A REGRESSION-TYPE ESTIMATOR BASED ON PRELIMINARY TEST OF SIGNIFICANCE

J. E. Grimes and B. V. Sukhatme
California Polytechnic State University and Iowa State University

1. Introduction. If data on an auxiliary characteristic X correlated with the characteristic Y under study is available, then it is customary to use this data to provide a more efficient estimate of $\bar{Y}$, the population mean. If Y and X are correlated and the relationship between the two variables is linear, but the relationship does not pass through the origin or the correlation between Y and X is not sufficiently high, quite often a regression type estimator is used. A frequently used estimator of this type is the so-called difference estimator suggested by Hansen, Horwitz and Madow (1953), defined as

$$\bar{y}_d = \bar{y} + \beta_0(\bar{X}-\bar{x}), \tag{1.1}$$

where β_0 is a fixed constant, assumed to be known, $\bar{x}$ and $\bar{y}$ are the mean per unit estimates of $\bar{X}$ and $\bar{Y}$, and $\bar{X}$ is the population mean of X. The value of β_0 that minimizes $V(\bar{y}_d)$ is easily seen to be $\beta_2 = \sigma_{12}/\sigma_1^2$, the regression coefficient of Y on X. If no reliable guess can be made about the value of the regression coefficient, the usual practice is to estimate it from the sample by

$$\hat{\beta}_2 = s_{12}/s_1^2 \tag{1.2}$$

where $s_{12} = \sum_i^n (x_i-\bar{x})(y_i-\bar{y})/(n-1)$,

and $s_1^2 = \sum_i^n (x_i-\bar{x})^2/(n-1)$.

and use as an estimator of $\bar{Y}$, the regression estimator $\bar{y}_\ell$ defined as,

$$\bar{y}_\ell = \bar{y} + \hat{\beta}_2(\bar{X}-\bar{x}) \ . \tag{1.3}$$

The difference estimator $\bar{y}_d$ is an unbiased estimator of the population mean $\bar{Y}$ and its variance is given by,

$$V(\bar{y}_d) = \sigma_2^2(1-\rho^2)(1+\delta^2)/n \tag{1.4}$$

where σ_1^2 and σ_2^2 are the variances of X and Y, σ_{12} is the covariance between X and Y, ρ is the correlation coefficient between X and Y and

$$\delta = \left(\rho - \frac{\beta_0\sigma_1}{\sigma_2}\right)/(1-\rho^2)^{1/2} \ . \tag{1.5}$$

The regression estimator on the other hand is generally biased, the bias vanishing when the relationship between Y and X is linear. Further its variance to terms of order n^{-2} is given by

$$V(\bar{y}_\ell) = \sigma_2^2(1-\rho^2)(n-2)/n(n-3) \ . \tag{1.6}$$

From past experience, we are often able to make an intelligent guess about β_2 the regression coefficient of Y on X. Let β_0 denote the guessed value of β_2. If β_0 is relatively close to β_2, it would appear from the above that $\bar{y}_d$ is more appropriate than $\bar{y}_\ell$ as an estimator of $\bar{Y}$, otherwise $\bar{y}_\ell$ would be more appropriate. We therefore propose an estimator which chooses between $\bar{y}_\ell$ and $\bar{y}_d$, based on a preliminary test of significance of the relative closeness of β_0 to β_2 and investigate its efficiency with respect to other regression-type estimators currently in use.

2. Proposed Regression-Type Estimator. A common method of making a test of the relative closeness of β_2 to β_0 is the usage of the statistic,

$$t = \sqrt{n-2}(\hat{\beta}_2-\beta_0)s_1 / (s_2^2 - \hat{\beta}_2^2 s_1^2)^{1/2} \tag{2.1}$$

where $s_2^2 = \sum_i^n (y_i-\bar{y})^2/(n-1)$. (2.2)

If from past experience, it is hypothesized that β_2 is β_0 but nothing further is known about β_2, the proposed estimator based on preliminary test of significance, to be called Sometimes Regression Estimator, may be defined as

$$\begin{aligned} \bar{y}_s &= \bar{y}_d && \text{if } t \in A \\ &= \bar{y}_\ell && \text{if } t \in A^c \end{aligned} \tag{2.3}$$

where A is the event $|t| \le t_0$ and A^c the complementary event $|t| > t_0$.

