A COMPARISON OF TWO METHODS OF ESTIMATING CALIBRATION LINES

BY
KATHLEEN LAMBORN

TECHNICAL REPORT NO. 22
FEBRUARY 27, 1970

PREPARED UNDER THE AUSPICES OF
PUBLIC HEALTH SERVICE GRANT USPHS-5TI GM 25-12

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
A COMPARISON OF TWO METHODS OF ESTIMATING
CALIBRATION LINES

BY

KATHLEEN LAMBORN

TECHNICAL REPORT NO. 22
FEBRUARY 27, 1970

PREPARED UNDER THE AUSPICIES

OF

PUBLIC HEALTH SERVICE GRANT USPHS-5TI GM 25-12

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
I. Introduction

Suppose it is necessary to estimate $W$ given $Z$ when there is known to be a linear relation between the two of the form

$$Z = \alpha + \beta W + e$$

where $\alpha$ and $\beta$ are unknown parameters and $e$ is a random variable with variance $\tau^2$. Then a calibration line, an estimate of the relation $\alpha + \beta W$, is used. To form this calibration line $n$ observations $y_i$ are taken where

$$y_i = \alpha + \beta x_i + \epsilon_i$$

and the $x$'s are chosen to be specific values. The variance of $\epsilon_i$ is $\sigma^2$ and is not necessarily equal to $\tau^2$. The question arises as to how, given the information from the pairs $(x_i, y_i)$, to get the best calibration line in some meaningful sense.

A few examples will indicate the variety of situations in which such a problem might arise. Suppose it is necessary to calibrate an instrument, a thermometer for example. Then readings $y_i$ are taken at certain known temperatures $x_i$. In the future the calibration line provides an estimate of the true temperature given the reading. Assuming that these future readings are taken under the same conditions, the same
type of reading error will be present for each new observation and \( \epsilon^2 = \sigma^2 \). On the other hand, \( K \) readings could be taken at the unknown temperature and their average used as \( Z \). In this case \( \tau^2 = \sigma^2/K \). Finally, there may be instances when \( Z \) is not actually a random variable and \( \tau^2 = 0 \). This would be the case if the calibration line was a dosage response curve and it was desired to find the drug level to administer in order to achieve a certain response. When the dose is given, there would be a variation in the response, but the concern might be to achieve a mean response \( Z \).

R. G. Krutchkoff ([2] and [3]) considered this question of the "best" calibration line in terms of minimizing the squared error for future observations. In other words, if \( Z \) is observed and the calibration line gives an estimate \( \hat{W} \) of \( W \), the true value, he was concerned with minimizing \( (\hat{W} - W)^2 \). Under the assumption that \( \epsilon \sim N(0, \sigma^2) \), the classical approach has been to find the maximum likelihood estimates \( \hat{\alpha}, \hat{\beta} \) for \( \alpha \) and \( \beta \), and then for a new observation \( Z \) estimate \( W \) by \( \hat{W} = \frac{Z - \hat{\alpha}}{\hat{\beta}} \). Krutchkoff used Monte Carlo techniques for the case where \( \epsilon \sim N(0, \sigma^2) \) to show that, in fact, a better method might be an inverse approach. Using the inverse method, one proceeds as if the underlying model was \( x = \gamma + \delta y + \epsilon \) and finds formal least squares estimates for \( \gamma \) and \( \delta \). Then \( \hat{W}_B = \hat{\gamma} + \hat{\delta} Z \). He compared Monte Carlo averages, denoted "Av.", of \( (\hat{W} - W)^2 \) and \( (\hat{W}_B - W)^2 \) for each of several values of \( W, \alpha, \beta \) and \( \sigma \). The inverse method seemed to produce consistently lower values.

E. J. Williams ([4]) questioned the whole idea of considering \( (\hat{W} - W)^2 \) since \( E(1/\hat{\beta}) \) and hence \( E(\hat{W} - W)^2 \) are infinite when using the
classical approach. Thus, he reasons, any estimate with a finite mean and variance would do better under this criterion. In doing his study, Krutchkoff bounded $\hat{\beta}$ away from zero when low values of $\hat{\beta}$ seemed to be inflating $Av(\hat{W}-W)^2$ to an extreme degree. This removes much of the force of William's argument. Suppose a rule is established to bound $\hat{\beta}$ away from zero at all times. Then $E(\hat{W}-W)^2$ is no longer infinite. This truncation is not unreasonable since anyone wishing to use a calibration line assumes that $\hat{\beta}$ and $\beta$ are sufficiently far from zero to make the line useful. In considering the mean squared error of the truncated estimate, one is ignoring those extreme cases which are unlikely to occur and comparing the two methods on the basis of the more probable range of events.

This paper will show that if $\hat{\beta}$ is bounded away from zero and certain assumptions are made on the parameters, series expansions can be used to provide theoretical explanations for the results Krutchkoff observed and the relative quality of the two estimates can be stated as a function of the parameters and the sample size. In general the paper will present the results and conclusions with derivations given in the appendix.

II. Models

The estimation of the line is based on $n$ observations of the form $(x_1, y_1)$ where $y_1 = \alpha + \beta x_1 + \epsilon_1$, $(x_1, y_1)$ are known and $\epsilon_1 \sim N(0, \sigma^2)$. It is assumed that $|\beta| > 0$. The calibration line will then be used to estimate $W$ of the pair $(W, Z)$ where $Z$ is known, $Z = \alpha + \beta W + e$, $e \sim N(0, \tau^2)$. Define
\[ S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 \]
\[ S_{xe} = \sum_{i=1}^{n} (x_i - \bar{x})(\epsilon_i - \bar{\epsilon}) \]
\[ S_{ee} = \sum_{i=1}^{n} (\epsilon_i - \bar{\epsilon})^2 \]

The results can be stated most clearly in terms of \( 1/n \) \( S_{xx} \) which will be denoted by \( s_x^2 \), where \( s_x^2 \) is a measure of the dispersion of the chosen \( x \) values about their mean. It is assumed that \( s_x^2 \) remains essentially constant as \( n \) increases. This will be true in general since under most designs additional sampling is done at certain \( x \) values rather than new values being used.

**Method A (Classical Approach)**

The classical approach discussed earlier gives an estimate \( W = \frac{\hat{y}}{\hat{\beta}} \)
where \( \hat{\gamma} = \bar{y} - \hat{\beta} \bar{x} \) and \( \hat{\beta} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{S_{xx}} = \beta(1 + \frac{S_{xe}}{\beta S_{xx}}) \). It is desired to define a \( \hat{\beta}' \) such that \( P(\hat{\beta}' \neq \hat{\beta}) \) is small, yet \( E(1/\hat{\beta}') \) is finite. Let \( \frac{\sigma^2}{\hat{\beta}^2} = \text{Var} \hat{\beta} = \frac{S^2_{xe}}{S_{xx}} \). Assume \( \sigma^2/\beta^2 \ll 1 \), which is not unreasonable if this is to be a calibration line. Then an \( \xi \) can be chosen so that

\[
\hat{\beta}' = \begin{cases} 
\hat{\beta} & \beta - \xi < \hat{\beta} < \beta + \xi \\
\beta + \xi \text{ sgn} (\hat{\beta} - \beta) & \text{otherwise}
\end{cases}
\]

has an absolute value greater than zero for all values of \( \hat{\beta} \) and \( P(\hat{\beta} \neq \hat{\beta}) \) is small, thus satisfying both requirements. Define
\[ \hat{\alpha}' = \bar{y} - \hat{\beta}' \bar{x} \] and \( S_{x|x}^{'} \) analogously. Let

\[ \hat{W}_A = \frac{Z - \hat{\alpha}'}{\hat{\beta}'} \]

