$ as follows. Taking equalities on both sides, we have

$$\begin{aligned} \frac{\log(T)}{T\epsilon} &= \frac{\epsilon}{\sqrt{T}} \\ \frac{\log(T)^{1/2}}{T^{1/4}} &= \epsilon \end{aligned}$$

so we get

$$h = \frac{\log(T)^{1/2}}{T^{3/4}}$$

so

$$k = \left[ \sqrt{\frac{T\sqrt{T}}{\log(T)}} \right]$$

**Simulations and Descriptive Statistics.**

We generate $s$ $(= 1000)$ datasets of size $T$ $(= 50, 200, 400)$ from the model we consider. Then with these data sets we estimate $\delta$ with the method described before and get the estimated value of $\delta$ $(\widehat{\delta}_T)$. We will denote by $\widehat{\delta}_T(i)$ as the estimated value of $\delta$ of the $i^{th}$ simulation (i.e. with the $i^{th}$ dataset). With these $\widehat{\delta}_T(i)$ we calculate the following summary descriptive statistics, to show how the estimator behaves. We shall now give a brief description of the summary statistics.

$$\begin{aligned} \text{Mean of } \widehat{\delta}_T(i)'s &= \overline{\widehat{\delta}}_T = \frac{1}{s}\sum_{i=1}^{s} \widehat{\delta}_T(i) \\ \text{Variance of } \widehat{\delta}_T(i)'s &= V(\widehat{\delta}_T) = \frac{1}{s}\sum_{i=1}^{s} (\widehat{\delta}_T(i) - \overline{\widehat{\delta}}_T)^2 \\ \text{MSE of } \widehat{\delta}_T(i)'s &= MSE(\widehat{\delta}_T) = \frac{1}{s}\sum_{i=1}^{s} (\widehat{\delta}_T(i) - \delta)^2 \end{aligned}$$

Further more we will look at the estimate of the $\Pr(-\frac{3}{\sqrt{T}}\sqrt{V} + \delta \,\square\, \widehat{\delta}_T \,\square\, \frac{3}{\sqrt{T}}\sqrt{V} + \delta)$, where $V = Var\left(D\left[m(X)\right]\right) + \sigma_u^2/\sigma_x^2$. This is a natural statistic to look at, since by the Theorem in the previous section we know that

$$\Pr(-\frac{3}{\sqrt{T}}\sqrt{V} + \delta \,\square\, \widehat{\delta}_T \,\square\, \frac{3}{\sqrt{T}}\sqrt{V} + \delta) \overset{asy}{\sim} \Phi(-\frac{3}{\sqrt{T}}\sqrt{V} + \delta \,\square\, \widehat{\delta}_T \,\square\, \frac{3}{\sqrt{T}}\sqrt{V} + \delta)$$

for large $T$.

So we can look at the following estimate of the above probability as

$$\widehat{\Pr} = \frac{1}{s}\sum_{i=1}^{s} I(-\frac{3}{\sqrt{T}}\sqrt{V} + \delta \square \hat{\delta}_T(i) \square \frac{3}{\sqrt{T}}\sqrt{V} + \delta)$$

This probability gives us an estimate of how accurately our $\hat{\delta}_T$ estimates $\delta$ in small samples.

**Results**

In the model described before, we vary the error variance ($\sigma_u^2 = 1, 4, 9$), so as the disturbance of the error increases we expect to see a larger variation about the mean of the estimates $\bar{\hat{\delta}}_T$ and the actual $\delta$ ($\delta = 0$ in this model). The results are tabulated in Table 1. We see as expected with the decrease of sample size the variation increases. The closeness of $V(\hat{\delta}_T)$ and $MSE(\hat{\delta}_T)$ tells us that our $\hat{\delta}_T$'s are close to the actual, with increasing $\sigma_u^2$ and sample size $T$.

$[Table 1]$

## 3.3 Comparisons with Härdle-Stoker Method

The Härdle-Stoker method (1989) uses the indirect estimate

$$\hat{\delta}_{hs} = \frac{1}{T}\sum_{t=1}^{T} y_t \frac{\widehat{Df}_h(\mathbf{x}_t)}{\hat{f}_h(\mathbf{x}_t)}$$

where $\hat{f}_h(\mathbf{x}_t)$ and $\widehat{Df}_h(\mathbf{x}_t)$ are the kernel density estimates with a bandwidth $h$. It has been shown in (Härdle-Stoker 1989) that under some assumptions

$$\sqrt{T}\left(\hat{\delta}_{hs} - \delta\right) \overset{asy}{\approx} N\left(0, Var\left(D\left[m\left(\mathbf{X}\right)\right]\right) + \sigma_u^2 E\left(\mathcal{L}(\mathbf{X})\right)^2\right) \tag{4}$$

when the error term $\mathbf{u}$ is uncorrelated with $\mathbf{X}$. In his article, Stoker (1991b), defines the "direct" estimator of $\delta$ defined as

$$\hat{\delta}_d = \frac{1}{T}\sum_{t=1}^{T} D\left[\widehat{m}\left(\mathbf{x}_t\right)\right] I\left(\hat{f}_h\left(\mathbf{x}_t\right) > b\right)$$

where $\widehat{m}(\mathbf{x}_t)$ is the (Nadaraya-Watson) kernel regression estimator of $m(\mathbf{x})$, is asymptotically eqivalent to the indirect estimator $\hat{\delta}_{hs}$. He also finds under similar conditions of the H-S estimator, $\sqrt{T}\left(\widehat{\delta}_d - \delta\right)$ has the same asymptotic distribution as in (4). Given this we shall compare only the $\widehat{\delta}_{hs}$ estimator with our proposed $\widehat{\delta}$.

