EFFECTS OF LINEAR INEQUALITY CONSTRAINTS ON DISTRIBUTIONS
OF PARAMETER ESTIMATES IN THE STANDARD LINEAR MODEL

by R. Jansen

GIANNINI FOUNDATION OF
AGRICULTURAL ECONOMICS

JAN 8 1976

June 1975

Preliminary and Confidential
This paper deals with the problem of estimating parameters in the standard linear model, while linear inequality constraints are imposed upon the parameter values. In particular, the distributions of the estimators are considered, enabling us to discuss the magnitude of their means and their variances in comparison with those derived from the estimators in unrestricted multiple regression. It is shown that the joint probability density function (p.d.f) of these estimators is a weighted sum of two conditional p.d.f's; that these estimates are biased, but have smaller variances than the corresponding ordinary least squares estimates, at least if the errors are normally distributed.

In section two of this paper the problem is formulated in terms of the standard linear model. In section three the probability density functions of the estimators are derived after which in section four their means and variances are considered.
1. INTRODUCTION

One may wish to restrict the range of parameter values for logical, theoretical or empirical reasons. For instance, regression estimates of the marginal propensity to consume may be limited to the range [0,1], while non-negativity may be imposed upon the elasticity of the demand for a commodity with respect to its own price (Theil [6]).

Quadratic programming enables us to find the point estimates of the parameters, subject to linear inequality constraints (see, e.g. Judge and Takayama [2]). For this purpose several well-known algorithms are available, such as those developed by: van de Panne and Whinston [5] or Hadley [1].

The sampling distribution of an estimator and its statistical properties, however, cannot be derived as easily as in the case of unrestricted estimation. An important contribution to this issue has been made by Lovell and Prescott [4], who examined the properties of the so-called two-step estimator in multiple regression with a non-negativity constraint imposed upon one of the parameters. This two-step approach had already been suggested by Zellner [8], in analyzing the case of simple regression.

This paper presents an extension of the two-step approach, proposed by Lovell and Prescott in the sense that the distribution of these estimators are derived, in the case when one parameter is restricted to the range: $\hat{\beta}_1 > c$, with $c$ a fixed number. Since restricting the value of one parameter may affect the estimates of one or more other parameters (preferably in a theoretically desired direction), the outcome of such an analysis of the joint statistical behaviour of these estimates is of more than purely theoretical interest.

2. FORMULATION OF THE PROBLEM

We start from the standard form of the linear model:
\[ y = X\beta + \varepsilon, \quad (2.1) \]

with \( y \) a column vector of \( N \) observations on the variable to be explained,

\( X \) an \( N \times K \) matrix of \( K \) explanatory variables.

\( \beta \) a column vector of \( K \) parameters

and \( \varepsilon \) a column vector of \( N \) disturbances.

Moreover, we make the usual assumptions that the disturbance terms have a multivariate normal distribution, with zero mean and covariance matrix \( \sigma^2 I \), while \( X \) is non-stochastic and of full column rank.

Under these assumptions, the unrestricted O.L.S. estimator of \( \beta; \ b = (X'X)^{-1}X'y \) has a normal distribution with mean \( \beta \) and covariance matrix \( \sigma^2 (X'X)^{-1} \).

Next we restrict one of the estimators, for instance \( \hat{\beta}_1 \), to a certain non-stochastic fixed number, say \( c \), as a lower bound. Then the problem becomes minimizing the sum of squared residuals subject to this restriction. In formula:

Minimize: \[ e'e = (y - X\hat{\beta})'(y - X\hat{\beta}) \]

subject to: \[ \hat{\beta}_1 \geq c, \]

with \( e \) the estimate of \( \varepsilon \) and \( \hat{\beta} \) the estimate of the parameter vector \( \beta \), with \( \hat{\beta}_1 \), its first element.

By means of quadratic programming, we may arrive at point estimates of all parameters. In addition, however, we wish to obtain the joint probability density function of the parameter estimates, and in particular their covariance matrix.

3. DISTRIBUTION OF THE PARAMETERS

We distinguish two cases: 1. the unrestricted O.L.S. estimator \( b_1 \) is not exceeding the value \( c \) and 2. the
unrestricted O.L.S. estimator $b_1$ exceeds the value $c$.
In case 1 the restricted estimator of $\beta_1$; viz. $\hat{\beta}_1$, equals
its lower bound $c$.

