INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or “target” for pages apparently lacking from the document photographed is “Missing Page(s)”. If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.

2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of “sectioning” the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.
Harter, Rachel Margaret

SMALL AREA ESTIMATION USING NESTED-ERROR MODELS AND
AUXILIARY DATA

Iowa State University

Ph.D. 1983

University
Microfilms
International 300 N. Zeeb Road, Ann Arbor, MI 48106
Small area estimation using nested-error models
and auxiliary data

by

Rachel Margaret Harter

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work
Signature was redacted for privacy.

For the Major Department
Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa
1983
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>I. INTRODUCTION</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>II. OVERVIEW OF SMALL AREA METHODOLOGY</strong></td>
<td>3</td>
</tr>
<tr>
<td>A. Early Methods</td>
<td>3</td>
</tr>
<tr>
<td>B. Symptomatic Accounting Techniques</td>
<td>4</td>
</tr>
<tr>
<td>C. Regression Methods</td>
<td>8</td>
</tr>
<tr>
<td>D. Categorical Data Analysis Approaches</td>
<td>14</td>
</tr>
<tr>
<td>E. Prediction Approaches</td>
<td>18</td>
</tr>
<tr>
<td>F. James-Stein Estimators</td>
<td>19</td>
</tr>
<tr>
<td><strong>III. PREVIOUS WORK IN COUNTY CROP AREA ESTIMATION</strong></td>
<td>22</td>
</tr>
<tr>
<td>A. Prediction of Crop Areas</td>
<td>22</td>
</tr>
<tr>
<td>B. The Battese-Fuller Predictor</td>
<td>25</td>
</tr>
<tr>
<td>1. The nested-error model and assumptions</td>
<td>25</td>
</tr>
<tr>
<td>2. Predictors when $\beta$ is known</td>
<td>27</td>
</tr>
<tr>
<td>3. Predictors when $\beta$ is unknown</td>
<td>30</td>
</tr>
<tr>
<td>4. Predictions for nonsampled areas</td>
<td>32</td>
</tr>
<tr>
<td>C. An Extension for Stratification</td>
<td>33</td>
</tr>
<tr>
<td>1. The model and assumptions</td>
<td>33</td>
</tr>
<tr>
<td>2. Predictors when $\beta$ is known</td>
<td>36</td>
</tr>
<tr>
<td>3. Predictors when $\beta$ is unknown</td>
<td>40</td>
</tr>
<tr>
<td><strong>IV. ESTIMATION OF VARIANCE COMPONENTS</strong></td>
<td>50</td>
</tr>
<tr>
<td>A. The Fitting-of-Constants Estimators</td>
<td>50</td>
</tr>
<tr>
<td>B. Effects of Estimating the Variance Components</td>
<td>54</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>-----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>1. Effects on the estimation of $\beta$</td>
<td>54</td>
</tr>
<tr>
<td>2. Effects on the Battese-Fuller predictor</td>
<td>55</td>
</tr>
<tr>
<td>C. Estimation with Prior Estimates</td>
<td>90</td>
</tr>
<tr>
<td>1. General theory</td>
<td>90</td>
</tr>
<tr>
<td>2. Prior estimates of $\sigma^2_v$ and $\sigma^2_e$</td>
<td>93</td>
</tr>
<tr>
<td>3. Prior estimate of $\sigma^2_v$</td>
<td>96</td>
</tr>
<tr>
<td>4. Prior estimate of $\sigma^2_v/\sigma^2_e$</td>
<td>97</td>
</tr>
<tr>
<td>V. Prediction under the Heterogeneous Nested-Error Model</td>
<td>100</td>
</tr>
<tr>
<td>A. The Model and Predictor</td>
<td>100</td>
</tr>
<tr>
<td>B. Estimation of Variance Components</td>
<td>105</td>
</tr>
<tr>
<td>1. The variance component estimators</td>
<td>106</td>
</tr>
<tr>
<td>2. Effects on the estimation of $\beta$</td>
<td>117</td>
</tr>
<tr>
<td>3. Effects on the predictor</td>
<td>178</td>
</tr>
<tr>
<td>4. Estimation with prior estimates</td>
<td>191</td>
</tr>
<tr>
<td>C. Tests of the Model</td>
<td>193</td>
</tr>
<tr>
<td>VI. Numerical Examples</td>
<td>196</td>
</tr>
<tr>
<td>A. The Homogeneous Nested-Error Model</td>
<td>196</td>
</tr>
<tr>
<td>B. The Heterogeneous Nested-Error Model</td>
<td>211</td>
</tr>
<tr>
<td>VII. Summary</td>
<td>219</td>
</tr>
<tr>
<td>VIII. References</td>
<td>224</td>
</tr>
<tr>
<td>IX. Acknowledgements</td>
<td>231</td>
</tr>
<tr>
<td>X. Appendix: An Alternative Estimator of $\gamma_1$ for the Homogeneous Nested-Error Model</td>
<td>232</td>
</tr>
</tbody>
</table>
I. INTRODUCTION