Comparing the asymptotic variances of the $\widehat{\delta}_{hs}$ or $\widehat{\delta}_d$ (4) and the asymptotic variance of $\widehat{\delta}$ (Theorem 1), we see that by Rao-Cramer inequality,

$$Var\left(D\left[m\left(\mathbf{X}\right)\right]\right) + \sigma_u^2 E\left(\mathcal{L}(\mathbf{X})\right)^2 > Var\left(D\left[m\left(\mathbf{X}\right)\right]\right) + \sigma_u^2 \Sigma^{-1}.$$

But does this implies that $\widehat{\delta}$ is asymptotically more efficient than $\widehat{\delta}_{hs}$ or $\widehat{\delta}_d$ ? The answer to that question is not necessarily so, since assumption (A1) used to derive the asumptotic violates the assumption of smoothness of $f$ needed for the asymptotic normality of Härdle-Stoker A.D.E [2]. Also H-S assumptions on $f$ violates assumption A1, since we need the assumption of compact support of $f$ for asymptotic normality of $\widehat{\delta}$.. Hence they can only be compared through simulation methods.

We use the same model as in (4) [3]. To compute the H-S estimator bandwidth of the Kernel, $h$ is taken to be $T^{-2/7}$, the optimal bandwidth obtained by minimising the MSE (Härdle, Hart, Marron and Tsybakov (1991)). We use the Gaussian Kernel to compute the density.

$$[Table 2]$$

Generally as expected our proposed A.D.E $\widehat{\delta}$ out performs $\widehat{\delta}_{hs}$ in this simulation, the reason being that model violates the condition for asymptotic normality of $\widehat{\delta}_{hs}$. So our estimator complements the H-S estimator.

# 4 Conclusion

The paper proposes an alternative method of estimating the Average Derivative Estimate (A.D.E). We propose the method of averaging the local OLS slopes to estimate the A.D.E.

We prove the asymptotic normality of our A.D.E under some regularity assumptions. These assumptions are similar but not same as the assumptions under which Härdle-Stoker

[2] Assumption 1, $f(x) = 0$, at the boundary of the support and.assumption 5 all derivatives of $f(x)$ of order $l+2$, exists. (Härdle-Stoker, 1989)

[3] Notice that the assumption that $X \sim U(0,1)$ violates the assumption of smoothness of $f$ needed for the asymptotic normality of Härdle-Stoker A.D.E.

(H-S) proved the (asymptotic) normality of their A.D.E. Stoker (1991b) also defines a "direct" estimator of $\delta$, and shows the assymptotic equivalence of the direct and the H-S estimator. The H-S estimator requires some smoothness conditions on the density of explanatory variable $f(\mathbf{x})$. Our method we do not require such assumptions but we need $f(\mathbf{x})$ to have compact support. It might be worthwhile to point out that by not requiring the density $f(\mathbf{x})$ to vanish, our method can be used to test for linearity or stability by dividing the data into different regions and calculating the ADE of each region and testing for equality like a Chow test.

The method described, is applied to a model with single regressor, assuming the density of $\mathbf{x}$ to be uniform. We simulate and compare the small sample results of H-S estimator with our estimator using various measures of performance. The results also indicate that our estimator performs better than H-S estimator under the given situation where asymptotic conditions of the Härdle Stoker method is not strictly applicable. Our method thus complements the H-S method.

## References

Ahn, H, 1997, Semiparametric Estimation of a Single-Index Model with Nonparametrically Generated Regressors Econometric-Theory; Vol 13(1), pp 3-31.

Banerjee, A.N, 1994, A Method of Estimating the Average Derivative, Universite Catholique de Louvain CORE Discussion Paper: 9403.

Härdle, W, 1992, Applied Non Parametric Regression, Cambridge, Cambridge University Press (Econometric Society Monographs).

Härdle, W., W.Hilderbrand. and M.Jerison, 1991, Emperical Evidence for the Law of Demand, Econometrica, Vol 59, pp 1525-1550.

Härdle, W, and T.M.Stoker, 1989, Investigating Smooth Multiple Regression by the Method of Average Derivatives, Journal of American Statistical Association, Vol 84, 408, pp 986-995.

Härdle, W, J.Hart, J.S.Marron, and A.B.Tsybakov, 1992, Bandwidth Choice for Average Derivative Estimation, Journal of the American Statistical Association, Vol 87, 417, pp. 218-226

Newey, W.K and T.M. Stoker, 1993, Efficiency of Weighted Average Derivative Estimators and Index Models, Econometrica, Vol 61, pp 1199-1223

Rao, C.R., 1973, Linear Statistical Inference and its Applications, (John Wiley, New York).

Serfling, R.J, 1980, Approximation Theorems of Mathematical Statistics, (John Wiley, New York).

Stoker, T.M., 1991a, Lectures on Semiparametric Econometrics, (CORE Lecture Series, Core Foundation, Louvain-la Neuve).

Stoker, T.M, 1991b, Equivalance of Direct, Indirect and Slope Estimators of Average Derivatives, Nonparametric and Semiparametric Methods in Econometrics and

Statistics, W.A. Barnett, J,Powell and G, Tauchen, eds., Cambridge University Press.