Consequently the probability that case 1 appears:

$$P[\hat{\beta}_1 = c] = \int_{-\infty}^{c} (2\pi \sigma^2 v_{11})^{-\frac{3}{2}} \exp\left[-\frac{1}{2}(z_1 - \frac{\beta_1}{\sigma})^2\right]dz_1$$

$$= \int_{-\infty}^{c} p(z_1)dz_1$$

$$= \alpha \text{ (say)},$$

with $p(\hat{\beta}_1)$, the single variate normal density function
with mean $\beta_1$ and variance $\sigma^2 v_{11}$, in which $v_{11}$ is the first
element of the partitioned matrix:

$$V = (X'X)^{-1} = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix}$$

with $v_{11}$ a scalar, $v_{12}$ a row vector of length $K - 1$, $v_{21}$
a column vector of length $K - 1$, equalling $v_{12}'$, and $v_{22}$
a matrix of order $(K - 1) \times (K - 1)$.

The matrix $R = (X'X)$ may be partitioned accordingly. Relating $V$ to $R$ and vice versa we can easily derive the
following identities:

a. $v_{11} = (r_{11} - r_{12}R_{22}^{-1}r_{21})^{-1}$

b. $v_{22} = (R_{22} - r_{21}v_{11}^{-1}r_{12})$

$\ast$

c. $r_{11} = (v_{11} - v_{12}v_{22}^{-1}v_{21})^{-1}$

d. $R_{22} = (v_{22} - v_{21}v_{11}^{-1}v_{12})^{-1}$

(3.3)

e. $v_{11}^{-1}v_{12} = -r_{12}R_{22}^{-1}$

f. $v_{22}^{-1}v_{21} = -r_{21}r_{11}^{-1}$

In case 2 the restricted estimator of $\beta_1$ equals the
unrestricted O.L.S. estimator $b_1$. The probability that
case 2 appears is of course:
\[
P[\hat{\beta}_1 \geq c] = \int_c^\infty (2\pi \sigma^2 v_{11})^{-\frac{1}{2}} \exp\left[-\frac{1}{2} \left(\frac{z_1 - \beta_1}{\sigma v_{11}}\right)^2\right] dz_1
\]
\[
= \int_c^\infty p(z_1) dz_1
\]
\[
= 1 - \alpha
\]

Together with (3.1) we specify the distribution function of \(\hat{\beta}_1\) by:

\[
P(\hat{\beta}_1 < c) = 0
\]
\[
P(\hat{\beta}_1 = c) = \alpha \quad (3.5)
\]
\[
P(\hat{\beta}_1 > c) = 1 - \alpha
\]

The mean of this distribution:

\[
E[\hat{\beta}_1] = ac + \int_c^\infty z_1 p(z_1) dz_1
\]
\[
= ac + \beta_1 - \int_{-\infty}^c z_1 p(z_1) dz_1
\]
\[
= \beta_1 - \int_{-\infty}^c (z_1 - c) p(z_1) dz_1 \quad (3.6)
\]
\[
= \beta_1 + \alpha (c - \beta_1) + \sigma^2 v_{11} p(c) \quad (3.6.1)
\]

with \(p(c) = (2\pi \sigma^2 v_{11})^{-\frac{1}{2}} \exp\left[-\frac{1}{2} \left(\frac{c - \beta_1}{\sigma v_{11}}\right)^2\right]\).

From (3.6) we find: bias \(\hat{\beta}_1\) = \(- \int_{-\infty}^c (z_1 - c) p(z_1) dz_1\).

For the variance we obtain:
\[ E[\hat{\beta}_1 - E[\hat{\beta}_1]]^2 = E[\hat{\beta}_1^2] - (E[\hat{\beta}_1])^2 \]
\[ = ac^2 + \int_c^\infty z_1^2 p(z_1)dz_1 - (E[\hat{\beta}_1])^2 \]
\[ = (1 - \alpha)\sigma^2 v_{11} - (\sigma^2 v_{11} p(c))^2 + \alpha(1 - \alpha)(c - \beta_1)^2 + \]
\[ + (1 - 2\alpha)(c - \beta_1)\sigma^2 v_{11} p(c) \]
\[ = (1 - \alpha)\sigma^2 v_{11} + \int_c^\infty (z_1 - c)p(z_1)dz_1 \cdot \int_c^\infty (z_1 - c)p(z_1)dz_1 \]  \hspace{1cm} (3.6.2)

Since
\[ \int_{-\infty}^c (z_1 - c)p(z_1)dz_1 = \int_{-\infty}^0 yt(y)dy, \text{ with } y = z_1 - c, \]

is never positive and \( \int_0^\infty yt(y)dy \) is non-negative, the bias of \( \hat{\beta}_1 \) is non-negative and the variance of \( \hat{\beta}_1 \) is less than or equal to the least squares variance.