Then for \( \hat{W}_A \) not only are the needed expectations finite, but, as shown in the appendix, series expansions in \( S_{x|x}^{'}/\beta S_{xx} \) yield the approximations

\[ E(\hat{W}_A) = W + (W - \bar{x}) \left( \frac{\sigma^2}{\beta^2 s_x^2} + \frac{3\sigma^4}{\beta^4 s_x^4} \right) \]

\[ E(\hat{W}_A - W)^2 = (W - \bar{x})^2 \left( \frac{\sigma^2}{\beta^2 s_x^2} + \frac{9\sigma^4}{\beta^4 s_x^4} \right) + \frac{n\sigma^2 + \sigma'^2}{n} \left( 1 + \frac{3\sigma^2}{\beta^2 s_x^2} + \frac{15\sigma^4}{2\beta^4 s_x^4} \right) \]

**Method B (Inverse Approach)**

Following the inverse approach, one first solves the equation

\[ Y = \alpha + \beta x + \epsilon \] for \( x \). This gives

\[ x = -\alpha/\beta + 1/\beta \; Y + \epsilon' = \gamma + \delta Y + \epsilon' \]

Then using formal least squares estimates, \( \hat{W}_B = \hat{\gamma} + \hat{\delta} Z \) where

\[ \hat{\gamma} = \bar{x} - \hat{\delta} \bar{y} \] and \( \hat{\delta} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma (y_i - \bar{y})^2} \). Using the fact that \( \Sigma_{i=1}^n (y_i - \bar{y})^2 \)

has a noncentral \( \chi^2 \) distribution with \( n-1 \) degrees of freedom, it can be shown that for \( n \geq 4 \) \( \hat{\delta} \) has a finite expectation and for \( n \geq 6 \) the second moment exists. Then, as shown in the appendix, if \( \sigma^2/\beta^2 s_x^2 \ll 1 \) and \( n \geq 6 \) series expansions yield the approximations
\[ E(\hat{w}_B) = w + (w-x) \left[ -\frac{(n-3)\sigma^2}{n\beta^2 s_x^2} + \frac{\gamma^4}{n^2\beta^4 s_x^4} (n^2 - 8n + 15) \right] \]  

\[ E(\hat{w}_B-W)^2 = (w-x)^2 \left[ \frac{\sigma^2}{n\beta^2 s_x^2} + \frac{(n-2)\sigma^4}{n\beta^4 s_x^4} + \frac{n^2 + \gamma^2}{n^2\beta^2 s_x^2} \left[ 1 + \frac{(7-2n)\sigma^4}{n^2\beta^4 s_x^4} + \frac{3n^2 - 26n + 55}{n^2\beta^4 s_x^4} \right] \right] \]  

(3)

III. Comparisons

Assuming that the proper conditions on \( \sigma/\beta s_x \) hold, equations (1) and (3) can now be used to compare the quality of the two methods. One important point about these equations: these are not expansions in powers of \( n \). The higher order terms which have been ignored are higher orders of \( \sigma/\beta s_x \). Thus the expansions can be used to compare the two techniques for small values of \( n \). Notice that in evaluating these equations the actual values of \( \gamma, \sigma, \) and \( \beta \) need not be known. These parameters affect the mean squared error only through the ratios \( \gamma/\sigma \) and \( \sigma/\beta \).

Small Sample Case

Most of Krutchkoff's initial work was done with small \( n \), specifically \( n = 6 \). Here it is clear why \( Av(\hat{w}-w)^2 \) might be expected to be smaller for method B than for method A. The key is in the \( \gamma^4 \) term. The coefficient of \( (w-x)^2 \) is smaller under B than under A, and even as it becomes slightly larger the negative term in the second part of equation (3) tends to counteract it if \( (w-x)^2 \) is small and \( \gamma^2 \) is
large. In fact the choice between the two methods depends very much on these two factors as \( n \) increases. However, for \( n \leq 8 \) method \( B \) would be preferable independent of any other factor.

The other question to be considered is the effect of the design on the quality of the two methods. It is clear that the design affects \( \text{Av}(\hat{\omega} - \omega)^2 \) through \( \bar{x} \) and \( s_x^2 \). Obviously, using either procedure it is best to have \( \bar{x} \) as close as possible to where the \( \omega \) values are expected to fall. The difference in sensitivity to design which Krutchkoff feels he has observed is due to the fact that the coefficient of \( \frac{1}{n} \frac{2}{\beta} \frac{1}{s_x} \) is larger under method \( A \) than under method \( B \) for \( n = 6 \). In general, however, the two methods are affected differently by design. Under method \( A \) the mean squared error is a decreasing function of \( s_x^2 \), so it would always be best to make \( s_x^2 \) as large as possible. The situation under the inverse procedure is rather interesting. While the coefficient of \( (\omega - \bar{x})^2 \) is a decreasing function of \( s_x^2 \), the second term is an increasing function, so that there will be an optimal value for \( s_x^2 \) which will depend on \( |\omega - \bar{x}| \).

Finally, equations (1) and (3) can be used to give the approximation

\[
\text{E}(\hat{\omega}_A - \omega)^2 - \text{E}(\hat{\omega}_B - \omega)^2 = \frac{\sigma^4}{2 \beta \frac{n^2}{\gamma} s_x^2} \left[ \frac{2}{s_x^2} (\frac{(\omega - \bar{x})^2}{s_x^2}) (-n^2 + 10n - 16) + (\frac{\tau^2}{\sigma^2} + 1)(2n - 10) \right]
\]

(4)

which can be used to make precise statements as to when one method would be preferable to the other.

**Large Sample Case**

For \( n \) large two cases must be considered: \( \tau = 0 \) and \( \tau > 0 \).
First, for \( \gamma = 0 \) the mean squared error under method A goes to zero while \( E(\hat{W}_B - W)^2 \geq (W - \bar{x})^2 \sigma_x/\beta s_x^2 \) which is small but does not vanish. Thus, for \( n \) large method A would be preferable and method B could always be improved by increasing \( s_x^2 \). The difference between the two methods is due to the fact that the classical estimates give consistent estimates for the regression coefficients, so that in the limit the line is estimated exactly, but the inverse estimates do not.

The situation is different when \( \gamma^2 > 0 \), since the estimate of the calibration line is no longer the only source of error. The second term in the two equations no longer goes to zero as \( n \to \infty \). For large \( n \)

\[
E(\hat{W}_A - W)^2 \approx \frac{\gamma^2}{\beta^2}
\]

\[
E(\hat{W}_B - W)^2 \approx (W - \bar{x})^2 \frac{\sigma_x^4}{\beta^4 s_x^2} + \frac{\gamma^2}{\beta^2} \left[ 1 - \frac{2\sigma_x^2}{\beta^2 s_x^2} \right]
\]

Then for \( W \) close to \( \bar{x} \), method B could be expected to produce a smaller squared error even when a large number of points are used to estimate the line.