# Tables

**Table 1: Simulation results:**

| $T = 400$ | $\bar{\hat{\delta}}_T$ | $V(\hat{\delta}_T)$ | $MSE(\hat{\delta}_T)$ | $\widehat{\Pr}$ |
|---|---|---|---|---|
| $\sigma_u^2 = 1$ | 0.0322 | 0.0853 | 0.08643 | 0.99 |
| $\sigma_u^2 = 4$ | -0.1044 | 0.2032 | 0.2141 | 0.98 |
| $\sigma_u^2 = 9$ | -0.0984 | 0.4790 | 0.4887 | 0.98 |

| $T = 200$ | $\bar{\hat{\delta}}_T$ | $V(\hat{\delta}_T)$ | $MSE(\hat{\delta}_T)$ | $\widehat{\Pr}$ |
|---|---|---|---|---|
| $\sigma_u^2 = 1$ | 0.0127 | 0.1909 | 0.1911 | 0.98 |
| $\sigma_u^2 = 4$ | 0.0119 | 0.4327 | 0.4329 | 0.98 |
| $\sigma_u^2 = 9$ | -0.0845 | 0.7585 | 0.7657 | 0.99 |

| $T = 50$ | $\bar{\hat{\delta}}_T$ | $V(\hat{\delta}_T)$ | $MSE(\hat{\delta}_T)$ | $\hat{P}$ |
|---|---|---|---|---|
| $\sigma_u^2 = 1$ | 0.0183 | 0.6835 | 0.6838 | 0.97 |
| $\sigma_u^2 = 4$ | -0.0712 | 1.8086 | 1.8137 | 0.97 |
| $\sigma_u^2 = 9$ | 0.1628 | 3.5835 | 3.6101 | 0.99 |

**Table 2: Comparison with Härdle-Stoker method**

| $T = 100$ | $\bar{\hat{\delta}}_{hs}$ | $\bar{\hat{\delta}}$ | $MSE(\hat{\delta}_{hs})$ | $MSE(\hat{\delta})$ |
|---|---|---|---|---|
| $\sigma_u^2 = 1$ | -1.128 | 0.0575 | 3.8334 | 0.2003 |
| $\sigma_u^2 = 4$ | 1.6077 | -0.0328 | 2.6459 | 0.6492 |
| $\sigma_u^2 = 9$ | -0.1951 | -0.1951 | 4.2662 | 1.4613 |

where

$$\text{Mean of } \hat{\delta}_{hs}(i)'s = \bar{\hat{\delta}}_{hs} = \frac{1}{100}\sum_{i=1}^{100}\hat{\delta}_{hs}(i)$$

$$\text{MSE of } \hat{\delta}_{hs}(i)'s = MSE(\hat{\delta}_{hs}) = \frac{1}{100}\sum_{i=1}^{100}(\hat{\delta}_{hs}(i) - \delta)^2$$

# Appendix.

**Lemma 1** *Under the assumptions, we have as $T \to \infty$,*

$$1) \sup_{1<r<k} T^{-\frac{1}{4}} \left\| \frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' I \{\mathbf{x}_t \in H_r\} - p_r \boldsymbol{\mu}_2 \right\| \xrightarrow{P} 0$$

$$2) \sup_{1<r<k} T^{-\frac{1}{4}} \left\| \frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t I \{\mathbf{x}_t \in H_r\} - p_r \boldsymbol{\mu}_1 \right\| \xrightarrow{P} 0$$

$$3) \sup_{1<r<k} T^{-\frac{1}{4}} \left\| \frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t u_t I \{\mathbf{x}_t \in H_r\} \right\| \xrightarrow{P} 0$$

$$4) \sup_{1<r<k} T^{-\frac{1}{4}} \left\| \frac{1}{T} \sum_{t=1}^{T} u_t I \{\mathbf{x}_t \in H_r\} \right\| \xrightarrow{P} 0$$

$$5) \sup_{1<r<k} T^{-\frac{1}{4}} \left\| \frac{1}{T} \sum_{t=1}^{T} I \{\mathbf{x}_t \in H_r\} - p_r \right\| \xrightarrow{P} 0$$

*where*

$$p_r = \int_{H_r} f(\mathbf{x})\, d\mathbf{x}, \ \boldsymbol{\mu}_1 = E(\mathbf{x}_t) \ \text{ and } \boldsymbol{\mu}_2 = E(\mathbf{x}_t \mathbf{x}_t').$$

**Proof of Lemma 1:** Observe that, if $\mathbf{M}_r, (1 \,\square\, r \,\square\, k)$ are a collection of independent random variables then,

$$\Pr\left\{ \sup_{1<r<k} \|\mathbf{M}_r\| > \varepsilon \right\} = 1 - \prod_{r=1}^{k} \left(1 - \Pr\left(\|\mathbf{M}_r\| > \varepsilon\right)\right)$$

so

$$\Pr\left\{ \sup_{1<r<k} \|\mathbf{M}_r\| > \varepsilon \right\} \to 0$$

$$\text{iif } \prod_{r=1}^{k} \left(1 - \Pr\left(\|\mathbf{M}_r\| > \varepsilon\right)\right) \to 1$$

$$\text{iif } \sum_{r=1}^{k} \Pr\left(\|\mathbf{M}_r\| > \varepsilon\right) \to 0$$

$$\text{if } \sum_{r=1}^{k} E\|\mathbf{M}_r\|^2 \to 0 \qquad \text{(using Chebyshev's inequality)} \tag{5}$$

Thus, if $\sum_{r=1}^{k} E\|\mathbf{M}_r\|^2 \to 0$, then $\sup_{1<r<k} \|\mathbf{M}_r\| \xrightarrow{P} 0$.