Next, we examine the manner in which the restriction on the estimate of \( \beta_1 \) affects the distributions of the estimates of the remaining parameters, denoted by the vector \( \hat{\beta}_- = \{\hat{\beta}_2, \ldots, \hat{\beta}_K\} \). First we consider the conditional distribution of \( \hat{\beta}_- \), given \( \hat{\beta}_1 = c \); viz. \( \mathbb{h}^{(1)}(\hat{\beta}_-|\hat{\beta}_1 = c) \).

It is obvious that the estimator \( \hat{\beta}_- \), in case \( \hat{\beta}_1 = c \), follows from estimating the parameters in the model:
\[ y = x_1c + X_\beta_- + \eta \]  \hspace{1cm} (3.7)

with \( x_1 \) the first column of the \( X \)-matrix, \( X_- \) the \( X \)-matrix without the first column and \( \eta \) the vector of disturbances, related to the original disturbances \( \epsilon \), according to:
\[ \eta = \epsilon + x_1(\beta_1 - c) \]
Consequently η has a multivariate normal distribution with mean \(x_1(\beta_1 - c)\) and covariance matrix \(\sigma^2 I\).

Given \(\hat{\beta}_1 = c\), \(\hat{\beta}_n = (X'X)^{-1}X'(y - x_1c)\) \hspace{1cm} (3.8)

Consequently, \(\hat{\beta}_n\) given \(\hat{\beta}_1 = c\) has a normal distribution with mean \((X'X)^{-1}X'(X\beta - x_1c)\) and covariance matrix \(\sigma^2(X'X)^{-1}\).

Using definition 3.2 and 3.3.e:

\[
E[\hat{\beta}_n | \hat{\beta}_1 = c] = (X'X)^{-1}X'(X\beta - x_1c) = (X'X)^{-1}X'_n(X_\beta - x_1(\beta_1 - c)) \hspace{1cm} (3.9)
\]

\[
= \beta_n + R_{22}^{-1}r_{21}(\beta_1 - c)
\]

Summarizing:

\[
h_1(\hat{\beta}_n | \hat{\beta}_1 = c) \approx n(\beta_n + R_{22}^{-1}r_{21}(\beta_1 - c), \sigma^2 R_{22}^{-1})
\hspace{1cm} (3.10)
\]

\[; -\infty < \hat{\beta}_2 < \infty, \ldots, -\infty < \hat{\beta}_K < \infty.\]

The next step is to find the distribution of \(\hat{\beta}_n\), given \(\hat{\beta}_1 > c\): viz. \(h_2(\hat{\beta}_n | \hat{\beta}_1 > c)\).

Starting from the conditional distribution of all estimates of \(\beta_i\), given \(\hat{\beta}_1 > 0\); \(i = 1, \ldots, K\), we may write:

\[
h(\hat{\beta} | \hat{\beta}_1 > c) = \frac{g(\hat{\beta})}{P(\hat{\beta}_1 > c)} = \frac{1}{1 - \alpha} g(\hat{\beta}): c < \hat{\beta}_1 < \infty, -\infty < \hat{\beta}_2 < \infty, \ldots,\]

\[; -\infty < \hat{\beta}_K < \infty, \hspace{1cm} (3.11)\]

by virtue of eq. (3.4), with \(g(\hat{\beta})\) the multivariate normal probability density function with mean \(\beta\) and covar. matrix.
From this expression we are able to derive $h_2(\beta_1 > c)$ by integrating both sides of the identity (3.11) over $\hat{\beta}_1$, which results in:

$$h_2(\beta_1 > c) = \frac{1}{1 - \alpha} \int_0^\infty g(z_1, \hat{\beta}_2, \ldots, \hat{\beta}_k) dz_1$$

with $h_2(\beta_1 > c)$ the multivariate normal p.d.f. of $\beta_1$, given a certain value of $\beta_1$, say $d$, with mean $(\beta_1 - \text{R}_{22}^{-1}(d - \beta_1))$ and covar. matrix $\sigma^2 \text{R}_{22}^{-1}$. The marginal distribution of $\beta_1$, denoted by $f_\beta(\hat{\beta}_1)$, could be derived from the two conditional distributions stated in eqs. (3.10) and (3.12). To see clearly what is going on we define the transformation function $\pi = \pi(\hat{\beta}_1)$ in such way that:

$$\pi = 0 \quad \text{for } \hat{\beta}_1 = c$$

$$\pi = 1 \quad \text{for } \hat{\beta}_1 > c$$

Hence, the p.d.f. of $\pi$ is of the Bernouilli kind:

$$q(\pi) = \alpha^{(1-\pi)}(1 - \alpha)^\pi; \quad \pi = 0, 1$$

$$= 0 \quad \text{; elsewhere}$$

By definition we write the joint distribution of $\hat{\beta}_1$ and $\pi$ as:

$$k(\hat{\beta}_1, \pi) \equiv l(\hat{\beta}_1 | \pi).q(\pi)$$

with $l(\hat{\beta}_1 | \pi)$ the conditional p.d.f. of $\hat{\beta}_1$, given a certain value of $\pi$. It follows immediate that:
\[ f_-(\hat{\beta}_-) = \sum_{\pi=0}^{1} l(\hat{\beta}_-|\pi).q(\pi) \]

\[ = l(\hat{\beta}_-|\pi = 0).\alpha + l(\hat{\beta}_-|\pi = 1).(1 - \alpha) \]

\[ = \alpha h_-^{(1)}(\hat{\beta}_-|\hat{\beta}_1 = c) + (1 - \alpha)h_-^{(2)}(\hat{\beta}_-|\hat{\beta}_1 > c); \quad (3.16) \]

\[ ; -\infty < \hat{\beta}_2 < \infty, \ldots, -\infty < \hat{\beta}_K < \infty \]

\[ = 0 \quad \text{; elsewhere} \]

in which \(h_-^{(1)}(\hat{\beta}_-|\hat{\beta}_1 = c)\) and \(h_-^{(2)}(\hat{\beta}_-|\hat{\beta}_1 > c)\) are defined in eq. 3.10 and 3.12, respectively.

We see that the joint distribution of the estimates \(\hat{\beta}_2, \ldots, \hat{\beta}_K\) may be considered as a weighted sum of the conditional p.d.f's with respect to the subsequent values of the estimate \(\hat{\beta}_1\), with the corresponding probabilities as weights.

4. MEAN AND VARIANCE-COVARIANCE

Having obtained the marginal density function of \(\hat{\beta}_-\), we are able to express mean and covariance matrix of this distribution. By eq. 3.16 we write for the mean:

\[ \mathbb{E}[\hat{\beta}_-) = \alpha \mathbb{E}[\hat{\beta}_-|\hat{\beta}_1 = c] + (1 - \alpha) \mathbb{E}[\hat{\beta}_-|\hat{\beta}_1 > c] \]

\[ = \alpha (\hat{\beta}_- + R^{-1}_{22}r_{21}(\beta_1 - c)) + (1 - \alpha) \mathbb{E}[\hat{\beta}_-|\hat{\beta}_1 > c] \quad (4.1) \]

by virtue of 3.9.

Using 3.12, the second part of the right term becomes:

\[ (1 - \alpha) \mathbb{E}[\hat{\beta}_-|\hat{\beta}_1 > c] = \int_c^\infty p(z_1)\mathbb{E}[\hat{\beta}_-|z_1]dz_1 \]

\[ = \int_c^\infty p(z_1)(\hat{\beta}_- - R^{-1}_{22}r_{21}(z_1 - \beta_1))dz_1 \quad (4.2) \]

\[ = (1 - \alpha)(\hat{\beta}_- + R^{-1}_{22}r_{21}\beta_1) - R^{-1}_{22}r_{21} \int_c^\infty z_1 p(z_1)dz_1 \]
From eq. (3.6) and (3.6.1):
\[ \int_c^\infty z_1 p(z_1)dz_1 = (1 - \alpha)\beta_1 + \sigma_v v_{11} p(c) \]  

(4.3)

Combining eqs. (4.3) and (4.2) with (4.1) yields:
\[ E[\hat{\beta}_-] = \beta_- + R_{22}^{-1}r_{21}\{\alpha(\beta_1 - c) - \sigma_v^2 v_{11} p(c)\} \]

(4.4)

Fortunately we may simplify this expression to:
\[ E[\hat{\beta}_-] = \beta_- + R_{22}^{-1}r_{21} \int_c^\infty (z_1 - c)p(z_1)dz_1 \]

(4.5)

\[ = \beta_- - R_{22}^{-1}r_{21} E[\hat{\beta}_1 - \beta_1] = \beta_- - R_{22}^{-1}r_{21} (\text{bias } \hat{\beta}_1) \]

by virtue of eq. (3.6.1).