The reason for this becomes clearer by considering that

\[
\hat{W}_B = W + \beta(W - \bar{x})(\hat{\delta} - \delta) + \delta(e - \bar{e})
\]

\[
\hat{W}_A = W + \beta(W - \bar{x})(1/\hat{\beta} - 1/\beta) + (e - \bar{e}) 1/\hat{\beta}^2
\]

If \( W \) is close to \( \bar{x} \) the last term is the significant one in both equations. As is seen from equations (1A) and (4A) of the appendix, \( \hat{\delta} \) is biased in such a way that it tends to be smaller in absolute value
than $\delta = 1/\beta$ while $1/\beta' \to 1/\beta$ as $n$ becomes large. Thus $\hat{W}$ tends to be closer to $W$ using the inverse procedure.

To determine explicit conditions under which method $B$ would be preferable to method $A$ for large $n$, take the limit over $n$ in equation (4) to get

$$E(\hat{W}_A - W)^2 - E(\hat{W}_B - W)^2 = \frac{\sigma^4}{\beta^4 s^2_x} \frac{2\tau^2}{\sigma^2} - \frac{(W-x)^2}{s^2_x}.$$  

Thus, method $B$ would be preferable to the classical approach if

$$\frac{\tau^2}{\sigma^2} > \frac{(W-x)^2}{2s^2_x}.$$  

As $\frac{\tau^2}{\sigma^2} \to 0$, $W$ must be much closer to $\bar{x}$ relative to the dispersion of the $x$'s. In arguing that Krutchkoff's assumptions are not likely to hold in real situations, Berkson points out that the variability for the future observations is likely to be greater than that when the original calibration points were taken, i.e., $\frac{\tau^2}{\sigma^2} > 1$. As the above inequality shows, this would simply mean an increase in the range of $W$'s for which method $B$ would be preferable over the case when $\frac{\tau^2}{\sigma^2} = 1$.

For $\tau/\sigma$ fixed the inequality implies that $B$ is preferable if $s^2_x > \frac{(W-x)^2}{2} \frac{\sigma^2}{\tau^2}$, i.e., a larger value of $s^2_x$ implies that $B$ is preferable to $A$ for a greater range of $W$'s. At the same time, it can be shown that the mean squared error under $B$ is minimized for $s^2_x = \frac{(W-x)^2}{2} \frac{\sigma^2}{\tau^2}$ so that increasing $s^2_x$ outside the range in which the new observations are expected would not only be unprofitable, but would in fact be detrimental. For the case $\tau^2 = \sigma^2$ the above inequality becomes $s^2_x = \frac{(W-x)^2}{2}$. Thus, if the $x$'s are taken at two extreme points, method $B$ would be preferable to $A$ for a range $\sqrt{2}$ times the size of the initial distance.
Robustness

Another criterion for choosing between the two methods is their sensitivity to the assumptions. In the discussions above, it was assumed that \( \varepsilon_i \sim N(0, \sigma^2) \). Suppose that \( \varepsilon_i \) has a mean zero and variance \( \sigma^2 \) but is not necessarily normally distributed. Let

\[
\theta_1 = \frac{E\varepsilon_i^3}{\sigma^2}, \quad \theta_2 = \frac{E\varepsilon_i^4}{\sigma^4} - 3 .
\]

If \( \varepsilon_i \) is normal then \( \theta_1 = \theta_2 = 0 \). Under the assumption the \( \varepsilon_i \) has a distribution close to that of a normal, perhaps normal plus contamination, equations (14A)-(17A) give the approximations for the mean and mean squared errors under the two methods. These equations show that even for small \( n \), the difference in the expected squared error due to non-normality is not large if the distribution is similar to the normal distribution. If the distribution differs significantly from the normal by having much larger tails, then \( P(\hat{\beta}' \neq \hat{\beta}) \) becomes large and the error in the equations for method A may be of the same order as the values of the equations themselves. However, if the tails are too large neither method is likely to produce a satisfactory estimate for small \( n \).

For \( n \) large and under the assumption that \( E|\hat{\delta}|^{2+\ell} \) is bounded over \( n \) for some \( \ell > 0 \), it can be shown that the squared errors are the same as in the normal case and the choice of methods would follow as in that case.

Effect of a Quadratic Term

Until now it has been assumed that the linear model was correct.
Suppose a linear model is assumed, but in fact

\[ y_i = \alpha + \beta x_i + \gamma (x_i - \bar{x})^2 + \epsilon_i \quad \epsilon_i \sim N(0, \sigma^2) \]

How does this affect the relative value of the two models? Suppose the \( x_i \) are chosen to be symmetric about \( \bar{x} \). Then \[ \sum_{i=1}^{n} (x_i - \bar{x})^k = 0 \] for \( k \) odd. With this symmetry \( \hat{\beta} \) under method A is unchanged by the addition of the quadratic term. For \( n \) large so that terms of order \( 1/n \) can be ignored its presence increases \( E(\hat{\beta}_A - W)^2 \) by the term \( \gamma^2/\beta^2 [E(\bar{x})^2 - s_x^2]^2 \). When using method B, several factors seem to determine the effect of the quadratic term on \( E(\hat{\beta}_B - W)^2 \), and, thus, the relative quality of the two methods. The absolute value of \( E(\hat{\beta}) \) is smaller when the quadratic term is present (see (20A) of the appendix). If \( \tau^2 > 0 \) and \( (W - \bar{x})^2 \) and \( [(W - \bar{x})^2 - s_x^2]^2 \) are small, the same reasoning presented for the true linear case implies that for \( n \) large \( E(\hat{\beta}_B - W)^2 \) will be smaller than \( E(\hat{\beta}_A - W)^2 \). Other than that it does not seem to be possible to define general rules for when one will have a smaller expected squared error than the other. For specific values of \( W, \bar{x}, s_x^2 \) and the parameters, equation (21A) in the appendix gives the approximate expected squared error for method B when \( n \) is large.

IV. Quality of the Approximations

Whenever approximations are used as the basis of a discussion, as they were throughout this paper, it is important to consider how the quality of the approximations varies with changes in the parameters. As would be hoped, the size of the approximation error relative to the value of the equation is a decreasing function of \( n \) and an increasing
function of $\sigma^2 / \beta^2 s_x^2$ in both equations. The error due to approximating the moments of the truncated random variable in (1) becomes insignificant compared to the error from ignoring higher order terms as $n$ increases, and both errors go to zero. The approximation error under method B is due to ignoring higher order terms and goes in the limit to

$$0 \left( \frac{[\sigma^2 / \beta^2 s_x^2]^3}{1 + \sigma^2 / \beta^2 s_x^2} \right).$$

An explicit upper bound can be put on this remainder term, a bound which is decreasing in $n$ and increasing in $\sigma^2 / \beta^2 s_x^2$. This can be used to show that the remainder is small comparatively even for small $n$. A more exact statement of the rates of convergence in both the normal and nonnormal cases is given in the appendix.

To get an indication of how close the approximations were for small $n$, a comparison was made between the values of the equations and the average squared errors Krutchkoff observed for a variety of the parameters he chose. They were quite close even for $n = 6$ and $\sigma / \beta = .2$. Since $\sigma / \beta = .2$ gave a standard error of .2 under method A even in the large sample situation when $s_x = 1/2$, this seems a large enough value to consider. The approximating equations in the nonnormal case were also compared to Krutchkoff's averages using Pearson distributions with a variety of values of skewness and kurtosis, and they were quite close (again for $n = 6$).