1) Let $\mathbf{M}_r = T^{-\frac{1}{4}} \frac{1}{T} \sum_{t=1}^{T} \left( \mathbf{x}_t \mathbf{x}_t' I \left\{ \mathbf{x}_t \in H_r \right\} - p_r \boldsymbol{\mu}_2 \right)$, so

$$
\begin{aligned}
& E \left\| \mathbf{M}_r \right\|^2 = \sqrt{T} E \left\| \frac{1}{T} \sum_{t=1}^{T} \left( \mathbf{x}_t \mathbf{x}_t' I \left\{ \mathbf{x}_t \in H_r \right\} - p_r \boldsymbol{\mu}_2 \right) \right\|^2 \\
& \leq \sqrt{T} \frac{1}{T^2} E \left\| \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \left( I \left\{ \mathbf{x}_t \in H_r \right\} - p_r \right) \right\|^2 + E \left\| p_r \left( \frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' - \boldsymbol{\mu}_2 \right) \right\|^2 \\
& \leq \sqrt{T} \frac{1}{T^2} E \left\| \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \left( I \left\{ \mathbf{x}_t \in H_r \right\} - p_r \right) \right\|^2 + p_r^2 E \left\| p_r \left( \frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' - \boldsymbol{\mu}_2 \right) \right\|^2 \\
& \leq \sqrt{T} \frac{1}{T^2} E \left( \sum_{t=1}^{T} \left\| \mathbf{x}_t \mathbf{x}_t' \right\| \left| I \left\{ \mathbf{x}_t \in H_r \right\} - p_r \right| \right)^2 + p_r^2 E \left\| \left( \frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' - \boldsymbol{\mu}_2 \right) \right\|^2 \\
& \leq \sqrt{T} \frac{1}{T^2} Const E \left( \sum_{t=1}^{T} \left| I \left\{ \mathbf{x}_t \in H_r \right\} - p_r \right| \right)^2 \quad \left( \begin{array}{l} \text{(since } \mathbf{x}_t \text{'s are bounded} \\ \text{and } E \left( \mathbf{x}_t \mathbf{x}_t' \right) = \boldsymbol{\mu}_2 \end{array} \right) \\
& = \sqrt{T} \frac{1}{T^2} Const \left[ \begin{array}{c} \sum_{t=1}^{T} E \left( I \left\{ \mathbf{x}_t \in H_r \right\} - p_r \right)^2 \\ + \sum_{t<t'=1}^{T} E \left| I \left\{ \mathbf{x}_t \in H_r \right\} - p_r \right| E \left| I \left\{ \mathbf{x}_t \in H_r \right\} - p_r \right| \end{array} \right]
\end{aligned}
$$

(since $\mathbf{x}_t$'s are independent)

$$
\begin{aligned}
& = \sqrt{T} \frac{1}{T^2} Const \left[ T p_r \left( 1 - p_r \right) + \binom{T}{2} \left( 2 p_r \left( 1 - p_r \right) \right)^2 \right] \\
& \leq Const. \left[ \frac{1}{\sqrt{T}} p_r + p_r \sqrt{T} h \right].
\end{aligned}
$$

Therefore summing over $r$ we get

$$
\sum_{r=1}^{k} E \left\| \mathbf{M}_r \right\|^2 < Const. \left[ \frac{1}{\sqrt{T}} + \sqrt{T} h \right].
$$

Hence as $T \to \infty$, the expression above goes to zero since $\sqrt{T} h \to 0$.

The proofs of 2) 3) 4) and 5) are similar to 1).

**Lemma 2** *Assume A(2) and A(4), if* $T_r \geq p$ *then ,*

$$
\hat{\beta}_r = D \left[ m(\bar{\mathbf{x}}_r) \right] + R_r^{(1)} + \ \boldsymbol{\theta}_r
$$

*where,*

$$
\begin{aligned}
\sup_{1 \leq r \leq k} R_r^{(1)} & \overset{P}{=} o \left( T^{-\frac{1}{2}} \right) \\
\text{and } \boldsymbol{\theta}_r &= \left[ \mathbf{S}_x^r \right]^{-1} \sum_{t=1}^{T} \left( \mathbf{x}_t - \bar{\mathbf{x}}_r \right) I \left\{ \mathbf{x}_t \in \mathbf{H}_r \right\} u_t.
\end{aligned}
$$

**Proof of Lemma 2:** We have

$$\hat{\beta}_r = [\mathbf{S}_x^r]^{-1} \sum_{t \in I_r} \widetilde{\mathbf{x}}_{t,r} y_t = [\mathbf{S}_x^r]^{-1} \sum_{t \in I_r} \widetilde{\mathbf{x}}_{t,r} m(\mathbf{x}_t) + \theta_r,$$

$$\text{where } \widetilde{\mathbf{x}}_{t,r} = (\mathbf{x}_t - \bar{\mathbf{x}}_r) I \{\mathbf{x}_t \in H_r\} \text{ and } I_r = \{t : \mathbf{x}_t \in H_r\}.$$

Take a Taylor series expansion around $\bar{\mathbf{x}}_r$ of $m(\mathbf{x}_t)$, for those $\mathbf{x}_t$ 's which are in $H_r$.