As expected, \( \lim_{c \to -\infty} E[\hat{\beta}_-] = \lim_{\alpha \to 0} E[\hat{\beta}_-] = \beta_- \), meaning that the bias will vanish if the restriction is raised.

It is more complicated to derive an expression of the covariance matrix related to the estimates \( \hat{\beta}_2, \ldots, \hat{\beta}_K \). We have seen from eq. 3.16 that the resulting distribution of \( \hat{\beta}_- \) could be written as a weighted sum of two conditional density functions. Because both conditional densities are known we would like to express the covariance matrix in terms of the covariance matrices related to the conditional estimates \( \hat{\beta}_- \); for that purpose we write:

\[ E[\hat{\beta}_- - E[\hat{\beta}_-]][\hat{\beta}_- - E[\hat{\beta}_-]]' = E[\hat{\beta}_-_1 \hat{\beta}_-_1] - E[\hat{\beta}_-]E[\hat{\beta}_-]' \]

\[ = \alpha E_{11} + (1 - \alpha)E_{22} - \{\alpha^2 E_{11}' + (1 - \alpha)^2 E_{22}' + \\
+ \alpha(1 - \alpha)E_{11}' + \alpha(1 - \alpha)E_{22}'\} \]

(4.6)

\[ = \alpha\{E_{11} - E_{11}'\} + (1 - \alpha)\{E_{22} - E_{22}'\} + \\
+ \alpha(1 - \alpha)(E_{1} - E_{2})(E_{1} - E_{2})' \]
with 

\[ E_{11} = E[\hat{\beta}_- \hat{\beta}^-_| \hat{\beta}_1 = c] \]

\[ E_{22} = E[\hat{\beta}_- \hat{\beta}^-_| \hat{\beta}_1 > c] \]

\[ E_1 = E[\hat{\beta}_-| \hat{\beta}_1 = c] \]

\[ E_2 = E[\hat{\beta}_-| \hat{\beta}_1 > c] \]

Equations (3.10), (4.2) and (4.3) yield expressions for 

\( (E_{11} - E_E') \), \( E_1 \) and \( E_2 \) respectively, while a formula 

for \( (E_{22} - E_E'E_2) \) is derived in Appendix A.

As final result we found for the covariance matrix 

of \( \hat{\beta}_- \):

\[ E[\hat{\beta}_- - E[\hat{\beta}_-]][\hat{\beta}_- - E(\hat{\beta}_-)]' = \sigma^2 V_{22} + (\Delta \text{var } \hat{\beta}_1)R_{22}^{-1}r_{21}r_{12}R_{22}^{-1} \]

(4.7)

Since \( \Delta \text{var } \hat{\beta}_1 \) is never positive and the main-diagonal 

elements of \( R_{22}^{-1}r_{21}r_{12}R_{22}^{-1} \) are non-negative, the variances 

of \( \hat{\beta}_2, ..., \hat{\beta}_k \) will always be less than or equal to the 

variances of the corresponding unrestricted least squares 

estimates, which are on the main-diagonal of \( \sigma^2 V_{22} \).

Efficiency of the estimators can be examined in terms 

of the mean squared errors, which correspond with the 

main-diagonal elements of \( E[\hat{\beta}_- - \beta_-][\hat{\beta}_- - \beta_-]' = \text{cov} (\hat{\beta}_-) + 

\text{bias bias}', with \( \text{cov} (\hat{\beta}_-) \) the covariance matrix derived 

by eq. 4.7 and with \( \text{bias} = E[\hat{\beta}_-]_\beta_- \), derived from eq. 4.5.

