MISSING VALUES IN MULTIVARIATE STATISTICS

II. POINT ESTIMATION IN SIMPLE LINEAR REGRESSION*

by

A. A. Afifi† and R. M. Elashoff §

University of California, Berkeley, and
University of Wisconsin, Madison;

and

University of California, Berkeley.

* This Research was supported in part by Public Health Grants 43-64-562
and GM-10525, by National Science Foundation Grant CP-2593, and ONR
Contract 1202-17, Project No. NR 042-222.

† Now at University of California, Los Angeles.

§ Now at University of California Medical Center, San Francisco.
0. Summary

We reviewed in [1] the literature dealing with missing observations in regression analysis. In this second paper, we study the probabilistic behavior of several ways to estimate the linear regression function between two random variables. We derive the mean square error of prediction for each method of estimation. Tables are given to characterize in terms of the correlation coefficient those situations where a given method has smaller mean square error than its competitors.

In Section 1, we specify the model and present some results concerning the classical least squares method. In Sections 2, 3, 4, we discuss the zero order regression method, a modified zero order method, and the first order regression method, respectively (see [1] for a general description of these methods.) Section 5 contains general conclusions and recommendations for the use of these estimation procedures.

1. The model and classical least squares.

Let $y$ denote the dependent variable and $x$ denote the independent variable. Assume that

$$y_{l} = \beta_0 + \beta_1 (x_{l} - \mu_x) + e_{l}; \quad l = 1, \ldots, n;$$

(1.1)

$$\text{ave} \quad e_{l} = 0; \quad \text{cov} (e_{l}, e_{l'}) = \begin{cases} \sigma^2 & \text{if } l = l' \\ 0 & \text{otherwise} \end{cases}$$

$x_{l}$ is $N(\mu_x, \sigma_x^2)$; $e_{l}$ and $x_{l}$ are independent.
We have \( n \) independent bivariate observations \( z^*_\ell = (y^*_\ell, x^*_\ell) \), where

\( n_y \) observations are recorded on \( y \), \( n_x \) observations are recorded on \( x \) and

\( n_c \) bivariate observations are complete. Thus, the mean and variance of \( y^*_\ell \),
conditional upon the observed \( x \)'s are

\[
\begin{align*}
\text{ave } y|x^*_\ell &= \mu_y + \beta(x^*_\ell - \mu_x) \\
\text{var } y|x^*_\ell &= \sigma^2
\end{align*}
\]

(1.2a) if \( x^*_\ell \) is observed,

\[
\begin{align*}
\text{ave } y|x^*_\ell &= \mu_y \\
\text{var } y|x^*_\ell &= \sigma^2 + \beta^2 \sigma_x^2 = \sigma_y^2
\end{align*}
\]

(1.2b) if \( x^*_\ell \) is not observed.

The classical way to estimate the parameters of (1.1) is least squares.
Using this method we estimate \( \mu_y \) and \( \beta \) by substituting

\[
\begin{align*}
\sum_{n_c}^* x^*_\ell / n_c & \quad \text{for } \mu_x \\
\sum_{n_c}^* [y^*_\ell - \mu_y - \beta(x^*_\ell - x.)]^2 & \quad \text{with respect to } \mu_y \text{ and } \beta.
\end{align*}
\]

These estimators are

* \( \sum_{n_c}^* \) denotes the summation over all indices \( \ell \) such that both \( x^*_\ell \) and \( y^*_\ell \) are observed, \( \sum_{n_x}^* \) denotes the summation over all indices \( \ell \) such that \( x^*_\ell \) is observed, etc.
\[ (1.3) \quad \hat{\mu}_{(ls)} = \frac{\sum_{c} y_{l}}{n_{c}} = y, \quad \hat{\beta} = \frac{\sum_{c} y_{l} (x_{l} - x_{c})}{\sum_{c} (x_{l} - x_{c})^{2}}. \]