V. Conclusions

Under what seem to be reasonable assumptions, there are instances when the squared error would be expected to be lower using method B than using method A. For small $n$ the difference in the magnitude of
the bias is not large and the difference in the variance may be more significant. For $n \leq 8$ this will be true for any values of the parameters which satisfy the assumptions. As $n$ increases the choice between the two methods will depend on the distance of $W$ from $\bar{X}$ relative to $s^2_x$, and on the size of $\tau^2/\sigma^2$. If $Z$ is observed without error, so that the only error is due to estimation of the calibration line, the classical method will be preferable in the asymptotic situation. If $Z$ is observed with error, then there is a range of values of $W$ around $\bar{X}$ where method B would be preferable in terms of mean squared error even asymptotically. As the error associated with $Z$ increases relative to $\sigma^2$, this range increases. Both methods seem to be robust to departures from normality. The relative merits of the two methods of the model should have been quadratic depending on the location of $W$ in relation to $\bar{X}$, the relative size of $(W-\bar{X})^2$ and $s^2_x$, and the values of the parameters.

VI. Acknowledgment
The author wishes to thank Professors Rupert Miller, Jr. and Lincoln Moses for their many helpful suggestions during the preparation of this paper.
APPENDIX

Suppose \( y_i = \alpha + \beta x_i + \epsilon_i \), \( Z = \alpha + \beta W + e \) where \( \epsilon_i \sim N(0, \sigma^2) \), \( e \sim N(0, \sigma^2) \) and \( |\beta| > 0 \).

Method A

\[
\hat{W}_A = \frac{Z - \hat{\alpha}}{\hat{\beta}} = \alpha + \beta W + e - (\alpha + \beta \bar{x} + \hat{\beta} \bar{x}) = \frac{\hat{\beta}(W - \bar{x}) + (e - \bar{e})}{\hat{\beta}} + \bar{x}
\]

\[
\hat{\beta} = \frac{\Sigma (x_i - \bar{x})(\alpha + \beta x_i + \epsilon_i - \bar{\epsilon}^*)}{S_{xx}} = \frac{\beta S_{xx} + S_{x\epsilon}}{S_{xx}}
\]

\[
1/\hat{\beta} = 1/\beta \left( 1 + \frac{S_{x\epsilon}}{\beta S_{xx}} \right)^{-1}
\]

Let \( S'_{x\epsilon} \) be the random variable defined by \( \hat{\beta}' = \frac{\beta S_{xx} + S'_{x\epsilon}}{S_{xx}} \). Then

\[
|S'_{x\epsilon}/\beta S_{xx}| \leq \delta/|\beta| = \epsilon < 1 \Rightarrow 1/\hat{\beta}' \text{ can be put into a series}
\]

expansion using the relation

\[
(1 + u)^{-1} = \sum_{i=0}^{\infty} (-u)^i + \frac{(-u)^5}{1+u}
\]

where \( u = \frac{S'_{x\epsilon}}{\beta S_{xx}} \). Similarly

\[
(1/\hat{\beta}')^2 = \sum_{i=0}^{\infty} (1+1)(-u)^i + R_5
\]

This gives the following approximations

\[
\frac{1}{\hat{\beta}'} = \frac{1}{\beta} \left( 1 - \frac{S'_{x\epsilon}}{\beta S_{xx}} + \left[ \frac{S'_{x\epsilon}}{\beta S_{xx}} \right]^2 - \left[ \frac{S'_{x\epsilon}}{\beta S_{xx}} \right]^3 + \left[ \frac{S'_{x\epsilon}}{\beta S_{xx}} \right]^4 \right) \quad (1A)
\]
\[
\left(\frac{1}{\beta'}\right)^2 = \frac{1}{\beta'} \left(1 - \frac{2s_x}{s_{xx}} + 3\left(\frac{s_x}{s_{xx}}\right)^2 - 4\left(\frac{s_x}{s_{xx}}\right)^3 + 5\left(\frac{s_x}{s_{xx}}\right)^4\right) \quad (2A)
\]

Using the Taylor expansion for \( u > 0 \) gives \( |R_5| < u^5 \). For \( u < 0 \)

\[
|R_5| = \sum_{i=5}^{\infty} (i+1)(i)\left|\frac{u}{u+1}\right|^i < \frac{6|u|^5}{(1+u)^2}
\]

so that \( \frac{6|\xi/\beta|^5}{(1-|\xi/\beta|)^2} \) gives an upper bound to the magnitude of the approximations in (1A) and (2A).

Under the assumption that \( e^2/\beta^2 s_x^2 \) is small, \( \xi \) can be chosen such that \( E\left[\frac{u^5}{(1+u)^2}\right] \) and \( ER_5 \) are small and such that \( E\left(\frac{s_x}{s_{xx}}\right)^k \) can be approximated by \( E\left(\frac{s_x}{s_{xx}}\right)^k = 0 \quad 0 < k \leq 4 \). For \( k \) odd

\[
E\left(\frac{s_x}{s_{xx}}\right)^k = E\left(\frac{s_x}{s_{xx}}\right)^k = 0 \quad \text{by symmetry}.
\]

\[
E(s_x^2) = s_{xx}^2 \\
E(s_x^4) = 3s_{xx}^2
\]

The following are good approximations:

\[
E\left(\frac{1}{\beta'}\right)^2 = \frac{1}{\beta'} \left[1 + \frac{\sigma^2}{n\beta s_x^2} + \frac{3\sigma^4}{n^2\beta^2 s_x^4}\right] \quad (3A)
\]

\[
E(\hat{w}_A) = w + (w-x)\left(-\frac{\sigma^2}{n\beta s_x^2} + \frac{3\sigma^4}{n^2\beta^2 s_x^4}\right) \quad (4A)
\]

\[
E(\hat{w}_A - w)^2 = E[\left(w-x\right)^2 - 2\left(w-x\right)(\beta(w-x) + (e-\bar{e}))) \frac{1}{\beta'} \]

\[
+ [\beta^2(w-x)^2 + 2\beta(e-\bar{e})(w-x) + (e-\bar{e})^2] \left(\frac{1}{\beta'}\right)^2 \}
\]

\[
\pm (w-x)^2\left(-\frac{\sigma^2}{n\beta s_x^2} + \frac{9\sigma^4}{n^2\beta^2 s_x^4}\right) + \frac{n^2\sigma^2}{n\beta s_x^2} \left[1 + \frac{3\sigma^2}{n\beta s_x^2} + \frac{15\sigma^4}{n^2\beta^2 s_x^4}\right] \quad (5A)
\]
These series expansions are written for fixed \( n \). Before limits can be taken over \( n \), one must examine the effect of varying \( n \) on the quality of the approximations.