Combining these results yield:

\[ E[\hat{\beta}_- - \beta_-][\hat{\beta}_- - \beta_-]' = \sigma^2 V_{22} + \Delta \text{var } \hat{\beta}_1R_{22}^{-1}r_{21}r_{12}R_{22}^{-1} + (\text{bias } \hat{\beta}_1R_{22}^{-1}r_{21})'x(\text{bias } \hat{\beta}_1R_{22}^{-1}r_{21})' \]

\[ = \sigma^2 V_{22} + \langle \Delta \text{var } \hat{\beta}_1 + (\text{bias } \hat{\beta}_1)^2 \rangle R_{22}^{-1}r_{21}r_{12}R_{22}^{-1} \]

\[ = \sigma^2 V_{22} + (\text{AMSE } \hat{\beta}_1)R_{22}^{-1}r_{21}r_{12}R_{22}^{-1} \]  

(4.8)
with $\Delta \text{MSE} \hat{\beta}_1$, the difference between the mean-squared error of $\hat{\beta}_1$, obtained by restricted estimation and the corresponding O.L.S. mean-squared error ($= \sigma^2 v_{11}$).

From eq. (3.6) and (3.6.2) we learn:

$$\Delta \text{MSE} \hat{\beta}_1 = -\sigma^2 v_{11} + \int_{-\infty}^{c} (z_1 - c)p(z_1)dz_1 \left\{ \int_{c}^{\infty} (z_1 - c)p(z_1)dz_1 + \int_{-\infty}^{c} (z_1 - c)p(z_1)dz_1 \right\}$$

$$= -\sigma^2 v_{11} + (\beta_1 - c) \int_{-\infty}^{c} (z_1 - c)p(z_1)dz_1$$

$$= -\sigma^2 v_{11} + (\beta_1 - c) \int_{-\infty}^{0} yt(y)dy \leq 0 \text{ if } \beta_1 \geq c.$$

Since the main-diagonal elements of $R^{-1}_{22}r_{21}r_{12}R^{-1}_{22}$ in 4.6 are non-negative, we may conclude that, in case $\beta_1 \geq c$, the respective mean-squared errors of $\hat{\beta}_2$, $\ldots$, $\hat{\beta}_K$ are less than or equal to the corresponding M.S.E.'s of the unrestricted least squares estimates.

5. CONCLUSIONS AND APPLICATIONS

As we have seen our estimates from multiple regression, subject to a linear inequality constraint, are biased but fortunately with less variance compared with unrestricted least squares estimates. Consequently, in some cases it might be favourable to reject an incorrect estimate and apply a quadratic programming procedure to find better estimates, depending on prior knowledge. It must be conceded, however, that this method is allowed only when we are sure of a correct specification of our model.

When two or more parameter estimates are restricted by inequality constraints the previous theory may be generalized; it will, however, become very complicated,
since the number of conditional probability density functions will increase progressively. Somermeyer, Jansen and Louter [7] applied the properties of the two-step estimator to a quadratic programming procedure discussing the standard-errors of parameter estimates in a regression model in which quarterly disposable income figures have been derived from annual data.

APPENDIX A

We want to solve \( E_{22} - E_{2}E_{2}' \), which in our original notation equals:

\[
E[(\hat{\beta}_- - E[\hat{\beta}_- | \beta_1 > c]) | \beta_1 > c][E[\hat{\beta}_- | \beta_1 > c]| \beta_1 > c]'
\]

Equation 3.12 implies in matrix notation:

\[
E_{22} - E_{2}E_{2}' = \frac{1}{1 - \alpha} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (z_- - E[z_- | z_1 > c])
\]

\[
(z_- - E[z_- | z_1 > c])'p(z_1)h_-(z_- | z_1)dz_- dx_1 \quad (A.1)
\]

From the remarks under (3.12) we learn:

\[
E[\hat{\beta}_- | \beta_1] = \beta_- - R_{22}^{-1}r_{21}(\hat{\beta}_1 - \beta_1) \quad \text{and by eqs. (4.2) and (4.3)}:
\]

\[
E[\hat{\beta}_- | \hat{\beta}] = \beta_- - R_{22}^{-1}r_{21} \frac{\sigma^2v_{11}}{1 - \alpha} p(c)
\]

with \( p(c) = \exp \left[ -\frac{(c - \beta_1)^2}{2\sigma^2v_{11}} \right] / \sqrt{2\pi\sigma^2v_{11}} \)

\[
= \beta_- - R_{22}^{-1}r_{21}(\hat{\beta}_1 - \beta_1) + R_{22}^{-1}r_{21}(\hat{\beta}_1 - \beta_1 - \frac{\sigma^2v_{11}p(c)}{1 - \alpha})
\]

\[
= E[\hat{\beta}_- | \beta_1] + R_{22}^{-1}r_{21}(\hat{\beta}_1 - \beta_1 - \frac{\sigma^2v_{11}p(c)}{1 - \alpha}).
\]
Using this result in A.1, we obtain:

\[
\mathbb{E}_{22} - \mathbb{E}_{22}^\prime = \frac{1}{1 - \alpha} \int_c^\infty \int_{-\infty}^\infty \{z - \mathbb{E}[z \mid z_1]\} \, \sigma_{\nu 11}^2 p(c) \{\text{ditto}\} p(z_1) h_-(z \mid z_1) \, dz \, dz_1
\]

\[
= \frac{1}{1 - \alpha} \int_c^\infty \int_{-\infty}^\infty \left( z - \mathbb{E}[z \mid z_1] \right) \left( z - \mathbb{E}[z \mid z_1] \right)^\prime \, dz + \int_{-\infty}^\infty R_{22}^{-1} r_{21} t(z_1) \left( z - \mathbb{E}[z \mid z_1] \right)^\prime \, dz +
\]

\[
+ \int_{-\infty}^\infty \left( z - \mathbb{E}[z \mid z_1] \right) \left( R_{22}^{-1} r_{21} t(z_1) \right)^\prime \, dz + \int_{-\infty}^\infty R_{22}^{-1} r_{21} r_{12} R_{22}^{-1} t(z_1)^2 \cdot h_-(z \mid z_1) \, dz_1
\]

with \( t(z_1) = z_1 - \beta_1 - \frac{\sigma_{\nu 11}^2}{1 - \alpha} p(c) \)

\[
= \frac{1}{1 - \alpha} \int_c^\infty p(z_1) \left\{ \text{cov}(z \mid z_1) + R_{22}^{-1} r_{21} r_{12} \cdot t(z_1)^2 \right\} \, dz_1
\]

\[
= \frac{1}{1 - \alpha} \int_c^\infty p(z_1) \text{cov}(z \mid z_1) \, dz_1 + \frac{1}{1 - \alpha} \int_c^\infty R_{22}^{-1} r_{21} r_{12} R_{22}^{-1} t(z_1)^2 p(z_1) \, dz_1
\]

(A.2)
We solve the integral in this expression, integrating the subsequent parts.

\[
\int_c (z_1 - \beta_1)^2 (2\pi\sigma^2 v_{11})^{-\frac{1}{2}} \exp \left\{-\frac{(z_1 - \beta_1)^2}{2\sigma^2 v_{11}}\right\} \, dz_1 =
\]

\[
\frac{1}{\sqrt{2\pi\sigma^2 v_{11}}} \int_{-\infty}^{\infty} w^2 \exp \left(-\frac{w}{\sigma\sqrt{v_{11}}}\right)^2 \, dw
\]

\[
= \frac{1}{\sqrt{2\pi\sigma^2 v_{11}}} \left\{[-w\sigma^2 v_{11} \exp \left(-\frac{w}{\sigma\sqrt{v_{11}}}\right)^2]_{-\infty}^{\infty} + \int_{c-\beta_1}^{c-\beta_1} \sigma^2 v_{11} \, \exp \left(-\frac{w}{\sigma\sqrt{v_{11}}}\right)^2 \, dw \right\}
\]

\[
= \frac{\sigma\sqrt{v_{11}}}{\sqrt{2\pi}} (c - \beta_1) \exp -\frac{1}{2\sigma\sqrt{v_{11}}} + (1 - \alpha)\sigma^2 v_{11}
\]

\[
\int_c (z_1 - \beta_1)(2\pi\sigma^2 v_{11})^{-\frac{1}{2}} \exp -\frac{(z_1 - \beta_1)^2}{2\sigma^2 v_{11}} =
\]

\[
= \frac{1}{\sqrt{2\pi\sigma^2 v_{11}}} \int_{-\infty}^{\infty} w \exp -\frac{w}{\sigma\sqrt{v_{11}}}^2 \, dw
\]

\[
= \frac{1}{\sqrt{2\pi\sigma^2 v_{11}}} \left[-\sigma^2 v_{11} \exp -\frac{w}{\sigma\sqrt{v_{11}}}^2 \right]_{c-\beta_1}^{\infty} = \sigma^2 v_{11} p(c).
\]
\[ \int _{c}^{\infty } \frac{(\sigma \sqrt{v_{11}})^4}{(1 - \alpha)^2} p(c)^2 p(z_1) dz_1 = \frac{(\sigma \sqrt{v_{11}})^4}{1 - \alpha} p(c)^2 \quad (A.5) \]