The conditional means and variances of the estimators are well known. The unconditional means and variances are

\[ \text{ave}_{y} \hat{\mu}_{y} = \mu_{y} ; \quad \text{ave}_{y} \hat{\beta}_{(ls)} = \beta ; \]

\[ \text{var}_{y} \hat{\mu}_{y} = \frac{\sigma_{y}^{2}}{n_{c}} ; \quad \text{var}_{y} \hat{\beta}_{(ls)} = \frac{\sigma^{2}}{\sigma_{x}^{2}(n_{c} - 3)} . \]

The unconditional variance of \( y_{(ls)} \) for a fixed \( x_{o} \) is

\[ (1.5) \quad \text{var}_{y|x_{o}} \hat{y}_{(ls)} = \frac{\sigma_{y}^{2}}{n_{c} - 3} \left( 1 - \frac{2}{n_{c}} + \frac{(x_{o} - \mu_{x})^{2}}{\sigma_{x}^{2}} \right) , \]

and, if \( x_{o} \) is random with mean \( \mu_{x} \) and variance \( \sigma_{x}^{2} \), (1.5) becomes

\[ (1.6) \quad \text{var}_{y} \hat{y}_{(ls)} = \frac{\sigma_{y}^{2}}{n_{c}} + \frac{\sigma_{x}^{2}(n_{c} + 1)}{n_{c}(n_{c} - 3)} + \beta^{2} \sigma_{x}^{2} . \]

The conditional and unconditional distributions for \( \hat{\mu}_{y} \), \( \hat{\beta} \), \( \hat{\mu}_{x} \)
were obtained by Karl Pearson [2] when \( z \) is bivariate normal. In addition, he found the first 8 central moments of \( y_{(ls)} \) conditional only on \( x_{o} \)
gave some approximations for the distribution. Alternatively, we may represent
the density of \( \hat{y}(z) \) by the first four terms of an Edgeworth expansion.

This expansion is facilitated by the fact that all the odd moments of \( \hat{y}(z) \) are equal to zero.

2. Zero order regression method.

The zero order technique is defined by the following steps. First, estimate each of the \( m_x \) missing \( x \) observations by \( x = \sum \frac{x_i}{n_x} \) and each of the \( m_y \) missing \( y \) observations by \( y = \sum \frac{y_i}{n_y} \). Next, substitute \( x \) for \( \mu_x \). Then, the zero order estimators for \( \nu_y \) and \( \beta \) are those which minimize the total sum of squares based on the completed sample.

The estimators are

\[
(2.1) \quad \hat{\nu_y} = y, \quad \hat{\beta} = \sum \frac{\left( \frac{x_{i,x} - \bar{x}_x}{n_x} \right) \left( \frac{y_{i,y} - \bar{y}_y}{n_y} \right)}{\sum \frac{(x_{i,x} - \bar{x}_x)^2}{n_x}}.
\]

We find the conditional means to be

\[
(2.2) \quad \text{ave}_{y|x} \hat{\nu_y} = \nu_y + \frac{n_c}{n_y} \beta \left( x - \mu_x \right),
\]

\[
(2.3) \quad \text{ave}_{y|x} \hat{\beta} = \frac{\beta}{\sum \frac{(x_{i,x} - \bar{x}_x)^2}{n_x}} \left[ \sum \frac{(x_{i,c} - \bar{x}_c)^2}{n_c} \right. \\
\left. + \frac{m_x n_c}{n_y} \left( \frac{(x_{i,c} - \bar{x}_c)}{(x_{i,c} - \mu_x)} \right) \left( \frac{(x_{i,c} - \mu_x)}{(x_{i,c} - \mu_x)} \right) \right].
\]
\[
(2.4) \quad \text{ave } y|x \hat{y}(0) = \text{ave } y|x \mu_y(0) + \left( x_o - x_\cdot \right) \text{ave } y|x \hat{\beta}(0)
\]

and the conditional variances to be

\[
(2.5) \quad \text{var } y|x \mu_y(0) = \frac{\sigma^2}{n_y} + \frac{m_x}{n_y^2} \beta^2 \sigma_x^2,
\]