$$m(\mathbf{x}_t) = m(\bar{\mathbf{x}}_r) + \widetilde{\mathbf{x}}'_{t,r} D [m(\bar{\mathbf{x}}_r)] + \frac{1}{2} \widetilde{\mathbf{x}}'_{t,r} D^2 [m(\boldsymbol{\xi}_{tr})] \widetilde{\mathbf{x}}_{t,r}$$

for some $\boldsymbol{\xi}_{tr}$ between $\mathbf{x}_t$ and $\bar{\mathbf{x}}_r$. Therefore

$$[\mathbf{S}_x^r]^{-1} \sum_{t \in I_r} \widetilde{\mathbf{x}}_{t,r} m(\mathbf{x}_t) = D [m(\bar{\mathbf{x}}_r)] + R_r^{(1)}$$

where

$$R_r^{(1)} = \frac{1}{2} \left[ \frac{\mathbf{S}_x^r / T}{tr (\mathbf{S}_x^r / T)} \right]^{-1} \frac{\sum_{t \in I_r} \widetilde{\mathbf{x}}_{t,r} \widetilde{\mathbf{x}}_{t,r}^T D^2 [m(\boldsymbol{\xi}_{tr})] \widetilde{\mathbf{x}}_{t,r}}{tr (\mathbf{S}_x^r)}$$

and

$$\mathbf{S}_x^r = \sum_{t \in I_r} \widetilde{\mathbf{x}}_{t,r} \widetilde{\mathbf{x}}'_{t,r}$$

using the previous lemmas. Taking the norm

Note that by lemma (1), we have

$$\frac{1}{T} \mathbf{S}_x^r \xrightarrow{P} p_r \left( \boldsymbol{\mu}_2 - \boldsymbol{\mu}_1^2 \right) = p_r \Sigma$$

therefore

$$tr \left( \frac{1}{T} \mathbf{S}_x^r \right) \xrightarrow{P} p_r tr (\Sigma)$$

combining we get

$$\frac{\mathbf{S}_x^r / T}{tr (\mathbf{S}_x^r / T)} \xrightarrow{P} \frac{\Sigma}{tr (\Sigma)}$$

$$\left[ \frac{\mathbf{S}_x^r / T}{tr (\mathbf{S}_x^r / T)} \right]^{-1} \xrightarrow{P} \left[ \frac{\Sigma}{tr (\Sigma)} \right]^{-1} \overset{P}{=} O(1)$$

Also

$$
\begin{aligned}
&\frac{\left\|\sum_{t\in I_r}\widetilde{\mathbf{x}}_{t,r}\widetilde{\mathbf{x}}'_{t,r}D^2\left[m(\boldsymbol{\xi}_{tr})\right]\widetilde{\mathbf{x}}_{t,r}\right\|}{tr\left(\mathbf{S}^r_x\right)} \\
&\leq \frac{\sum_{t\in I_r}\left\|\widetilde{\mathbf{x}}_{t,r}\right\|\left\|\widetilde{\mathbf{x}}'_{t,r}D^2\left[m(\boldsymbol{\xi}_{tr})\right]\widetilde{\mathbf{x}}_{t,r}\right\|}{tr\left(\mathbf{S}^r_x\right)} \\
&\leq \frac{\sum_{t\in I_r}\left\|\widetilde{\mathbf{x}}_{t,r}\right\|\left\|\widetilde{\mathbf{x}}'_{t,r}\widetilde{\mathbf{x}}_{t,r}\right\|}{tr\left(\mathbf{S}^r_x\right)}Const \\
&\leq \sqrt{T}h\frac{\sum_{t\in I_r}\left\|\widetilde{\mathbf{x}}'_{t,r}\widetilde{\mathbf{x}}_{t,r}\right\|}{tr\left(\mathbf{S}^r_x\right)}Const \\
&= O\left(\sqrt{T}h\right)
\end{aligned}
$$

Hence $R^{(1)}_r \overset{P}{=} o\left(\sqrt{T}\right)$ since by assumption $\sqrt{T}h \to 0$.

**Lemma 3** *We have for a given* $p$,

$$
\sum_{r=1}^{k} w_r D[m(\bar{\mathbf{x}}_r)]I\{T_r > p\} = \frac{1}{T}\sum_{t=1}^{T} D[m(\mathbf{x}_t)] + R^{(2)} + R^{(3)}
$$

*where* $R^{(2)}, R^{(3)} \overset{P}{=} o\left(T^{-\frac{1}{2}}\right)$.

**Proof of Lemma 3:** Let us define

$$
\begin{aligned}
R^{(2)} &= \sum_{r=1}^{k} w_r D[m(\bar{\mathbf{x}}_r)]I\{T_r > p\} - \sum_{r=1}^{k} w_r D[m(\bar{\mathbf{x}}_r)]I\{T_r > 0\} \\
R^{(3)} &= \sum_{r=1}^{k} w_r D[m(\bar{\mathbf{x}}_r)]I\{T_r > 0\} - \frac{1}{T}\sum_{t=1}^{T} D[m(\mathbf{x}_t)]
\end{aligned}
$$

then

$$
\sum_{r=1}^{k(T)} w_r D[m(\bar{\mathbf{x}}_r)]I\{T_r > p\} = \frac{1}{T}\sum_{t=1}^{T} D[m(\mathbf{x}_t)] + R^{(2)} + R^{(3)}
$$