Combining (A.3), (A.4), (A.5) and (A.2) yield:

\[ E_{22} - E_{2}E_{2}^{'} = \sigma^{2}R_{22}^{-1} + R_{22}^{-1}R_{21}r_{12}R_{22}^{-1} \sigma^{2}v_{11}(1 - \alpha + (c - \beta_{1})p(c) \]

\[ - \frac{\sigma^{2}v_{11}p(c)^2}{1 - \alpha}(1 - \alpha)^{-1}. \]

We already found:

\[ (E_{11} - E_{1}E_{1}^{'}) = \sigma^{2}R_{22}^{-1} \text{ by (3.10)} \]

\[ E_{1} = \beta_{-} + R_{22}^{-1}R_{21}(\beta_{1} - c) \text{ by (3.9)} \]

\[ E_{2} = \beta_{-} - R_{22}^{-1}R_{21} \frac{\sigma^{2}v_{11}p(c)}{1 - \alpha} \text{ by (4.2) and (4.3)}. \]

Substituting these results in eq. (4.6) we obtain:

\[ E[\beta_{-} - E[\beta_{-}][\beta_{-} - E[\beta_{-}]]]^{'} = \alpha \sigma^{2}R_{22}^{-1} + (1 - \alpha) \left\{ \sigma^{2}R_{22}^{-1} + \right. \]

\[ \left. \frac{R_{22}^{-1}R_{21}r_{12}R_{22}^{-1}}{1 - \alpha} \sigma^{2}v_{11}(1 - \alpha + \right. \]

\[ \left. + (c - \beta_{1})p(c) - \frac{\sigma^{2}v_{11}p(c)^2}{1 - \alpha} \right\} + \alpha(1 - \alpha). \]

\[ \left. \frac{R_{22}^{-1}R_{21}(\beta_{1} - c + \frac{\sigma^{2}v_{11}p(c)}{1 - \alpha})}{R_{22}^{-1}R_{21}(\beta_{1} - c + \frac{\sigma^{2}v_{11}p(c)}{1 - \alpha})} \right\}. \]

\[ \left. \frac{(R_{22}^{-1}R_{21}(\beta_{1} - c + \frac{\sigma^{2}v_{11}p(c)}{1 - \alpha}))}{R_{22}^{-1}R_{21}(\beta_{1} - c + \frac{\sigma^{2}v_{11}p(c)}{1 - \alpha})} \right\}. \]

\[ = \sigma^{2}R_{22}^{-1} + R_{22}^{-1}r_{21}r_{12}R_{22}^{-1} \sigma^{2}v_{11}(1 - \alpha + (c - \beta_{1})p(c) \]

\[ - \frac{\sigma^{2}v_{11}p(c)^2}{1 - \alpha} \right\} + \]
\[ + R_{22}^{-1}r_{21}r_{12}R_{22}^{-1}(\alpha - 1)(\beta_1 - c)^2 + \frac{\alpha}{1 - \alpha} (\sigma \sqrt{v_{11}})^4 \rho(c)^2 + 2\alpha(\beta_1 - c)\sigma v_{11}^2 \rho(c) \]

\[ = \sigma^2 R_{22}^{-1} + R_{22}^{-1}r_{21}r_{12}R_{22}^{-1}v_{11}^{-1}(1 - \alpha + (1 - 2\alpha)(c - \beta_1)\rho(c) - \sigma^2 v_{11}^2 \rho(c)^2 + \alpha(1 - \alpha)\frac{(\beta_1 - c)^2}{\sigma \sqrt{v_{11}}} \]

\[ = \sigma^2 R_{22}^{-1} + \sigma v_{11}^2 R_{22}^{-1}r_{21}r_{12}R_{22}^{-1} + \Delta \text{var} \hat{\beta}_1 R_{22}^{-1}r_{21}r_{12}R_{22}^{-1} \]

with \( \Delta \text{var} \hat{\beta}_1 \) the difference between the variance of the restricted estimate and the variance of the unrestricted estimate of \( \beta_1 \) (eq. (3.6.2)).

\[ = \sigma^2 v_{22} + \Delta \text{var} \hat{\beta}_1 R_{22}^{-1}r_{21}r_{12}R_{22}^{-1} \text{ by eq. (3.3d) and (3.3e).} \]

**FOOTNOTE**

* The notation \( h(y) = n(\mu, \Omega) \) means that the probability density function of \( y \) equals the normal density with mean \( \mu \) and var-covar. matrix \( \Omega \).

**REFERENCES**