\[
(2.6) \quad \text{var } y|x \hat{\beta}(0) = \frac{\sigma^2}{\sum_{n_c} \left( x_\cdot - x_\cdot \right)} \left\{ \frac{\sigma^2}{\sum_{n_x} \left( x_\cdot - x_\cdot \right)^2} \right\} \left( \frac{m_x}{n_y} \left( m_y - n_x \right) \right)^2 + \frac{m_x m_y}{n_c n_y} \left( \frac{m_y}{n_y} \left( m_y - n_x \right) \right)^2 \beta^2 \frac{\sigma_x^2}{\sigma^2} \right\},
\]

\[
(2.7) \quad \text{var } y|x \hat{y}(0) = \text{var } y|x \mu_y(0) + \left( x_o - x_\cdot \right)^2 \text{var } y|x \hat{\beta}(0)
\]

\[
+ \frac{2m_x m_y}{n_y^2} \left( x_o - x_\cdot \right) \left( \frac{m_y}{n_y} \left( m_y - n_x \right) \right) \beta^2 \frac{\sigma_x^2}{\sigma^2}.
\]

The mean square errors may be obtained by adding the square of the bias to the variance.

A few remarks about these formulas are in order. If \( n_c = n \), that is, if all the observations are complete, these formulas reduce to the corresponding ones for least squares. If only \( y_n \) is missing, so that \( n_c = n_y = n - 1 \) and \( n_x = n \), the conditional means become
\[ \text{ave}_{y|x} \mu_y^{(0)} = \mu_y + \beta(x_{(n)} - x_x) , \]

\[ (2.8) \]

\[ \text{ave}_{y|x} \beta^{(0)} = \beta \left[ 1 - \left( \frac{n}{n-1} \right) \frac{(x_n - x_{(n)})^2}{\sum_{k=1}^{n-1} (x_k - x_{(n)})^2} \right] , \]

\[ \text{ave}_{y|x} y^{(0)} = \text{ave}_{y|x} \mu_y^{(0)} + (x_{(n)} - x_x) \text{ave}_{y|x} \beta^{(0)} . \]

In this special situation, \( \hat{\beta}^{(0)} \) will usually underestimate \( \beta \) on the average.

The size of the bias depends on how far \( x_n \) deviates from \( x_{(n)} \). If we are interested in predictions of \( y \) for \( x \)-values near \( x_{(n)} \), then, on the average, the bias in our prediction will be comparatively small.

Now, suppose that \( n_c = n - 1 = n_x \) and \( n_y = n \). Then we have

\[ \text{ave}_{y|x} \mu_y^{(0)} = \mu_y + \frac{n-1}{n} \beta (x_{(n-1)} - x_x) , \]

\[ (2.9) \]

\[ \text{ave}_{y|x} \beta^{(0)} = \beta , \]

\[ \text{ave}_{y|x} y^{(0)} = \mu_y + \beta(x_{(n)} - x_x) + \frac{1}{n} \beta(x_{(n-1)} - x_x) . \]

The bias in the prediction of \( y \) comes from \( \mu_y^{(0)} \), since \( \hat{\beta}^{(0)} \) in this situation is simply the usual least squares estimator. In fact, \( \hat{\beta}^{(0)} = \hat{\beta}(\text{ls}) \) for arbitrary \( n_x \) is only \( x \)'s are missing. If the mean \( x \) observation is close to \( \mu_x \), the bias will be relatively small, which concurs with intuition.

Next, we turn our attention to the unconditional means and variances in the general problem. The unconditional means are
\begin{align*}
\text{ave}_{y} \hat{\mu}_{y} &= \mu_{y}, \\
\text{ave}_{y} \hat{\beta}(0) &= \beta \left[ \frac{n_{x} - 1}{n_{x}} + \frac{m_{x} m_{y}}{n_{x} n_{y} (n_{x} - 1)} \right], \\
\text{ave}_{y|x_{o}} \hat{\mu}_{y} &= \mu_{y} + (x_{o} - \mu_{x}) \text{ ave}_{y} \hat{\beta}(0), \\
\text{ave}_{y} \hat{\gamma}(0) &= \mu_{y}. 
\end{align*}