Now we need to look at a criterion for deciding whether or not the proposed estimator $\bar{y}_s$ has any advantages over $\bar{y}_d$ and $\bar{y}_\ell$. The most commonly used loss function is the squared error. This then leads to considering the variance of the estimator $\bar{y}_s$ if it is unbiased, or the mean square error of $\bar{y}_s$ if it is biased. We then have the expected value of $\bar{y}_s$ given by

$$E(\bar{y}_s) = E(\bar{y}_d|A)P(A) + E(\bar{y}_\ell|A^c)P(A^c), \tag{2.4}$$

and the mean square error of $\bar{y}_s$ is given by

$$\begin{aligned} M.S.E.(\bar{y}_s) &= E(\bar{y}_s-\bar{Y})^2 = E[(\bar{y}_d-\bar{Y})^2|A]P(A) \\ &+ E[(\bar{y}_\ell-\bar{Y})^2|A^c]P(A^c) \ . \end{aligned} \tag{2.5}$$

3. Expected Value and Variance of $\bar{y}_s$. It is necessary to make suitable assumptions about the joint distribution of X and Y in order to obtain a closed form for the expected value and the variance of $\bar{y}_s$. In what follows, we assume that the population is infinite and that X and Y have a bivariate normal distribution function.

Theorem 3.1: $\bar{y}_s$ is an unbiased estimator of the population mean $\bar{Y}$.

Proof: Using the fact that $\bar{x}$ and (s_1^2, s_2^2, s_{12}) are statistically independent, it can be easily seen that $E(\bar{y}_s) = \bar{Y}$. Q.E.D.

Since $\bar{y}_s$ is an unbiased estimator, we now obtain the variance of $\bar{y}_s$. As $(\bar{x},\bar{y})$ and (s_1^2, s_2^2, s_{12}) are statistically independent, we have from (2.5)

$$V(\bar{y}_s) = V(\bar{y}_d) - \frac{2\sigma_{12}}{n} E[(\hat{\beta}_2 - \beta_0)|A^c]P(A^c) + \frac{2\beta_0\sigma_1^2}{n} E[(\hat{\beta}_2 - \beta_0)|A^c]P(A^c) + \frac{\sigma_1^2}{n} E[(\hat{\beta}_2 - \beta_0)^2|A^c]P(A^c) \quad . \tag{3.1}$$

In order to further evaluate this, we need an expression for $E[(\hat{\beta}_2 - \beta_0)^h|A^c]P(A^c)$ for $h = 0,1,2$. It will be assumed that the sample size is $n \geq 4$.

Lemma 3.2: $KP(|t|>t_0)E[(\hat{\beta}_2-\beta_0)^h||t|>t_0]$

$$= \sum_{i=0}^{\infty} (2\theta)^{2i} \frac{\Gamma(\frac{h+2i+1}{2})\Gamma(\frac{n+2i-h-1}{2})}{\Gamma(2i+1)} I(h+2i+1) \text{ if } h \text{ is even,}$$

$$= \sum_{i=0}^{\infty} (2\theta)^{2i+1} \frac{\Gamma(\frac{h+2i+2}{2})\Gamma(\frac{n+2i-h}{2})}{\Gamma(2i+2)} I(h+2i+2), \text{ if } h$$

is odd where $m_0 = (n-2)/[t_o^2 + (n-2)]$,

$$K = \sqrt{\pi}\,\Gamma(\tfrac{n-1}{2})(1+\delta^2)^{\frac{n-h-1}{2}} (\sigma_1/\sigma_2\sqrt{1-\rho^2})^h, \quad \theta = \frac{\delta}{\sqrt{1+\delta^2}},$$

$I.(\cdot,\cdot)$ is the incomplete beta distribution function and $I(x)$ denotes $I_{m_0}(\frac{n-2}{2}, \frac{x}{2})$.

Proof: It is well-known that the joint density function for s_1, s_2 and $r = s_{12}/s_1s_2$ is given by

$$f(s_1,s_2,r) = K_1(s_1^2s_2^2)^{\frac{n-2}{2}}(1-r^2)^{\frac{n-4}{2}} \times \exp[-\frac{n-1}{2(1-\rho^2)}(\frac{s_1^2}{\sigma_1^2} - \frac{2\rho s_1s_2r}{\sigma_1\sigma_2} + \frac{s_2^2}{\sigma_2^2})]$$

if $0<s_1<\infty$, $0<s_2<\infty$, and $r^2<1$,

= 0 otherwise,

where $K_1 = (n-1)^{n-1}/\pi\Gamma(n-2)[(1-\rho^2)\sigma_1^2\sigma_2^2]^{\frac{n-1}{2}}$.