We shall now show $R^{(2)}, R^{(3)} \overset{P}{=} o\left(T^{-\frac{1}{2}}\right)$

$$
\begin{aligned}
\left\|R^{(2)}\right\| &= \left\|\sum_{r=1}^{k} w_r D[m(\bar{\mathbf{x}}_r)]\left(I\{T_r > p\} - I\{T_r > 0\}\right)\right\| \\
&= \left\|\sum_{r=1}^{k} w_r D[m(\bar{\mathbf{x}}_r)]I\{p > T_r > 0\}\right\| \\
&\leq M_1\left\|\sum_{r=1}^{k} w_r I\{p > T_r > 0\}\right\|,
\end{aligned}
$$

where $M_1$ is the upper bound for $D[m(\bar{\mathbf{x}}_r)]$. Since the random variable$\sum_{r=1}^{k} w_r I\{p > T_r > 0\}$ is positive we have to show that

$$E\left(\sum_{r=1}^{k} w_r I\{p \geq T_r > 0\}\right) \to 0$$

therefore

$$\begin{aligned} E\left(\sum_{r=1}^{k} w_r I\{p \geq T_r > 0\}\right) &= \sum_{r=1}^{k} E\left(w_r I\{p \geq T_r > 0\}\right) \\ = \frac{1}{T}\sum_{r=1}^{k} E\left(T_r I\{p \geq T_r > 0\}\right) &= \frac{1}{T}\sum_{r=1}^{k}\sum_{j=1}^{p} j p_r^j \left(1 - p_r\right)^{T-j} \\ = \frac{1}{T}\sum_{r=1}^{k} p_r \left\{\sum_{j=1}^{p} j p_r^{j-1} \left(1 - p_r\right)^{T-j}\right\} \end{aligned}$$

Since $\sum_{j=1}^{p} j p_r^{j-1} \left(1 - p_r\right)^{T-j}$is bounded we have

$$E\left(\sum_{r=1}^{k} w_r I\{p > T_r > 0\}\right) = O\left(T^{-1}\right)$$

hence $\left\|R^{(2)}\right\| = o\left(T^{-1}\right)$

Now let us consider

$$\begin{aligned} \left\|R^{(3)}\right\| &= \left\|\sum_{r=1}^{k} w_r D[m(\bar{\mathbf{x}}_r)] I\{T_r > 0\} - \frac{1}{T}\sum_{t=1}^{T} D[m(\mathbf{x}_t)]\right\| \\ &\Box \left\|\sum_{r=1}^{k} w_r \left(D[m(\bar{\mathbf{x}}_r)] - \frac{1}{T_r}\sum_{i\in I_r} D[m(\mathbf{x}_t)]\right) I\{T_r > 0\}\right\| \\ &\Box \sum_{r=1}^{k} w_r \left\|\frac{1}{T_r}\sum_{i\in I_r} \left(D[m(\bar{\mathbf{x}}_r)] - D[m(\mathbf{x}_t)]\right)\right\| I\{T_r > 0\} \\ &\Box \sum_{r=1}^{k} w_r \frac{1}{T_r}\sum_{i\in I_r} \left\|D^2\left[m(\xi'_{t,r})\right]\right\| \left\|\mathbf{x}_t - \bar{\mathbf{x}}_r)\right\| I\{T_r > 0\} \\ &= O\left(\sum_{r=1}^{k} w_r \frac{1}{T_r}\sum_{i\in I_r} \left\|\mathbf{x}_t - \bar{\mathbf{x}}_r)\right\|\right) \\ &= O\left(\sum_{r=1}^{k} w_r h\right) = O\left(h\right) \ \text{(Since } \sum_{r=1}^{k} w_r = 1) \end{aligned}$$

then $\left\|R^{(3)}\right\| = o\left(T^{-\frac{1}{2}}\right)$ since by assumption $h = o\left(T^{-\frac{1}{2}}\right)$.

**Lemma 4** *Under the assumptions, we have*

$$\sup_{1<r<k} \sqrt{T} \left\| \boldsymbol{\theta}_r - \Sigma^{-1} S_{xu} \right\| \xrightarrow{P} 0$$

*where*

$$\boldsymbol{\theta}_r = [S_x^r]^{-1} \sum_{t=1}^{T} (\mathbf{x}_t - \overline{\mathbf{x}_t})\, I\{\mathbf{x}_t \in H_r\}\, u_t$$

$$\textit{and } S_{xu} = \frac{1}{T} \sum_{t=1}^{T} (\mathbf{x}_t - \overline{\mathbf{x}_t})\, u_t.$$

**Proof of Lemma 4:** We have,

$$\begin{aligned} \boldsymbol{\theta}_r &= [\mathbf{S}_x^r]^{-1} S_{xu}^r \\ &= \left[ \frac{1}{T_r} \sum_{t \in I_r} \widetilde{\mathbf{x}}_{t,r} \widetilde{\mathbf{x}}_{t,r}^T \right]^{-1} \frac{1}{T_r} \sum_{t=1}^{T} \widetilde{\mathbf{x}}_{t,r} u_t \end{aligned}$$

where

$$\widetilde{\mathbf{x}}_{t,r} = (\mathbf{x}_t - \overline{\mathbf{x}_t})\, I\{\mathbf{x}_t \in H_r\}$$