The unconditional variances are

\begin{align*}
(2.11) \quad \text{var}_{y} \hat{\mu}_{y} &= \text{mse}_{y} \hat{\mu}_{y} = \frac{\sigma_{y}^{2}}{n_{y}}, \\
(2.12) \quad \text{var}_{y} \hat{\beta}(0) &= \frac{\sigma_{y}^{2}}{\sigma_{x}^{2} (n_{x} - 3)} \left[ \frac{n_{c} - 1}{n_{x} - 1} + \frac{m_{x} m_{y}}{n_{x} n_{y} (n_{x} - 1)} \right] \\
&\quad + \beta^{2} \left( \frac{m_{x} m_{y} n_{c}}{n_{x} n_{y} (n_{x} - 1) (n_{x} - 3)} + \frac{m_{y}^{2} m_{n} n_{c}}{n_{y}^{2} n_{x} (n_{x} - 1) (n_{x} - 3)} \right) \\
&\quad + \frac{2}{(n_{x} - 1) (n_{x} + 1)} \left[ m_{y} (n_{c} - 1) + \frac{m_{y}^{2} m_{n} (n_{x} - 2)}{n_{x} y_{x}} - \frac{2 m_{y} m_{n} (n_{x} - 1)}{n_{x} y_{x}} \right].
\end{align*}
(2.13) \[ \text{var}_{y \mid x_0} \hat{y}^{(0)} \]

\[
= \sigma^2 \left\{ \frac{1}{n_y} + \frac{1}{(n_x - 1)(n_x - 3)} \left[ \frac{1}{n_x^2} + \frac{(x_0 - \mu_x)^2}{\sigma_x^2} \right] \left[ n_x - 1 + \frac{m_m \cdot m_y}{n_x \cdot n_y} \right] \right\}
\]

\[
+ \beta^2 \sigma^2_x \left\{ \frac{1}{n_y} + \frac{m_m \cdot m_y}{n_x^2 (n_x - 1)^2} + \frac{2m_m \cdot m_y (n_x - 1)}{n_x^2 n_y (n_x - 1)^2} \right. \\
- \frac{2m_m \cdot m_x \cdot n_y \cdot c}{n_x^3 n_y (n_x - 1)(n_x - 3)} - \frac{2n_c (n_x - 1)}{n_x n_y (n_x - 1)} - \frac{4m_m \cdot m_y \cdot n_c}{n_x^2 n_y (n_x - 1)} \\
- \frac{(x_0 - \mu_x)^2}{\sigma_x^2} + \frac{(n_x - 1)^2}{(n_x - 3)} \left. \left[ \frac{1}{n_x^2} + \frac{(x_0 - \mu_x)^2}{\sigma_x^2} \right] \right. \\
- \left. \left[ \frac{m_m \cdot m_y \cdot n_y \cdot c}{n_x^2 n_y} + \frac{(n_x - 1)(n_x + 1)}{(n_x - 3)(n_x + 1)} + \frac{2m_m \cdot m_y (n_x - 2)}{n_y^2 n_x (n_x - 1)^2(n_x + 1)} \right. \\
- \frac{4m_m \cdot m_y \cdot n_c}{n_x n_y (n_x - 1)(n_x + 1)} + \frac{m_m \cdot m_y \cdot n_c}{n_x^2 n_y (n_x - 1)(n_x - 3)} \right\}.
\]
\[ \text{(2.14)} \quad \text{var } y \hat{\gamma}'(0) = \text{mse}_y \hat{\gamma}'(0) \]