Making the transformation

$u = (n-1)s_1^2/2\sigma_1^2(1-\rho^2)$, $v = (n-1)rs_1s_2/2\sigma_1\sigma_2(1-\rho^2)$ and

$t' = t/\sqrt{n-2} = (\hat{\beta}_2-\beta_0)s_1/(s_2^2-\hat{\beta}_2^2s_1^2)^{1/2}$, we get

$$f(u,v,t') = \frac{K_3}{ut'^{n-1}} \exp[-u(1-\rho^2)(1+\delta^2) - \frac{1+t'^2}{ut'^2}(v - \frac{\beta_0u\sigma_1}{\sigma_2})^2] \times \sum_{i=0}^{\infty} \frac{2^i(v-\frac{\beta_0u\sigma_1}{\sigma_2})^{n+i-2}\cdot\delta^i(1-\rho^2)^{\frac{i}{2}}}{\Gamma(i+1)}$$

in $R_1 = (0\leq u<\infty,\ 0\leq t'<\infty,\ v\geq u\sigma_1\beta_0/\sigma_2)$

$$= \frac{K_3}{u|t'|^{n-1}} \exp[-u(1-\rho^2)(1+\delta^2) - \frac{1+t'^2}{ut'^2}(v-\frac{\beta_0u\sigma_1}{\sigma_2})^2] \times \sum_{i=0}^{\infty} \frac{(-1)^i2^i\left|v-\frac{\beta_0u\sigma_1}{\sigma_2}\right|^{n+i-2}\delta^i(1-\rho^2)^{\frac{i}{2}}}{\Gamma(i+1)},$$

in $R_2 = (0\leq u<\infty,\ -\infty<t'<0,\ v< u\sigma_1\beta_0/\sigma_2)$

= 0, otherwise,

where $K_3 = 2^{n-2}(1-\rho^2)^{\frac{n-1}{2}}/\pi\Gamma(n-2)$.

We have $P(|t'|>t_0')E((\hat{\beta}_2-\beta_0)^h\big|\,|t'|>t_0'$

$$= \int_{R_4}[(v-\frac{\beta_0u\sigma_1}{\sigma_2})\frac{\sigma_2}{u\sigma_1}]^h f(u,v,t')du\,dv\,dt' + \int_{R_5}[(v-\frac{\beta_0u\sigma_1}{\sigma_2})\frac{\sigma_2}{u\sigma_1}]^h f(u,v,t')du\,dv\,dt'$$

$$= I_4 + I_5$$

where $R_4 = \{0<u<\infty,\ -\infty<t'<t_0',\ v<u\sigma_1\beta_0/\sigma_2\}$, and

$R_5 = \{0\leq u\leq\infty,\ v\geq u\sigma_1\beta_0/\sigma_2,\ t_0'<t'<\infty\}$.

To obtain the desired result the following lemmas are needed.

Lemma 3.3: $\int_{-\infty}^{0}|x|^n e^{-\frac{1}{2}x^2} = 2^{\frac{n-1}{2}}\Gamma(\frac{n+1}{2})$.

Lemma 3.4: $\Gamma(\frac{j-2}{2})\,\Gamma(\frac{j-1}{2})\,2^{j-3} = \Gamma(j-2)\sqrt{\pi}$.

Now, $I_4 = K_3 \int_{-\infty}^{t_0'}\int_{0}^{\infty}\int_{-\infty}^{u\beta_0\sigma_1/\sigma_2} (\frac{\sigma_2}{\sigma_1})^h (\frac{1}{u})^{h+1} \frac{1}{|t'|^{n-1}}$

$$x \exp[-u(1-\rho^2)(1+\delta^2) - \frac{1+t'^2}{ut'^2}(v-\beta)^2]$$

$$\sum_{i+0}^{\infty} (-1)^{h+i} \frac{\left|v - \frac{\beta_0 u\sigma_1}{\sigma_2}\right|^{n+h+i-2} (2\delta\sqrt{1-\rho^2})^i dv\,du\,dt'}{\Gamma(i+1)}$$

$$= \frac{1}{2K} \sum_{i=0}^{\infty} (-1)^{h+i} (2\theta)^i \frac{\Gamma(\frac{h+i+1}{2})\Gamma(\frac{n+i-h-1}{2})}{\Gamma(i+1)} I(h+i+1).$$

Similarly I_5 can be obtained. Q.E.D.

Using Lemma 3.2 and substituting into (3.1) we obtain the following theorem.

Theorem 3.5: $V(\bar{y}_s) - V(\bar{y}_d)$

$$= \frac{2\sigma_2^2(1-\rho^2)}{n} \sum_{i=0}^{\infty} \frac{\Gamma(\frac{n+2i-1}{2})\delta^{2i+2}}{\Gamma(i+1)\Gamma(\frac{n-1}{2})(1+\delta^2)^{\frac{n+2i-1}{2}}} I(2i+3)$$

$$+ \frac{\sigma_2^2(1-\rho^2)}{n} \sum_{i=0}^{\infty} \frac{(2i+1)\Gamma(\frac{n+2i-3}{2})\delta^{2i}}{2\Gamma(i+1)\Gamma(\frac{n-1}{2})(1+\delta^2)^{\frac{n+2i-3}{2}}} I(2i+3) .$$

As t_0 tends to infinity, $V(\bar{y}_s)$ tends to $V(\bar{y}_d)$ as is to be expected since the estimator $\bar{y}_s$ becomes $\bar{y}_d$. Similarly, as t_0 tends to zero, $V(\bar{y}_s)$ tends to $V(\bar{y}_\ell)$ since the estimator $\bar{y}_s$ becomes $\bar{y}_\ell$.