First, notice that the truncation of \( \hat{\beta} \) to \( \hat{\beta}' \) becomes less important as \( n \) increases. Let \( f_n(u) \) be the density function of
\[
\frac{1/n \cdot x_6}{\beta s_x^2}.
\]
Then
\[
P(\hat{\beta} \neq \hat{\beta}') = \int_{\frac{|\xi/\beta|}{|\xi/\beta|}}^{\infty} f_n(u) \, du.
\]

From the normality,
\[
\frac{f_{n+1}(u)}{f_n(u)} = \frac{\sqrt{n+1}}{\sqrt{n}} \cdot \frac{1}{\sigma^2} \cdot e^{-\frac{2\sigma^2}{x_6^2} u^2}
\]
under the assumption that \( s_x^2 \) remains constant over \( n \). For \( u \geq |\xi/\beta| = c \)
\[
\frac{f_{n+1}(u)}{f_n(u)} \leq \frac{\sqrt{n+1}}{\sqrt{n}} \cdot e^{\frac{c^2 \cdot s_x^2 \cdot \beta^2}{2\sigma^2}}
\]
and if \( \hat{\beta} \) is based on \( n \) observations
\[
P(\hat{\beta}_n \neq \hat{\beta}'_n) \leq P(\hat{\beta}_0 \neq \hat{\beta}'_0) \sqrt{n/6} \cdot e^{-\frac{(n-6) c^2 \cdot s_x^2 \cdot \beta^2}{2\sigma^2}}
\]
Next consider the approximations used in (5A). By definition
\[
|E(\frac{1/n \cdot x_6^k}{\beta s_x^2}) - E(\frac{1/n \cdot x_6^k}{\beta s_x^2})| \leq \frac{1}{|\xi/\beta|} \cdot 2 \cdot \int_{\frac{|\xi/\beta|}{|\xi/\beta|}}^{\infty} f_n(u) \, du \cdot |c^k - c^k|
\]
Denote this error as a function of \( n \) by \( r_n^k \). Then using the same reasoning as above

16
\[ 0 \leq r_n^k \leq r_6^k \sqrt{n/6} e^{-\frac{(n-6)c^2 s_x^2 \beta^2}{2 \sigma^2}} \]

The error due to ignoring the higher order terms in expanding \( 1/\beta \) equals \( u^5/1+u \). For \( |u| < c \) this is less in magnitude than \( u^5/1-c \).

\[ E^2(u^5) \leq E^2(|u|^5) \leq E(u^{10}) \]

Thus

\[ |E_{u^5/1+u}| \leq d \frac{\sigma^5}{|\beta| s_x^5 n^{5/2}} \]

where \( d \) is a constant independent of \( n \) and of \( \frac{\sigma^2}{\beta^2 s_x^2} \). A similar proof shows that the error from the approximation of \( E(1/\beta')^2 \) is also \( O(n^{-5/2}) \).

**Method B**

\[ \hat{w}_B = \hat{\gamma} \ast \hat{\delta} Z \]

\[ \hat{\gamma} = \bar{x} - \hat{\delta} y \]

\[ \hat{w}_B = \bar{x} - \hat{\delta}(\alpha + \beta \bar{x} + \bar{c}) + \hat{\delta}(\alpha + \beta \bar{w} + \bar{c}) \]

\[ = \bar{x} + \hat{\delta}[\beta(\bar{w} - \bar{x}) + (\bar{e} - \bar{c})] \]

\[ \hat{\delta} = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (y_i - \bar{y})^2} \]

\[ |\hat{\delta}|^2 < \frac{s_{xx}}{\Sigma (y_i - \bar{y})^2} \]  \( (6A) \)

17
Then \( E\delta \) and \( E\delta^2 \) are finite for all \( n \geq 6 \) using the fact that 
\[ \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - \bar{y})^2 \]
is distributed as a noncentral \( \chi^2 \) with \( n-1 \) degrees of freedom.

Now rewrite \( \delta \) to give

\[
\hat{\delta} = \frac{\sum (x_i - \bar{x})(\alpha + \beta x_i + e_i - (\alpha + \beta \bar{x} + e))}{\sum [\beta(x_i - \bar{x}) + (e_i - \bar{e})]^2} = \frac{\beta S_{xx} + S_{xxe}}{\beta^2 S_{xx} + 2\beta S_{xe} + S_{ee}}
\]

\[= \frac{1}{\beta} \left( 1 + \frac{S_{xxe}}{\beta S_{xx}} \right) \left( 1 + \frac{2\beta S_{xe} + S_{ee}}{\beta^2 S_{xx}} \right)^{-1}
\]

Let \( u = \frac{2\beta S_{xe} + S_{ee}}{\beta^2 S_{xx}} \). Since \( \beta^2 S_{xx} + 2\beta S_{xe} + S_{ee} \) is greater than zero with probability 1, \( u > -1 \) with probability 1 and

\[
\hat{\delta} = \frac{1}{\beta} \left( 1 + \frac{S_{xxe}}{\beta S_{xx}} \right) (\sum_{i=0}^{4} (-u)^i + \frac{u^5}{1+u}) \quad (7A)
\]

Substituting for \( u \) this gives the approximation

\[
\hat{\delta} \approx \frac{1}{\beta} \left( 1 + \frac{S_{xxe}}{\beta S_{xx}} \right) \left( 1 - \frac{2S_{xxe}}{\beta S_{xx}} - \frac{S_{ee}}{\beta S_{xx}} + \frac{4S_{xxe}^2}{\beta S_{xx}} + \frac{4S_{xxe}S_{ee}}{\beta S_{xx}} + \frac{S_{ee}^2}{\beta S_{xx}} \right)
\]

\[= \frac{8S_{xxe}^3}{\beta S_{xx}^2} - \frac{12S_{xxe}S_{ee}}{\beta S_{xx}^2} - \frac{6S_{xxe}S_{ee}^2}{\beta S_{xx}^2} - \frac{S_{ee}^3}{\beta S_{xx}^2} + \frac{16S_{xxe}}{\beta S_{xx}^2}
\]

\[+ \frac{32S_{xxe}^3 S_{ee}}{\beta S_{xx}^4} + \frac{4S_{xxe}S_{ee}^2}{\beta S_{xx}^2} + \frac{8S_{xxe}^3 S_{ee}}{\beta S_{xx}^4} + \frac{S_{ee}^4}{\beta S_{xx}^4} \quad (8A)
\]

In order to find the expected value of \( \hat{\delta} \) one needs the following relationships:
\[ E(S_{x}^{k}S_{e}^{l}) = 0 \quad \text{for} \quad k \text{ odd} \]

\[ E(S_{x}^{2}S_{e}^{2}) = (n+1)S_{xx}^{4} \]

\[ ES_{e}^{2} = (n^2 - 1)\sigma^4 \]

Then taking the expected value of (8A) yields

\[ \beta E(\delta) = 1 - \left( \frac{(n-3)\sigma^2}{n\beta s_x^2} + \frac{(n-5)(n-3)\sigma^4}{n^2 \beta s_x^4} \right) \left( \frac{\sigma^2}{\beta s_x^2} - \frac{\sigma^4}{\beta s_x^4} \right)(1 + O(1/n)) \]

Then

\[ E(\hat{\delta}^2) = \hat{W} + (W-\bar{x})\left[ \frac{-\left(\frac{(n-3)\sigma^2}{n\beta s_x^2}\right)}{n^2 \beta s_x^4} + \frac{(n-5)(n-3)\sigma^4}{n^2 \beta s_x^4} \right] \]

Following a similar procedure for (\delta)^2 gives

\[ \delta^2 = \frac{1}{\beta^2} \left(1 + \frac{S_{xx}}{\beta s_x^2}\right)^2 \sum_{i=0}^{4} (i+1)(-u)^i + B_5 \]

\[ \beta^2 E(\delta^2) = 1 - \left( \frac{-2(n-7)\sigma^2}{n\beta s_x^2} \right) + \left( \frac{3(n-11)(n-5)\sigma^4}{n^2 \beta s_x^4} \right) \]

\[ E(\hat{\delta}^2) = (W-\bar{x})^2 \left[ \frac{\sigma^2}{n\beta s_x^2} + \frac{(n-5)\sigma^4}{n^2 \beta s_x^4} \right] + \left( \frac{n+2+\sigma^2}{n\beta^2} \right)(\beta^2 E(\delta^2)) \]  \hspace{1cm} (10A) \]

As in the evaluation of method A, the above series expansions were done for fixed values of the parameters and of \( n \), and it is necessary to consider how the quality varies as these quantities change. A discussion will be given below for \( E\delta \). The case of \( E\delta \) is similar.