\[
= \sigma^2 \left\{ \frac{1}{n_y} + \left[ \frac{(n\_x+1)}{n\_x(n\_x-1)(n\_x-3)} \right] \left[ (n\_c-1) + \frac{m\_m}{n\_x n\_y} \right] \right\} \\
+ \beta^2 \sigma^2 \left\{ \frac{1}{n_y} + \frac{m\_m n\_c (n\_x+1)}{n\_x^2 n\_y^2} + \frac{(n\_c-1)(n\_x+1)}{n\_x(n\_x-1)} + \frac{3m^2 m^2}{n\_x^3 n\_x n\_y (n\_x-1)} \right. \\
+ \frac{2m\_m n\_c (n\_x-1)}{n\_x^2 n\_y (n\_x-1)} + \left. \frac{3m^2 m n\_c (n\_x+1)}{n\_x^3 n\_y (n\_x-1)(n\_x-3)} \right. \\
- \frac{2n\_c (n\_x-1)}{n\_x n\_y (n\_x-1)} - \frac{4m\_m n\_c}{n\_x^2 n\_y (n\_x-1)} \left\} \right.
\]

Note that if only \( y_n \) is missing, then \( \text{ave } y \hat{\gamma}'(0) = \frac{n\_x-1}{n} \beta \). Thus, \( \hat{\gamma}'(0) \)
underestimates \( \beta \), on the average, with small bias for moderate sample sizes.
If only \( x \)'s are missing, then the estimated regression line is an unbiased estimator of the population regression line.

We now discuss briefly the distribution theory of the estimators
\( \hat{\gamma}'(0) \), \( \hat{\beta}'(0) \), and \( \hat{\gamma}'(0) \). If we assume that \( z = (y,x)' \) is bivariate normal,
then, conditional upon the \( x \)'s, \( \hat{\mu}'_y \), \( \hat{\beta}'(0) \), and \( \hat{\gamma}'(0) \) are normal with mean
and variance as given by (2.2)-(2.11). Unconditionally, it is clear that \( \hat{\gamma}'(0) \)
is also normal with the unconditional mean and variance given by (2.10) and (2.11). The unconditional distribution of \( \hat{\beta}'(0) \) involves a troublesome integration problem, since both the conditional mean and variance depend upon \( x \).
3. A modified zero order regression method

In the classical least squares method, we substitute a parameter for each missing observation and minimize the sum of squares to get the estimators. In the zero order method, we substitute \( y \) for each missing \( y \) and \( x \) for each missing \( x \). Extending this approach we substitute functions \( u \) and \( v \) of the observed \( x \)'s and \( y \)'s for each missing \( y \) and \( x \) observation, respectively. Our object is to choose \( u, v, \alpha, \beta \) to minimize

\[
(3.1) \quad \sum_{n_c} (y - \alpha - \beta x)^2 + \sum_{m_y} (u - \alpha - \beta x)^2 + \sum_{m_x} (y - \alpha - \beta v)^2.
\]

The solutions to the minimization problem are

\[
(3.2) \quad \hat{\mu} = \frac{\sum_{n_c} (x - \hat{\mu}) y_n}{\sum_{n_c} (x - \hat{\mu})^2 + \sum_{n_c} (y - \hat{\mu})^2}, \quad \hat{\beta} = \frac{\sum_{n_c} (x - \hat{\mu}) (y_n - \hat{\mu})}{\sum_{n_c} (x - \hat{\mu})^2 + \sum_{n_c} (y_n - \hat{\mu})^2}
\]

\[
(3.3) \quad \hat{y} = \hat{\mu} + \hat{\beta} (x - \hat{\mu}).
\]

In the special situations where only \( x \)'s are missing or where only one \( y \) is missing, these estimators are identical to the classical least squares estimators. The missing values are estimated by

\[
(3.4) \quad \hat{u} = \hat{\mu} + \hat{\beta} (x - \hat{\mu}), \quad \hat{v} = \hat{\mu} + \frac{\hat{\beta} (y - \hat{\mu})}{\hat{\beta}}.
\]
In what follows we assume that \( m_y > 1 \). The conditional means are

\[
(3.5) \quad \text{ave } y | x \{ \hat{\mu}_y \} = \mu_y + \beta \left( \frac{(n_c)}{x - \mu_x} \right),
\]