4. Comparison of Different Estimators

A. Comparison of the sometimes regression estimator with the difference estimator.

Consider

$$D_2(\theta,m_0) = n\Gamma(\tfrac{n-1}{2})(1+\delta^2)^{\frac{n-3}{2}}(V(\bar{y}_s) - V(\bar{y}_d))/\sigma_2^2(1-\rho^2) \tag{4.1}$$

Then, we have from Theorem 3.5

$D_2(\theta,m_0)$

$$= \sum_{j=0}^{\infty} \frac{\Gamma(\frac{n+2j-3}{2})\theta^{2j}}{2\Gamma(j+1)} I(2j+3)[(2j+1)-2(n+2j-3)\theta^2] . \tag{4.2}$$

Let $j=i-1$ in the first summation of (4.2), then we have $D_2(\theta,m_0) = \frac{1}{2}\Gamma(\frac{n-3}{2})I(3)$

$$+ 2\sum_{j=0}^{\infty} \frac{\Gamma(\frac{n+2j-1}{2})\theta^{2j+2}}{\Gamma(j+1)} I(2j+5)[\frac{2j+3}{4(j+1)} - \frac{I(2j+3)}{I(2j+5)}] \tag{4.3}$$

$$= \sum_{j=0}^{\infty} C_j(m_0)\theta^{2j} , \tag{4.4}$$

where

$$C_0(m_0) = \frac{1}{2}\Gamma(\tfrac{n-3}{2})I_{m_0}(\tfrac{n-2}{2}, \tfrac{3}{2}) \tag{4.5}$$

and

$$C_{j+1}(m_0) = \frac{2\Gamma(\frac{n+2j-1}{2})I(2j+5)}{\Gamma(j+1)} [\frac{2j+3}{4(j+1)} - \frac{I(2j+3)}{I(2j+5)}],\ j=0,1,2,\ldots . \tag{4.6}$$

Consider first the effect of variation in θ. θ will vary over the interval $(-1,1)$ since δ may vary over the interval $(-\infty, \infty)$.

Lemma 4.1: For $a = 1, \frac{3}{2}, 2, \frac{5}{2} \ldots$ and $c = 1, \frac{3}{2}, 2, \frac{5}{2}, \ldots$

$$\frac{I_x(a,c)}{I_x(a,c+1)} \Big/ \frac{I_x(a,c+\frac{1}{2})}{I_x(a,c+\frac{3}{2})} \le 1, \quad \text{for} \quad 0 < x \le 1 .$$

Proof: L'Hospital's rule may be used to show that the lemma holds in a positive neighborhood of zero. Then the lemma may be proved for the entire interval by defining $\phi(x) = I_x(a, c+1)$. $I_x(a,c+\frac{1}{2}) - I_x(a,c)I_x(a,c+\frac{3}{2})$, and showing that there exists an x_1 such that

$$\phi'(x) \ge 0 \quad 0 < x \le x_1 ,$$
$$< 0 \quad x_1 < x \le 1 .$$

Lemma 4.2: For $0 < m_0 \le 1$ $C_0(m_0)$, $C_1(m_0)$, $C_2(m_0)$, ... is a sequence of numbers such that for some $K > 0$

$$C_j(m_0) \ge 0 \quad j \le K$$
$$< 0 \quad j > K .$$

Proof: Since $(2j+3)/4(j+1)$ is a decreasing function of j, and by Lemma 4.1, $I(2j+3)/I(2j+5)$ $j = 0,1,2,\ldots$ is an increasing function of j, $(2j+3)/4(j+1) - I(2j+3)/I(2j+5)$ $j = 0,1,2,\ldots$ is a decreasing function of j. Now, $C_0(m_0) > 0$. Suppose that $C_j(m_0) \ge 0$ for $j = 1,2,\ldots$. This implies that $D_2(\theta, m_0) \ge 0$ for $0 \le \theta < 1$. But from (4.2) for $\theta = 1\sqrt{2} + \varepsilon$ with $\varepsilon > 0$, $D_2(1/\sqrt{2} + \varepsilon, m_0) < 0$ for $0 < m_0 \le 1$ leading to contradiction. Hence the lemma is proven. Q.E.D.