This can be written as

$$\square \frac{1}{w_r} \mathbf{S}_x^r \Big]^{-1} \frac{1}{w_r} \frac{1}{T} \sum_{t=1}^{T} \widetilde{\mathbf{x}}_{t,r} u_t$$

we have from previous lemmas

$$\frac{\mathbf{S}_x^r}{T} \stackrel{P}{=} p_r \Sigma \text{ and } w_r \stackrel{P}{=} p_r$$

by assumption as $p_r > 0$, for all $r$ and $\Sigma$ is positive semi-definite, we have

$$\square \frac{1}{w_r} \mathbf{S}_x^r \Big]^{-1} \xrightarrow{P} \Sigma^{-1}, \text{ uniformly} \tag{6}$$

also notice that

$$\begin{aligned} &\sqrt{T} \frac{1}{T^2} E \left\| \frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t u_t I\{\mathbf{x}_t \in H_r\} - \frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t u_t p_r \right\|^2 \\ &= \sqrt{T} \frac{1}{T^2} E \left( \sum_{t=1}^{T} \|\mathbf{x}_t\| \left| I\{\mathbf{x}_t \in H_r\} - p_r \right| u_t \right)^2 \\ &\square\ \sqrt{T} \frac{1}{T^2} Const \left[ \begin{array}{c} \sum_{t=1}^{T} \left[ E\left(I\{\mathbf{x}_t \in H_r\} - p_r\right)^2 E\left(u_t^2 | \mathbf{x}_t\right) \right] \\ + \sum_{t<t'=1}^{T} E\left[ \left| I\{\mathbf{x}_t \in H_r\} - p_r \right| E \left| I\{\mathbf{x}_t \in H_r\} - p_r \right| E\left( |u_t u_{t'}| | \mathbf{x}_t \right) \right] \end{array} \right] \end{aligned}$$

following similar steps in the proof of lemma (1)

$$\sqrt{T}\frac{1}{T^2}E\left\|\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_tu_tI\left\{\mathbf{x}_t\in H_r\right\}-\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_tu_tp_r\right\|^2=O\left(\frac{1}{\sqrt{T}}p_r+p_r\sqrt{T}h\right)$$

since $E\left(u_t^2|\,\mathbf{x}_t\right)$ and $E\left(|u_tu_{t'}|\,|\,\mathbf{x}_t\right)$ are bounded. Therefore

$$\frac{1}{\sqrt{T}}\sum_{r=1}^{k}E\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{x}_t\left(I\left\{\mathbf{x}_t\in H_r\right\}-p_r\right)u_t\right\|^2=O\left(\frac{1}{\sqrt{T}}+\sqrt{T}h\right)$$

Using the same techniques as in lemma 1, we proof that

$$\sup_{1<r<k}\left\|\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_tu_tI\left\{\mathbf{x}_t\in H_r\right\}-p_r\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_tu_t\right\|\xrightarrow{P}0$$

we can proof similarly for

$$\sup_{1<r<k}\left\|\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_tI\left\{\mathbf{x}_t\in H_r\right\}-p_r\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\right\|\xrightarrow{P}0$$

Therefore

$$\sup_{1<r<k}\left\|S_{xu}^r-S_{xu}\right\|\xrightarrow{P}0$$

Combining with (6) we get the result.

**Lemma 5** *Under the assumptions,*

$$\sqrt{T}\sum_{r=1}^{k}w_r\boldsymbol{\theta}_rI\{T_r>p\}\xrightarrow{D}N\left(0,\sigma_u^2\Sigma^{-1}\right).$$

**Proof of Lemma 5:** Using the previous lemma we can show that

$$\sqrt{T}\boldsymbol{\theta}_rI\{T_r>p\}\overset{P}{=}\sqrt{T}\Sigma^{-1}S_{xu}I\{T_r>p\}+\sqrt{T}R_r^{(4)},$$

where $\sup_{1\square r\square k}R_r^{(4)}\overset{P}{=}o\left(T^{-\frac{1}{2}}\right)$. After multiplying and both sides by $w_r's$ and summing across all the bins we have

$$\sqrt{T}\sum_{r=1}^{k}w_r\boldsymbol{\theta}_rI\{p>T_r\}\overset{P}{=}\sqrt{T}\Sigma^{-1}S_{xu}\sum_{r=1}^{k}w_rI\{T_r>p\}+\sqrt{T}\sum_{r=1}^{k}w_rI\{T_r>p\}R_r^{(4)}.$$

We can easily show that $\left|\sum_{r=1}^{k}w_rI\{T_r>p\}-1\right|\xrightarrow{P}0$, and since

$$\left\|\sum_{r=1}^{k}w_rI\{T_r>p\}R_r^{(4)}\right\|\ \square\ \left\|\sup_{1\square r\square k}R_r^{(4)}\right\|$$

We have $\left\|\sum_{r=1}^{k} w_r I\{T_r > p\} R_r^{(4)}\right\| \xrightarrow{P} 0.$

Observe that by central limit theorem and as $\overline{u} \xrightarrow{P} 0$ we have,

$$\begin{aligned}\sqrt{T} S_{xu} &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\mathbf{x}_t - \boldsymbol{\mu}_1) u_t - \sqrt{T} (\overline{\mathbf{x}_t} - \boldsymbol{\mu}_1) \overline{u} \\ &\overset{D}{=} N\left(0, \Sigma\sigma^2\right).\end{aligned}$$

Therefore

$$\sqrt{T} \Sigma^{-1} S_{xu} \overset{D}{=} N\left(0, \Sigma^{-1}\sigma^2\right).$$

giving us

$$\sqrt{T} \sum_{r=1}^{k} w_r \boldsymbol{\theta}_r I\{T_r > p\} \xrightarrow{D} N\left(0, \sigma_u^2 \Sigma^{-1}\right).$$