Rewriting \( \hat{\delta} \) and \( u \) to show their dependence on \( n \) gives
\[
    u_n = \frac{2\beta(1/n S_{x\xi}) + 1/n S_{\xi \xi}}{\beta^2 S_x^2}
\]

\[
    \hat{\delta}_n = \frac{1}{\beta} \left( 1 + \frac{1/n S_{x\xi}}{\beta S_x^2} \right) (1 + u_n)^{-1}
\]

From (7A)

\[
    E(\hat{\delta}_n^5) = \frac{1}{\beta} E(1 + \frac{1/n S_{x\xi}}{\beta S_x^2}) \sum_{i=0}^{4} (-u_n)^i - \frac{1}{\beta} E(1 + \frac{1/n S_{x\xi}}{\beta S_x^2})(\frac{u_n^5}{1+u_n})
\]

Let \( R_n = (1 + \frac{1/n S_{x\xi}}{\beta S_x^2})(\frac{u_n^5}{1+u_n}) \). Calculating the expected value of the right hand side of (8A) and taking the limit gives

\[
    \lim_{n \to \infty} E(\hat{\delta}_n^5) = \frac{1}{\beta} \sum_{i=0}^{4} (-\frac{\sigma^2}{\beta^2 S_x^2})^i + \lim_{n \to \infty} \frac{1}{\beta} E R_n
\]

Using (6A) \( E(\hat{\delta}_n^5) < \frac{ns_x^2}{\sigma^2(n-3)} \) so that \( E(\hat{\delta}_n^5) \) is bounded over \( n \).

From this and the law of large numbers it follows that

\[
    \left( \frac{\sigma^2}{\beta^2 S_x^2} \right)^5 \lim_{n \to \infty} E R_n = \frac{\sigma^2}{1 + \frac{\sigma^2}{\beta^2 S_x^2}}
\]

It remains to determine an upper bound for \( E R_n \) as a function of \( n \) and \( \sigma^2/\beta^2 S_x^2 \).

\[
    E|B_n| = E[\left| 1 + \frac{1/n S_{x\xi}}{\beta S_x^2} \right| \left| u_n^5 \right|]
\]

\[
    (1 + \frac{S_{x\xi}}{\beta S_x^2})^2 = 1 + \frac{2\beta S_{x\xi}}{\beta^2 S_x^2} + \frac{S_{x\xi}^2}{\beta^2 S_x^2}
\]
\[
= 1 + u_n + \frac{s_{\xi}^2 - s_{xx}s_{\xi}}{\beta^2 s_{xx}^2} \leq 1 + u_n
\]

\[
E|B_n| \leq E \left| \frac{u_n}{\sqrt{1 + u_n}} \right|^5
\]

Let \( v_n = \frac{1}{\sigma^2} \sum_1^n (y_i - \bar{y})^2 = \frac{n\beta^2}{\sigma^2} (1 + u_n) \). Consider 3 cases \( u \in (-1, -\xi), \)

\( u \in (-\xi, 0) \), and \( u \geq 0 \), where \( \xi > 0 \) will be defined below.

The contribution to \( E|B_n| \) from \( u \in [-1, -\xi] \) is less than

\[
\int_0^{(1-\xi) \frac{n\beta^2}{\sigma^2} \frac{s_{xx}^2}{\sigma^2} \frac{1/2}{\sqrt{2} \sigma_v} f_n(v_n) dv_n \quad (11A)
\]

where \( f_n(v_n) \) is the density function of the noncentral \( \chi^2 \) distribution with \( n-1 \) degrees of freedom. For \( v_n \leq n-1 \) \( f_n(v_n) \) is less than the density function for \( \chi^2_{n-1} \). Choose \( \xi \) so that \( (1-\xi)n(\beta^2 s_{xx}^2/\sigma^2) \leq n-1 \).

For simplicity let \( \xi = 1 - \frac{\sigma}{2\beta s_{xx}^4} \). Let \( W_n = \frac{v_n}{n} \). Then (11A) is less than

\[
\int_0^{\frac{1}{\beta s_{xx}^4 \left(\sqrt{2} \sigma / \sqrt{2} \sigma_v\right)}} \left(\frac{nW_n}{2} \right)^{n-4} e^{-\frac{nW_n}{2}} \frac{dW_n}{\Gamma\left(n-1 - \frac{1}{2}\right)}
\]

For fixed \( n \) the integrand attains its maximum at \( W_n = \frac{n-4}{n} \). Thus, for \( n \) such that \( \frac{n-4}{n} > \frac{1}{2} \frac{\sigma^2}{\beta^2 s_{xx}^2} \) the integrand is less than
\[
\frac{1}{\sqrt{2}} \left[ n^{-\frac{1}{2}} \left( \frac{\sigma^2}{\beta^2 s_x^2} \right)^{\frac{1}{2}} e^{-\frac{1}{2} \left( \frac{\sigma^2}{\beta^2 s_x^2} \right)} \right] \left[ \Gamma \left( \frac{n-1}{2} \right) \right]^{-1}
\]

\[
< \frac{1}{\sqrt{2}} \left[ n^{-\frac{1}{2}} \left( \frac{\sigma^2}{\beta^2 s_x^2} \right)^{\frac{1}{2}} e^{-\frac{n}{4} \frac{\sigma^2}{\beta^2 s_x^2}} \right] \left[ \Gamma \left( \frac{n-1}{2} \right) \right]^{-1} \text{ for } \frac{\sigma^2}{\beta^2 s_x^2} < 1.
\]

For \( u \in (-\xi, 0) \) \( \frac{u^5}{(1+u)^{1/2}} \) < \( \frac{u^5}{(1-\xi)^{1/2}} \) and \( |u| < \frac{2 \xi}{\beta s_{xx}} \). Therefore

\[
\int_{-\xi}^{0} u^5 dF_n(u) < \frac{\xi}{\beta s_{xx}} \frac{5}{10} \sigma_{10}^4 \left( \frac{\sigma^2}{\beta^2 s_x^2} \right)^{10}
\]

where \( d \) is a constant independent of \( n \) and \( \frac{\sigma^2}{\beta^2 s_x^2} \). Then

\[
\int_{-\xi}^{0} \frac{u^5}{(1+u)^{1/2}} dF_n(u) < \frac{2}{3} d \frac{\sigma^2}{\beta^2 s_x^2} n^{-\frac{5}{2}} \left( \frac{\sigma^2}{\beta^2 s_x^2} \right)^3
\]

For \( u > 0 \) \( \int_0^\infty B_n(u) dF_n(u) < \int_0^\infty u^{10} dF_n(u) \). Expanding \( u^{10} \) and taking the expected value provides an upper bound for this term, so that for any specific value of \( n \) and the parameters an upper bound for \( |EB_n| \) can be given.