Define the relative efficiency of $\bar{y}_s$ with respect to $\bar{y}_d$ as $e_2(\delta, m_0) = V(\bar{y}_d)/V(\bar{y}_s)$.

Theorem 4.3: For m_0 fixed such that $0 < m_0 \le 1$, there exists a θ_0 where $0 < \theta_0 < 1$ and $D_2(\theta,m_0) > 0$ and hence $e_2(\delta,m_0) < 1$, $-\theta_0 < \theta < \theta_0$; ≤ 0 and hence $e_2(\delta,m_0) \ge 1$ otherwise.

Proof: Since from (4.3) $D_2(\theta, m_0)$ is symmetric in θ, it is necessary only to consider $D_2(\theta, m_0)$ for θ positive.

From (4.2), $D_2(0, m_0) > 0$ for $0 < m_0 \leq 1$. Further for $\theta = 1\sqrt{2} + \epsilon$ with $\epsilon > 0$, $[(2j+1) - 2(n+2j-3)\theta^2] < 0 \quad j = 0,1,2,\ldots,$ and we have $D_2(1/\sqrt{2} + \epsilon, m_0) < 0, \ 0 < m_0 \leq 1$. Since $D_2(\theta, m_0)$ is continuous then there exists θ_0 such that $0 < \theta_0 < 1$ and $D_2(\theta_0, m_0) = 0$.

We now show that $D_2(\theta, m_0) < 0$ for $\theta > \theta_0$. By Lemma 4.2 there exists a K such that

$$C_j(m_0) > 0 \quad \text{for } j < K$$
$$\leq 0 \quad \text{for } j \geq K .$$

Hence $\sum_{j=0}^{\infty} C_j(m_0)\theta_0^{2j} = 0$ i.e., $\sum_{j=0}^{\infty} C_j(m_0)\theta_0^{2j-1} = 0$;

and since $D_2(\theta, m_0)$ is a power series in θ which converges for $-1 < \theta < 1$, we get

$$\frac{\partial D_2(\theta, m_0)}{\partial \theta} = \sum_{j=0}^{\infty} 2j \ C_j(m_0)\theta^{2j-1} \quad \text{for } 0 \leq \theta < 1 ,$$

and therefore $\left.\frac{\partial D_2(\theta, m_0)}{\partial \theta}\right|_{\theta_0}$

$$\leq 2K \sum_{j=1}^{\infty} C_j(m_0)\theta_0^{2j-1} = \frac{-2K \ C_0(m_0)}{\theta_0} < 0 .$$

It can be similarly shown that if $D_2(\theta^*, m_0) < 0$, then $\left.\frac{\partial D_2(\theta, m_0)}{\partial \theta}\right|_{\theta^*} < 0$. Therefore for m_0 fixed, as θ increases, $D_2(\theta, m_0)$ becomes negative and stays negative. Q.E.D.

Next consider the variation of $D_2(\theta, m_0)$ due to m_0 with θ fixed.

<u>Lemma 4.4</u>: If for fixed θ, there exists an $m_0^* \epsilon$ (0.1) such that

$$\left.\frac{\partial D_2(\theta, m_0)}{\partial m_0}\right|_{m^*} = 0$$

then
$$\frac{\partial D_2(\theta, m_0)}{\partial m_0} \begin{cases} > 0 & 0 \leq m_0 < m_0^* \\ = 0 & m_0 = m_0^* \\ < 0 & m_0^* < m_0 < 1 . \end{cases}$$

The proof of this lemma follows in a manner analogous to the proof of Theorem 4.3.

<u>Theorem 4.5</u>: There exists $\theta_1^* > 0$ and $\theta_2^* > 0$ defined by $D_2(\theta_1^*, 1) = 0$, and $\theta_2^* = \operatorname*{Inf}_S \theta$, where $S = \{\theta : \theta > 0, \ D_2(\theta, m_0) \leq 0$ for all $m_0 \ni 0 < m_0 \leq 1\}$; such that

a) for θ fixed and $\epsilon[-\theta_1^*, \theta_1^*]$ $D_2(\theta, m_0) \geq 0$ and hence $e_2(\delta, m_0) \leq 1$ for $0 < m_0 < 1$,

b) for θ fixed and $\epsilon\{(-\theta_2^*, -\theta_1^*) \cup (\theta_1^*, \theta_2^*)\}$, $\exists m_0^* \ni 0 < m_0^* < 1$, and
$D_2(\theta, m_0) \geq 0$ and hence
$$e_2(\delta, m_0) \leq 1 \qquad 0 < m_0 \leq m_0^* ,$$
$D_2(\theta, m_0) < 0$ and hence
$$e_2(\delta, m_0) > 1 \qquad m_0^* < m_0 \leq 1 ;$$

c) for θ fixed and $\epsilon\{(-1, -\theta_2^*] \cup [\theta_2^*, 1)\}$ $D_2(\theta, m_0) \leq 0$, and hence $e_2(\delta, m_0) \geq 1$ for $0 < m_0 \leq 1$.