\[
(3.6) \quad \text{ave } y | x \{ \hat{\beta} \} = \frac{\sum_{n_c} \left( \frac{(n_c)}{x - \mu_x} \right)^2}{\left[ \sum_{n_c} \left( \frac{(n_c)}{x - \mu_x} \right)^2 + \sum_{m_y} \left( \frac{(m_y)}{x - \mu_x} \right)^2 \right]^2} \beta,
\]

\[
(3.7) \quad \text{ave } y | x, x_0 \{ \hat{\beta} \} = \mu_y + \beta \left( \frac{(n_c)}{x - \mu_x} \right)
\quad + \left( \frac{(x_0 - \mu_x)}{\sum_{n_c} \left( \frac{(n_c)}{x - \mu_x} \right)^2} + \sum_{m_y} \left( \frac{(m_y)}{x - \mu_x} \right)^2 \right) \beta,
\]

and the conditional variances are

\[
(3.8) \quad \text{var } y | x \{ \hat{\mu}_y \} = \frac{\sigma^2}{n_c},
\]

\[
(3.9) \quad \text{var } y | x \{ \hat{\beta} \} = \sigma^2 \frac{\sum_{n_c} \left( \frac{(n_c)}{x - \mu_x} \right)^2}{\left[ \sum_{n_c} \left( \frac{(n_c)}{x - \mu_x} \right)^2 + \sum_{m_y} \left( \frac{(m_y)}{x - \mu_x} \right)^2 \right]^2},
\]

\[
(3.10) \quad \text{var } y | x, x_0 \{ \hat{\beta} \} = \frac{\sigma^2}{n_c} + \frac{\left( \frac{(x_0 - \mu_x)}{\sum_{n_c} \left( \frac{(n_c)}{x - \mu_x} \right)^2} + \sum_{m_y} \left( \frac{(m_y)}{x - \mu_x} \right)^2 \right)^2}{\sum_{n_c} \left( \frac{(n_c)}{x - \mu_x} \right)^2 + \sum_{m_y} \left( \frac{(m_y)}{x - \mu_x} \right)^2} \sigma^2,
\]
The unconditional means are

\[
(3.11) \quad \text{ave}_y \hat{y} = y, \\
(3.12) \quad \text{ave}_y \hat{\beta} = \frac{n_c - 1}{n_x - 2} \beta, \\
(3.13) \quad \text{ave}_y | x_o \ y' = u + \beta(x_o - u_x) \frac{n_c - 1}{n_x - 2}, \quad \text{ave}_y y' = u_y
\]

and the unconditional variances are

\[
(3.14) \quad \text{var}_y \hat{y} = \frac{\sigma_y^2}{n_c}, \\
(3.15) \quad \text{var}_y \hat{\beta} = \frac{n_c - 1}{n_x - 2} \frac{s_y^2}{s_x^2} + \frac{2(n_c - 1)(m - 1)}{n_x (n_x - 2)^2} \beta^2, \\
(3.16) \quad \text{var}_y | x_o y' = \frac{\sigma_y^2}{n_c} + \left[ \frac{1}{n_c} + \frac{(x_o - u_x)^2}{s_x^2} \right] \frac{n_c - 1}{n_x - 2} \frac{s_y^2}{s_x^2} + \frac{(m - 1)(m + 1)}{n_c n_x (n_x - 2)} \beta^2 \sigma_x^2 + \frac{2(m - 1)(n_c - 1)}{n_x (n_x - 2)^2} \beta^2 (x_o - u_x)^2, \\
(3.17) \quad \text{var}_y y' = \left[ \frac{1}{n_c} + \frac{(n_c - 1)(n_c + 1)}{n_c (n_x - 2)(n_x - 4)} \right] \sigma^2 + \frac{(n_c + 1)(m - 1)(m + 1)}{n_c n_x (n_x - 2)} \beta^2 \sigma_x^2.
\]
4. A first order regression method

We define a first order regression method as follows. First, compute the regression of $y$ on $x$ by classical least squares and estimate the missing $y$'s by this regression function. Second, compute the regression of $x$ on $y$ by classical least squares and estimate the missing $x$'s by this last regression function. Third, estimate $\mu_x$ by the sample mean of the observed and estimated $x$ values. Fourth, apply least squares to the completed sample, with $\mu_x$ estimated as above, to obtain estimators for $\mu_y$ and $\beta$.