**Robustness**

Suppose \( e_i \) has mean zero and variance \( \sigma^2 \). Let

\[
\theta_1 = \frac{E e_i^3}{\sigma^3} \quad \theta_2 = \frac{E e_i^4}{\sigma^4} - 3
\]

\[
s_k = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^k
\]

Assume \( s_k \) is constant over \( n \). Equations (4A) and (5A) now have
additional terms due to $\theta_1$ and $\theta_2$ (which were zero in the normal case) and to the fact that $\bar{e}$ is no longer independent of $(\varepsilon_i - \bar{e})$.

As in the normal case, the truncation of $\hat{\beta}$ becomes less significant as $n$ increases. Using Chebychev's inequality

$$P(\beta_n \neq \hat{\beta}_n) < \frac{\sigma^2}{\frac{n s_x^2}{\hat{\beta}^2}}.$$

For $k$ even $E\left(\frac{1}{\beta s_x^2} \right)^k < E\left(\frac{1}{\beta s_x^2} \right)^k$.

$$|E\left(\frac{1}{\beta s_x^2} \right)^{1/n s_x^2} | \leq E\left(\frac{1}{\beta s_x^2} \right)^{1/2} < \frac{\sigma}{\sqrt{n} |\beta| s_x^2}$$

For $k$ odd, greater than 1,

$$|E\left(\frac{1}{\beta s_x^2} \right)^{1/n s_x^2} | = O(n^{-k})$$

Similar inequalities hold for $E(\varepsilon_i s_x^\beta)$ depending on whether $(k + \ell)$ is even or odd. With no further assumptions on the distribution of $\varepsilon_i$, the error in the equations may now be of the same order as the terms themselves. However, it is possible to write

$$E(\hat{\theta}_A) = W + (W - \bar{X}) 0(\frac{1}{\sqrt{n}}) \quad (12A)$$
\[ E(\hat{W}_A - W)^2 = (W - \bar{x})^2 0(1/n) + 2(W - \bar{x})0(n^{-2}) + \frac{\sigma^2}{\beta^2} \left[ 1 + o\left( \frac{1}{\sqrt{n}} \right) \right] + \frac{\sigma^2}{np^2} \left[ 1 + o\left( \frac{1}{\sqrt{n}} \right) \right] \]

(13A)

Suppose the distribution of \( e_1 \) is such that the error incurred by replacing \( S'_{xx} \) by \( S_{xx} \) in computing expectations involving the first four moments of \( e \) are small compared to the expected values.

\[
E(\hat{e}S_{xx}') = 0
\]

\[
E(\hat{e}S^2_{xx}') = \theta \sigma^2 x
\]

\[
E(\hat{e}S^3_{xx}') = \theta s^4 \sigma^4
\]

\[
E(\hat{e}S^2_{xx}') = \theta s^4 \sigma^4 / n^2
\]

Then under these conditions

\[
E(\hat{W}_A) = W + (W - \bar{x})\left( \frac{\sigma^2}{n\beta s_x^2} + \frac{3\sigma^4}{2n\beta s_x^4} - \frac{1}{2n\beta s_x^6} + \frac{\theta_1 s^3}{n\beta s_x^8} + \frac{\theta_2 s^4}{n\beta s_x^8} \right)
\]

\[
- \frac{1}{2n\beta s_x^8} + \frac{\theta_2 s^4}{n\beta s_x^8}
\]

(14A)

\[
E(\hat{W}_A)^2 = (W - \bar{x})^2 \left( \frac{\sigma^2}{n\beta s_x^2} - \frac{2\sigma^4}{n\beta s_x^4} + \frac{9\sigma^4}{2n\beta s_x^4} + \frac{3\theta_1 s^3}{n\beta s_x^8} \right)
\]

\[
+ 2(W - \bar{x})\left[ - \frac{2\sigma^3}{n\beta s_x^2} + \frac{3\theta_2 s^4}{n\beta s_x^8} \right]
\]

\[
+ \frac{\sigma^2}{\beta^2} \left[ 1 + \frac{3\sigma^2}{n\beta s_x^2} - \frac{4\theta_1 s^3}{n\beta s_x^8} + \frac{\sigma^4}{n\beta s_x^4} + \frac{5\theta_2 s^4}{n\beta s_x^8} \right]
\]
\[ + \frac{\sigma^2}{n\beta^2} \left[ 1 + \frac{3\sigma^2}{n\beta^2 s_x^2} + \frac{3\theta_2 \sigma^2}{n \beta^2 s_x^2} \right] \]  

(15A)

where the error due to ignoring higher order terms is \( O(n^{-5/2}) \).

To recalculate \( E(\hat{W}_B) \) and \( E(\hat{W}_B - W)^2 \) one needs:

\[
E \varepsilon_S \varepsilon_S = \theta_1 \sigma^3 \left( \frac{n-1}{n} \right) \\
E (\varepsilon_S \varepsilon_S \varepsilon_S) = 0 \\
E \varepsilon_S \varepsilon_S^2 = \left( \frac{n-1}{n} \right) \sigma^4 + \left( \frac{n-1}{n^2} \right) \theta_2 \sigma^4
\]

Then again ignoring terms \( (\sigma/\beta s_x)^k \) for \( k > 4 \)

\[
E(\hat{W}_B) = W + (W-x) \left[ -\frac{(n-3)\sigma^2}{n\beta^2 s_x^2} - \frac{4\theta_1 s_x^2 \sigma^3}{2 \beta^2 s_x^3} + \frac{(n-5)(n-3)\sigma^4}{n^2 \beta^4 s_x^4} \\
+ \frac{8\theta_2 \sigma^4 s_x^4}{n \beta^4 s_x^8} + \frac{(n-9)(n-1)\theta_2 \sigma^4}{n \beta^4 s_x^8} \right]
\]

\[
\theta_1 \sigma^3 (n-3) + \frac{4\theta_2 s_x^2 \sigma^4}{n \beta^4 s_x^8}
\]  

(16A)

\[
E(\hat{W}_B - W)^2 = (W-x)^2 \left[ \frac{\sigma^2}{n\beta^2 s_x^2} - \frac{4\theta_1 s_x^2 \sigma^3}{2 \beta^2 s_x^3} + \frac{(n-5)^2 \sigma^4}{2 \beta^4 s_x^4} + \frac{(n-11)(n-1)\theta_2 \sigma^4}{n^2 \beta^4 s_x^4} \\
+ \frac{12\theta_2 s_x^4 \sigma^4}{n \beta^4 s_x^8} \right] + 2(W-x) \left[ -\frac{(n-4)\theta_1 \sigma^3}{2 \beta^2 s_x^3} + \frac{8\theta_2 s_x^2 \sigma^4}{n \beta^4 s_x^8} \right]
\]

\[
+ \frac{\sigma^2}{n\beta^2} + \frac{5\theta_2 \sigma^4}{n \beta^4 s_x^2} - \frac{2(n-1)\sigma^4}{n \beta^4 s_x^2} - \frac{2(n-1)\theta_2 \sigma^4}{n \beta^4 s_x^2}
\]

25
\[
\frac{r^2}{\beta^2} \left[ 1 - \frac{(2n-n) \sigma^2}{n \beta^2 s^2_x} - \frac{126 \frac{n-3}{n} \sigma^3}{n \beta^2 s^3_x} + \frac{(3n-5)(n-11) \sigma^4}{n \beta^4 s^4_x} \right] + \frac{(3n-29)(n-1) \sigma^4}{n \beta^4 s^4_x} + \frac{286 \sigma^4}{n \beta^4 s^4_x}
\]
(17A)

Assuming \( E_n \) has a finite upper bound over \( n \) for some \( a > 2 \), the strong law of large numbers can be used to show that for large \( n \) the expected values of the remainder terms are
\[
0 \left( \frac{\sigma^6}{\beta^6 s^6_x} \right). \]
The rate of convergence depends on the distribution of \( u_n = (2 \beta s^2_x + s^2_{\varepsilon_x})/\beta^2 s^2_{xx} \) near -1.