<u>Proof</u>: Since $D_2(\theta, m_0)$ is symmetric in θ, it is necessary only to consider $D_2(\theta, m_0)$ for $\theta > 0$. Suppose for θ fixed $\ni 0 < \theta < 1, \exists m_0^* \ni m_0^* \in (0,1)$ and $D_2(\theta, m_0^*) = 0$. Since

$$\lim_{m_0 \to 0} D_2(\theta, m_0) = \lim_{m_0 \to 0} \frac{\partial D_2(\theta, m_0)}{\partial m_0} = 0,$$ it follows

from Lemma 4.4 that if $\frac{\partial D_2(\theta, m_0)}{\partial m_0} < 0$ in the

in the neighborhood of $m_0 = 0$, then $\frac{\partial D_2(\theta, m_0)}{\partial m_0} < 0$ $0 < m_0 \leq 1$. Under that condition there could not be a point $m_0^* \ni 0 < m_0^* \leq 1$ and $D_2(\theta, m_0) = 0$.

Hence in order that $D_2(\theta, m_0^*) = 0$ it follows that there must exist an m_0^{**} such that $0 < m_0^{**} < m_0^* \leq 1$ and

$$\frac{\partial D_2(\theta, m_0)}{\partial m_0} \begin{cases} > 0 & 0 < m_0 < m_0^{**} \\ = 0 & m = m_0^{**} \\ < 0 & m_0^{**} < m_0 \leq 1 . \end{cases}$$

Hence if $D_2(\theta, m_0^*) = 0$ then for $m_0 > m_0^*$, $D_2(\theta, m_0) < 0$. By above if for $\theta = \theta_1$, $D_2(\theta_1, 1) \geq 0$ then $D_2(\theta_1, m_0) \geq 0, \ 0 < m_0 \leq 1$. If further for $\theta = \theta_2$, $D_2(\theta_2, 1) < 0$, then by Theorem 4.3, $\theta_2 > \theta_1$. Hence $\theta_1^* = \{\theta : \theta > 0$ and $D_2(\theta, 1) = 0\}$. If $D_2(\theta, 1) < 0$ then either $\theta = \theta_3$ and $D_2(\theta_3, m_0) \leq 0$, $0 < m_0 \leq 1$ or $\theta = \theta_4$ and

$\exists\, m_0^* \ni D_2(\theta_4, m_0) > 0 \qquad 0 < m_0 < m_0^*$

$= 0 \qquad m_0 = m_0^*$

$< 0 \qquad m_0^* < m_0 \leq 1 .$

Now for $m_{0_1} < m_0^*$, $D_2(\theta_3, m_{0_1}) \leq 0$ and $D_2(\theta_4, m_{0_1}) \geq 0$, then by Theorem 4.3 $\theta_4 \leq \theta_3$.

Hence $\theta_2^* = \underset{S}{\mathrm{Inf}}\ \theta$ and theorem is proved. Q.E.D.

Theorem 4.6: For e_0 fixed such that $0 < e_0 < 1$, there exists an m_0^* such that for $m_0 \leq m_0^*$, $e_2(\delta, m_0) \geq e_0$.

Proof: By Lemma 4.4, for fixed θ or equivalently for fixed δ, $\exists\, m_0(\theta) \ni$

$$e_2(\delta, m_0) = 1 \Big/ \left[1 + \frac{D_2(\theta, m_0)}{\Gamma(\frac{n-1}{2})(1+\delta^2)^{\frac{n-1}{2}}}\right] \geq e_0 \quad \text{for}$$

$0 < m_0 \leq m_0(\theta)$. Pick $m_0^* = \underset{0 \leq \theta < 1}{\mathrm{Inf}} m_0(\theta)$. Hence

$e_2(\delta, m_0) \geq e_2(m_0^*, \delta) \geq e_0$ for $0 < m_0 \leq m_0^*$ and for any $\delta \in [0, \infty)$. Q.E.D.