Specifically, we substitute for a missing $y_i$

$$\hat{y}_i = y_{i.} + \hat{\beta}(ls)(x_i - x_{i.}),$$

(4.1)

$$\hat{\beta}(ls) = \frac{\sum (x_i - x_{i.})y_i}{\sum (x_i - x_{i.})^2};$$

and for a missing $x_i$

$$\hat{x}_i = x_{i.} + \hat{\delta}(ls)(y_i - y_{i.}),$$

(4.2)

$$\hat{\delta}(ls) = \frac{\sum (x_i - x_{i.})y_i}{\sum (y_i - y_{i.})^2}.$$

When least squares is applied to the completed sample, the following estimators of $\mu_y$ and $\beta$ are obtained:
\( \hat{\mu}_y(1) = \bar{y} = \frac{n_y}{n} \mu_{y*} \left( \frac{n_c}{m_x} \right) + \frac{m_x}{n} \mu_{x*} \left( \frac{m_y}{m_x} \right) \)

\[ + \frac{m_y}{n} \hat{\beta}(\text{LS}) \left( \begin{pmatrix} m_y \\ n_c \end{pmatrix} \right) \left( \begin{pmatrix} x_{y*} - x_{c*} \end{pmatrix} \right), \]

\( \hat{\beta}(1) = \frac{\sum_{i=1}^{n} (x_{i*} - \bar{x})(y_{i*} - \bar{y})}{\sum_{i=1}^{n} (x_{i*} - \bar{x})^2} \),

where

\( \bar{x} = \frac{n_x}{n} x_{c*} + \frac{m_x}{n} x_{y*} + \frac{m_x}{n} \hat{\beta}(\text{LS}) \left( \begin{pmatrix} m_y \\ n_c \end{pmatrix} \right) \left( \begin{pmatrix} y_{y*} - y_{c*} \end{pmatrix} \right) \).

The estimated regression equation at \( x_o \) is

\( \hat{y}(1) = \bar{y} + \hat{\beta}(1) (x_o - \bar{x}). \)

The conditional mean and variance of \( \hat{\mu}_y \) are

\( \text{ave } y|x \hat{\mu}_y = \mu_y + \frac{n_x}{n} \hat{\beta} \left( \frac{n_c}{m_x} \right) \left( x_{y*} - \mu_x \right) \),

\( \text{var } y|x \hat{\mu}_y = \frac{\sigma^2}{n^2} \left[ \frac{n_x^2}{n_c} + \frac{m_x^2 (x_{y*} - \mu_y)^2}{n_c} \right] + \frac{m_x}{n^2} \beta^2 \sigma^2 \)

while unconditionally these become.
\[ (4.9) \quad \text{ave}_{y} \hat{\mu}_{y}^{(1)} = \mu_{y}, \]

\[ (4.10) \quad \text{var}_{y} \hat{\mu}_{y}^{(1)} = \frac{s^2}{n^2} \left[ \frac{n_{x}^2}{n_{c}} + \frac{n_{x} m_{y}}{n_{c} (n_{c} - 3)} + \frac{s^2 o^2}{n} \right]. \]

The moments of \( \hat{\beta}^{(1)} \) and \( \hat{\gamma}^{(1)} \) seem to be hopeless to evaluate exactly, since both the numerator and the denominator in (4.4) contain the \( y \)'s. Using a Taylor expansion of \( \hat{\beta}^{(1)} \), we obtain its unconditional mean as

\[ (4.11) \quad \text{ave}_{y} \hat{\beta}^{(1)} = \beta \frac{n - 1 + \frac{m_{x} m_{y}}{n n_{c}} \left[ 1 - \frac{(1-\rho^2)^2}{n_{c} - 1} \right]}{(n_{x} - 1) + m_{x} \rho^2 + \left( \frac{m_{x} m_{y}}{n n_{c}} + \frac{m_{x}}{n_{c} - 3} + \frac{m_{x} m_{y}}{n n_{c} (n_{c} - 3)} \right) (1-\rho^2)} + o(n^{-2}). \]

Some remarks on expression (4.11) are in order. If \( m_{x} = m_{y} = 0 \), (4.11) reduces to \( \beta \) as expected. Clearly, this same result obtains if only \( m_{x} = 0 \), since the first order regression is the same as classical least squares when \( m_{x} = 0 \).