Finally, for \( n \) increasing and \( \frac{\sigma^2}{\beta^2 s^2_x} \) small,
\[
\hat{E}_{\hat{W}_B} = \hat{W} + (\hat{W} - \hat{X})(- \frac{\sigma^2}{\beta^2 s^2_x} + \frac{\sigma^4}{\beta^4 s^4_x} + O(b_n) + O\left(\frac{\sigma^6}{\beta^6 s^6_x}\right))
\]
(18A)

\[
E(\hat{W}_B - \hat{W})^2 = (W - \bar{X})^2 \left[ \frac{\sigma^4}{\beta^4 s^4_x} + O(b_n) + O\left(\frac{\sigma^6}{\beta^6 s^6_x}\right) \right]
\]

\[
+ 2(W - \bar{X}) O(b_n) + \frac{\sigma^2}{n \beta^2 s^2_x} \left[ 1 + O\left(\frac{\sigma^2}{\beta^2 s^2_x}\right) + O(b_n) \right]
\]

\[
+ \frac{\sigma^2}{\beta^2} \left[ 1 - \frac{2 \sigma^2}{\beta^2 s^2_x} + \frac{3 \sigma^4}{\beta^4 s^4_x} + O(b_n) + O\left(\frac{\sigma^6}{\beta^6 s^6_x}\right) \right]
\]
(19A)

where \( \lim_{n \to \infty} b_n = 0 \) and \( O\left(\frac{\sigma}{\beta^2 s^2_x}\right)^k \) is the limit over \( n \) of the appropriate remainder term.
Effect of Quadratic Term

Suppose it has been assumed that \( y_i = \alpha + \beta x_i + \epsilon_i \), but in fact
\[ y_i = \alpha + \beta x_i + \gamma(x_i - \bar{x})^2 + \epsilon_i \]
and \( Z = \alpha + \beta \bar{W} + \gamma(\bar{W} - \bar{x})^2 + \epsilon \) where
\( \epsilon_i \sim N(0, \sigma^2) \) and \( \epsilon \sim N(0, \tau^2) \). The \( x_i \) are chosen symmetrically about \( \bar{x} \), so that \( \sum_{i=1}^{n} (x_i - \bar{x})^k = 0 \) for \( k \) odd. Let
\[
S_{x_2^2} = \sum_{i=1}^{n} (x_i - \bar{x})^2 - s_{x_2}^2
\]
\[
S_{x_2^e} = \sum_{i=1}^{n} \left[ (x_i - \bar{x})^2 - s_{x}^2 \right] \left[ \epsilon_i - \bar{\epsilon} \right]
\]
Because of the symmetry of the \( x \)'s, \( \mathbb{E}[S_{x_2^e}/S_{x_2}] = 0 \) for \( k \) odd. Using method A
\[
\hat{\beta} = \frac{\mathbf{S} (x_i - \bar{x})[\beta(x_i - \bar{x}) + \gamma((x_i - \bar{x})^2 - s_{x}^2) + \epsilon_i - \bar{\epsilon}]}{S_{xx}}
\]
\[
= \frac{\beta S_{xx} + S_{x_2^e}}{S_{xx}}
\]
so \( \hat{\beta} \) is unaffected by the presence of the quadratic.
\[
\hat{W}_A = \frac{\beta(\bar{W} - \bar{x}) + \gamma(\bar{W} - \bar{x})^2 - s_{x}^2 + (\epsilon - \bar{\epsilon})}{\hat{\beta}^2} + \bar{x}
\]
For large \( n \), \( 1/\hat{\beta} \approx 1/\beta \) and
\[
\hat{W}_A \approx \bar{W} + \gamma/\beta \left[ (\bar{W} - \bar{x})^2 - s_{x}^2 \right] + \frac{(\epsilon - \bar{\epsilon})}{\beta}
\]
so that \( \gamma = 0 \) increases \( \mathbb{E}[\hat{W}_A - \bar{W}]^2 \) by the term \( \frac{\gamma^2}{\beta^2} \left[ (\bar{W} - \bar{x})^2 - s_{x}^2 \right]^2 \).

Using method B
\[
\hat{\delta} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (y_i - \bar{y})^2}
\]
The quadratic term does
affect this estimate. Now

\[ \delta = \left\{ \frac{\beta S_{xx} + S_{xc}}{\beta^2 S_{xx} + \gamma S_{x_c} S_{x_2}} \right\} \left\{ 1 + \frac{2(\beta S_{x_c} + \gamma S_{x_2}) + S_{cc}}{\beta^2 S_{xx} + \gamma S_{x_2} S_{x_2}} \right\}^{-1} \]

Let \( c = 1 + \frac{\sigma^2}{\beta^2 S_{xx}} \). Under the same assumptions on \( \frac{\sigma^2}{\beta^2 S_{xx}} \) as before, the expansions for \( \frac{1}{1+u} \) and \( \frac{1}{(1+u)^2} \) give good approximations where

\[ 2(\beta S_{x_c} + \gamma S_{x_2}) + S_{cc} \]

now \( u = \frac{n^2 \sigma^2}{\beta^2 S_{xx}} \). Then

\[ E\delta \approx \frac{1}{\beta c} \left\{ 1 - \frac{(n-3)\sigma^2}{n^2 \beta^2 S_{xx}} - \frac{(n-3n+15)\sigma^4}{n^4 \beta^4 S_{xx}^2} \right\} \]  

(20A)

For large \( n \)

\[ E\delta^2 \approx \frac{1}{\beta^2 c^2} \left\{ 1 + \frac{\sigma^2}{\beta^2 S_{xx}} \right\}^{-2} \]

and hence

\[ E(W_0 - W)^2 = \frac{(W - \bar{x})^2}{(W - \bar{x})^2} + 2(W - \bar{x})[\beta(W - \bar{x}) + \gamma ((W - \bar{x})^2 - S_{xx})][1 + \frac{\sigma^2}{\beta^2 S_{xx}}]^{-1} \frac{1}{\beta c} \]

\[ + \left[ \beta(W - \bar{x}) + \gamma ((W - \bar{x})^2 - S_{xx}) \right] \left[ 1 + \frac{\sigma^2}{\beta^2 S_{xx}} \right]^{-2} \frac{1}{\beta^2 c^2} \]

\[ + \frac{\sigma^2}{\beta^2 c} \left[ 1 + \frac{\sigma^2}{\beta^2 S_{xx}} \right]^{-2} \frac{1}{\beta^2 c^2} \]  

(21A)
REFERENCES