B. Comparison of the Sometimes Regression Estimator with the Regression Estimator.

$$\text{Let } D_1(\theta, m_0) = n(V(\bar{y}_s) - V(\bar{y}_\ell)) \,/\, \sigma_2^2(1-\rho^2)(1-\theta^2)^{\frac{n-3}{2}} \tag{4.7}$$

Then using (1.4), (1.6), (4.1) and (4.3), it can be seen that

$$D_1(\theta, m_0) = \theta^2(1-\theta^2)^{-\frac{n-1}{2}} - \frac{(1-\theta^2)^{-\frac{n-3}{2}}}{n-3} + I(3)/(n-3)$$

$$+ \sum_{j=0}^{\infty} \frac{2\Gamma(\frac{n+2j-1}{2})\theta^{2j+2}}{\Gamma(j+1)\Gamma(\frac{n-1}{2})} I(2j+5)$$

$$\times \left[\frac{2j+3}{4(j+1)} - \frac{I(2j+3)}{I(2j+5)}\right] . \tag{4.8}$$

Define the relative efficiency of $\bar{y}_s$ with respect to $\bar{y}_\ell$ as $e_1(\delta, m_0) = V(\bar{y}_\ell) / V(\bar{y}_s)$. Consider first the effect of variation of θ.

Theorem 4.7: For m_0 fixed such that $0 < m_0 \leq 1$, there exists a θ_0 such that $0 < \theta_0 < 1$ and $D_1(\theta, m_0) < 0$ and hence $e_1(\delta, m_0) > 1$, $-\theta_0 < \theta < \theta_0$

≥ 0 and hence $e_1(\delta, m_0) \leq 1$ otherwise.

The theorem can be proved by using techniques similar to those used in proving Theorem 4.3.

Next consider the effect of m_0 with θ fixed. The result is given without proof in Theorem 4.8.

Theorem 4.8: With θ fixed, $D_1(\theta, m_0)$ varies with $D_2(\theta, m_0)$ as a function of m_0. For θ fixed such that $0 \leq \theta < 1$, $D_1(\theta, m_0)$ falls in one of the following three categories:

(a) $D_1(\theta, m_0)$ is always increasing as a function of m_0 for $0 < m_0 \leq 1$;

(b) $\exists\, m_0^*$ such that $0 < m_0^* < 1$ and $D_1(\theta, m_0)$ is increasing as a function of m_0 for $m_0 < m_0^*$ and decreasing for $m_0 > m_0^*$;

(c) $D_1(\theta, m_0)$ is always decreasing as a function of m_0 for $0 < m_0 \leq 1$.

5. Conclusions and Recommendations Regarding the Use of the Sometimes Regression Estimator.

If conditions are such that the use of regression type estimators is warranted, the question arises as to when the sometimes regression estimator would be most appropriate. Actually, the sometimes regression estimator includes both the difference estimator $\bar{y}_d$ and the regression estimator $\bar{y}_\ell$ as special cases. Hence the sometimes regression estimator may be used whenever it is appropriate to use regression type estimators.

Consider the effect of change in the relative closeness of β_0 to β_2. Theorem 4.3 gives the result that for fixed m_0, $V(\bar{y}_s)$ is greater than $V(\bar{y}_d)$ for β_0 close to β_2, but this relationship reverses itself as the distance of β_2 from β_0 increases and it remains reversed. Theorem 4.7 illustrates that the situation is reversed for the relationship of the variance of the sometimes regression estimator to the variance of the regression estimator. Analogous results hold for the relative efficiencies. These results are illustrated in Figures 1 and 2 for n equal to 6. The relative distance between β_2 and β_0 is a fixed unknown quantity. However on the basis of past experience, it may be possible to have some idea about the likely range of values it can take on.

Now m_0 can be fixed in any manner we please. If m_0 is fixed such that the probability of using $\bar{y}_d$ is very high, then the relative efficiency of $\bar{y}_s$ with respect to $\bar{y}_d$ is close to 1. On the other hand, if m_0 is such that the probability of using $\bar{y}_\ell$ is high, then the relative efficiency of $\bar{y}_s$ with respect to $\bar{y}_\ell$ is close to 1. The effect of changing the level of significance of the test when the relative distance between β_0 and β_2 is fixed is illustrated in Figures 3 and 4 for n equal to 6.

If there is a priori information that β_0 may be the actual value of β_2, the guidelines for using the sometimes regression estimator may be stated as follows:

1) If β_0 is considered a very reliable guess for β_2 then t_0 may be chosen so that the likelihood that $\bar{y}_s$ results in using $\bar{y}_d$ is high. This would tend to minimize the loss in efficiency of $\bar{y}_s$ with respect to $\bar{y}_d$.

2) If β_0 is not considered a very reliable choice for β_2 then t_0 may be chosen so that the likelihood that $\bar{y}_s$ results in using $\bar{y}_\ell$ is very high. This would tend to minimize the loss in efficiency of $\bar{y}_s$ with respect to $\bar{y}_\ell$.

3) If no further information is available about the reliability of the choice of β_0, a middle range value for m_0 may be used.

6. Acknowledgement. This work was supported under Contract No. OEC-0-73-6640 by the U.S. Office of Education, Department of Health, Education and Welfare.