Next, we obtain the unconditional variance of \( \hat{\beta}^{(1)} \) by a Taylor expansion to terms of \( o(n^{-2}) \). The expression for this variance is several pages long and, therefore, will not be given.
\[
\text{mse } y|x_0 \leq \text{mse } y|x_0^* \quad \text{iff } b \leq r_3^- \ ,
\]

where \( r_1 \), \( r_2 \), and \( r_3 \) are obtained from (1.5), (2.13), and (3.16).

Computer programs were used to evaluate \( r_1 \), \( r_2 \), and \( r_3 \) for

\[
x_0' = 0.0(0.1)2.0 \ ; \quad n = 10(10)100 \ ; \quad m_x = 0(5)n - 10 \ ; \quad m_y = 5(5)n_x - 5.
\]

Examples of the values of \( r_1 \) are shown in Tables 5.5 and 5.6, of \( r_2 \) in Table 5.7, and of \( r_3 \) in Tables 5.8 and 5.9.

It is important to point out that, when no \( y \)'s are missing, the zero order regression method gives smaller mean square error of prediction than least squares, regardless of the number of missing \( x \)'s or the magnitude of the correlation coefficient.

No estimation techniques is uniformly best. For any given situation, we are able to make qualitative as well as quantitative comparisons by using the formulas for conditional and unconditional mean square errors given in this paper.
Table 5.3
Method Giving Smallest $\text{mse}_y \hat{\beta}$: $\rho^2 = 0.5$, $n = 40$

<table>
<thead>
<tr>
<th>$m_x$</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
<th>35</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.4
Method Giving Smallest $\text{mse}_y \hat{\beta}$: $\rho^2 = 0.75$, $n = 40$

<table>
<thead>
<tr>
<th>$m_x$</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
<th>35</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5.8
Comparison of $\text{mse } y|x_o$ $\hat{y}(0)$ and $\text{mse } y|x_o$ $\hat{y}'$

Values of $r_3$ for $x_o' = 0$, $n = 40$

<table>
<thead>
<tr>
<th>$m_y$</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
<th>35</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.032</td>
<td>0.025</td>
<td>0.021</td>
<td>0.019</td>
<td>0.017</td>
<td>0.016</td>
<td>0.015</td>
</tr>
<tr>
<td>5</td>
<td>0.396</td>
<td>0.369</td>
<td>0.396</td>
<td>0.498</td>
<td>0.791</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.490</td>
<td>0.527</td>
<td>0.648</td>
<td>0.983</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.553</td>
<td>0.673</td>
<td>0.989</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.624</td>
<td>0.899</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.741</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.9
Comparison of $\text{mse } y|x_o$ $\hat{y}(0)$ and $\text{mse } y|x_o$ $\hat{y}'$

Values of $r_3$ for $x_o' = \pm 1$, $n = 40$

<table>
<thead>
<tr>
<th>$m_y$</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
<th>35</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.126</td>
<td>0.068</td>
<td>0.045</td>
<td>0.033</td>
<td>0.026</td>
<td>0.021</td>
<td>0.017</td>
</tr>
<tr>
<td>5</td>
<td>0.169</td>
<td>0.117</td>
<td>0.106</td>
<td>0.118</td>
<td>0.169</td>
<td>0.409</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.200</td>
<td>0.156</td>
<td>0.164</td>
<td>0.236</td>
<td>0.619</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.241</td>
<td>0.219</td>
<td>0.300</td>
<td>0.809</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.303</td>
<td>0.360</td>
<td>0.949</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.419</td>
<td>0.972</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.790</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
REFERENCES