**Proof of Theorem 1**: From lemma (2) we can write

$$\begin{aligned}\sqrt{T}\widehat{\delta} &= \sqrt{T} \sum_{r=1}^{k} w_r \widehat{\boldsymbol{\beta}}_r I\{T_r > p\} \\ &= \sqrt{T} \sum_{r=1}^{k} w_r \left(D\left[m(\bar{\mathbf{x}}_r)\right] + R_r^{(1)}\right) I\{T_r > p\} + \sqrt{T} \sum_{r=1}^{k} w_r I\{T_r > p\} \boldsymbol{\theta}_r\end{aligned}$$

where $\sup_{1 \Box r \Box k} R_r^{(1)} \overset{P}{=} o\left(T^{-\frac{1}{2}}\right)$ (since by assumption $\sqrt{T} h \to 0$).Therefore $\sum_{r=1}^{k} w_r R_r^{(1)} I\{T_r > p\} \overset{P}{=} o\left(T^{-\frac{1}{2}}\right)$. Further using lemma 3 we can write

$$\sqrt{T} \sum_{r=1}^{k} w_r D\left[m(\bar{\mathbf{x}}_r)\right] I\{T_r > p\} \overset{P}{=} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} D[m(\mathbf{x}_t)] + \sqrt{T} R^{(2)} + \sqrt{T} R^{(3)}.$$

As $\left(\sqrt{T} R^{(2)}, \sqrt{T} R^{(3)}, \sum_{r=1}^{k} w_r R_r^{(1)} I\{T_r > p\}\right) \overset{P}{=} o\left(T^{-\frac{1}{2}}\right)$ ,we have

$$\sqrt{T} \sum_{r=1}^{k} w_r \widehat{\boldsymbol{\beta}}_r I\{T_r > p\} \overset{P}{=} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} D[m(\mathbf{x}_t)] + \sqrt{T} \sum_{r=1}^{k} w_r I\{T_r > p\} \boldsymbol{\theta}_r.$$

Using central limit theorem we show that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} D[m(\mathbf{x}_t)] \overset{D}{=} N\left(0, Var\left(D\left[m\left(\mathbf{x}\right)\right]\right)\right)$$

and by lemma (5) we have

$$\sqrt{T} \sum_{r=1}^{k} w_r \boldsymbol{\theta}_r I\{T_r > p\} \xrightarrow{D} N\left(0, \sigma_u^2 \Sigma^{-1}\right).$$

Since $\sum_{r=1}^{k} w_r \boldsymbol{\theta}_r I\{T_r > p\}$and $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} D[m(\mathbf{x}_t)]$ are uncorrelated by assumption, we prove that

$$\sqrt{T} \sum_{r=1}^{k} w_r \widehat{\boldsymbol{\beta}}_r I\{T_r > p\} \xrightarrow{D} N\left(0, Var\left(D\left[m\left(\mathbf{x}\right)\right]\right) + \sigma_u^2 \Sigma^{-1}\right)$$

**Proof of Theorem 2:** We shall use lemma (2) and the fact that $x \to xx'$ is continuous mapping, to show that

$$\widehat{\boldsymbol{\beta}}_r \widehat{\boldsymbol{\beta}}_r' \overset{P}{=} D\left[m(\bar{\mathbf{x}}_r)\right] D\left[m(\bar{\mathbf{x}}_r)\right]' + \boldsymbol{\theta}_r \boldsymbol{\theta}_r'.$$

Then we use the proof of lemma (3) and the fact $D\left[m(\bar{\mathbf{x}}_r)\right] D\left[m(\bar{\mathbf{x}}_r)\right]'$ is differentiable to get

$$\sum_{r=1}^{k} w_r \widehat{\boldsymbol{\beta}}_r \widehat{\boldsymbol{\beta}}_r' I\{T_r > p\} \overset{P}{=} \frac{1}{T} \sum_{t=1}^{T} D\left[m(\mathbf{x}_t)\right] D\left[m(\mathbf{x}_t)\right]' + \sum_{r=1}^{k} w_r \boldsymbol{\theta}_r \boldsymbol{\theta}_r' I\{T_r > p\}$$

By weak law of large numbers we have,

$$\frac{1}{T} \sum_{t=1}^{T} D\left[m(\mathbf{x}_t)\right] D\left[m(\mathbf{x}_t)\right]' \xrightarrow{P} E\left(D\left[m(\mathbf{x})\right] D\left[m(\mathbf{x})\right]'\right)$$

As in lemma (4) we can show that

$$\sup_{1<r<k} T \left\| \boldsymbol{\theta}_r \boldsymbol{\theta}_r' - \Sigma^{-1} S_{xu} S_{xu}' \Sigma^{-1} \right\| \xrightarrow{P} 0,$$

and again by weak law of large numbers we have,

$$T S_{xu} S_{xu}' \xrightarrow{P} \sigma_u^2 \Sigma \text{ as } E\left(S_{xu} S_{xu}'\right) = \sigma_u^2 \Sigma$$

implying

$$\sup_{1<r<k} \left\| T \boldsymbol{\theta}_r \boldsymbol{\theta}_r' - \sigma_u^2 \Sigma^{-1} \right\| \xrightarrow{P} 0$$

Therefore

$$\sum_{r=1}^{k} w_r \boldsymbol{\theta}_r \boldsymbol{\theta}_r' I\{T_r > p\} \overset{P}{=} \sigma_u^2 \Sigma^{-1} \sum_{r=1}^{k} w_r I\{T_r > p\}$$

$$\overset{P}{=} \sigma_u^2 \Sigma^{-1} \text{ since}$$

$$\sum_{r=1}^{k} w_r I\{T_r > p\} \overset{P}{=} 1$$

From previous theorem we also know that

$$\widehat{\delta} \stackrel{P}{\rightarrow} \delta \quad = E\left(D\left[m(\mathbf{x})\right]\right)$$

Hence

$$\widehat{\mathbf{V}} \xrightarrow{P} Var\left(D\left[m\left(x\right)\right]\right) + \sigma_u^2 \Sigma^{-1}.$$