REFERENCES

Cochran, W.G. 1963. Sampling Techniques, Second Edition, Wiley, New York.

Cramer, Harold, 1946. Mathematical Methods of Statistics. Princeton, N.J., Princeton University Press.

Hansen, M.H., Hurwitz, W.N. and Madow, W.G. 1953. Sample survey methods and theory Vols. I and II: Methods and applications. Wiley, N.Y.

Sukhatme, P.V. and Sukhatme, B.V. 1970. Sampling theory of surveys with applications. Ames, Iowa, Iowa State University Press.

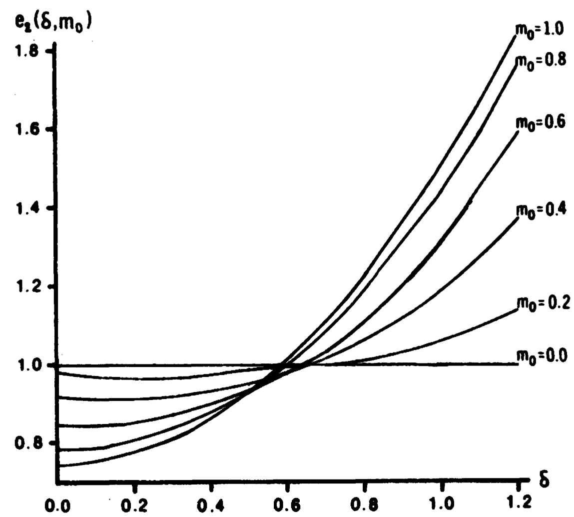

FIGURE 1

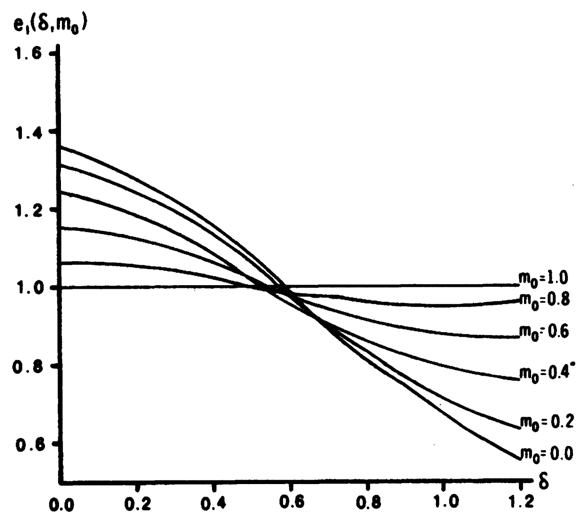

FIGURE 2

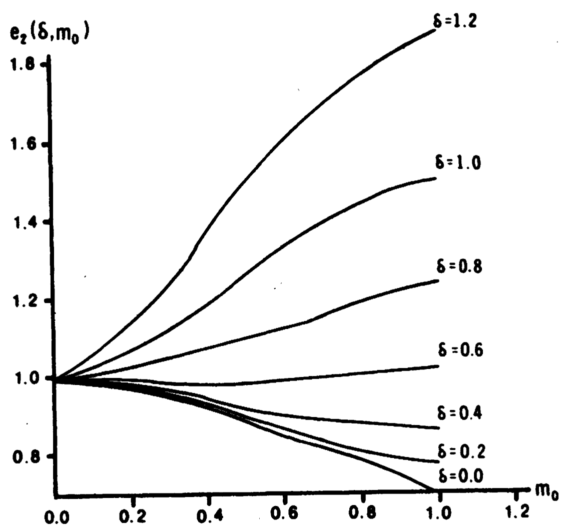

FIGURE 3

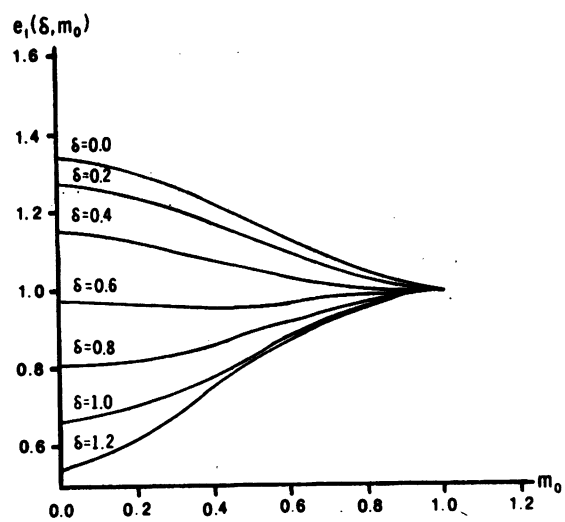

FIGURE 4