Often when a sample survey is taken of a large population, the sample units are widely scattered over the entire domain. In addition to the estimation of parameters for the whole population, there may be interest in making inferences about subdomains of the population. However, data on the individual subdomains may be meager or nonexistent. The problem of making estimates for sparcely sampled subdomains from large survey data is often called "small area estimation."

The problem of small area estimation is not new. For decades, the U.S Bureau of the Census and similar agencies in other countries have made small area population estimates for intercensal years. Interest in small area estimation has increased as government agencies have come to rely more and more on small area estimates. Federal, state and local administrators require estimates for smaller geographic or demographic areas, but census data for these areas are available only at the time of a decennial census. Between census periods, estimates for the smaller areas are often based on national survey data. Increased federal funding of programs at the state and local levels has spurred the development of efficient small area methodology. For example, under the General Revenue Sharing Program, the U.S. Treasury Department allocates monies to state and local governments based on population, per capita income, and taxation data for the areas involved. The Census Bureau provides current estimates of per capita income for state and local
areas. The sampling error of unbiased survey estimators for areas with a small population is relatively large, so alternative methods are appropriate. Many small area estimation procedures have been suggested over the years. A historical overview of the major methods and the problems to which they have been applied is given in Chapter II.

The particular small area estimation problem of interest in this work is the estimation of county crop areas using a national survey by the U.S. Department of Agriculture and auxiliary data from LANDSAT satellites. Battese and Fuller (1981, 1982) proposed a prediction procedure for county crop area estimation based on a nested-error model. The crop estimation problem, the nested-error model, and the Battese-Fuller procedure are discussed in Chapter III. Variations of the prediction procedure for nonsampled areas and for a stratified design are also considered.

In practice, the variance components of the nested-error model usually must be estimated. In Chapter IV, estimators of the variance components are suggested, and the theoretical implications of using estimated variance components in the prediction procedure are explored.

A nested-error model with heterogeneous error variances is considered in Chapter V. A small area prediction procedure analogous to the Battese-Fuller procedure is presented. Estimation of the variance components of this model is also discussed.

Finally, numerical examples are given in Chapter VI, illustrating the prediction procedures for both nested-error models.
II. OVERVIEW OF SMALL AREA METHODOLOGY

Over the years, a variety of methods have been proposed for estimating characteristics of small subpopulations. Some methods take advantage of symptomatic auxiliary data such as births, deaths, school enrollments, and automobile registrations. It is assumed that the auxiliary variables are highly correlated with the dependent variable. Other estimators use data from neighboring small areas or assume a model structure for the population. Since each method was developed for a specific purpose, the choice of a method for a particular problem depends on the available data and the assumptions one is willing to make. A summary of most small area techniques is given in this chapter. Purcell and Kish (1979, 1980) have also provided an overview of small area methodology.

A. Early Methods

One of the first problems requiring small area techniques was that of estimating intercensal or postcensal populations of cities and counties. Some of the earliest methods used were not very statistical in nature. Snow (1911) reported that around the turn of the century the American government made population estimates for small areas by simple linear extrapolation based on the two previous censuses. About the same time, the Registrar-General of England made postcensal estimates using arithmetic and geometric progressions. The population of each city, town, or other small region was estimated by arithmetic progression,
then all small area estimates were multiplied by a factor so that they summed to the geometric progression estimate for the whole country.

B. Symptomatic Accounting Techniques

The problem of estimating intercensal populations of local areas has always been one of the major problems in small area estimation. Not surprisingly, the U.S. Bureau of the Census is responsible for several of the small area techniques that have been developed in the last 40 years. A recent study by the Panel on Small-Area Estimates of Population and Income (1980) examined the small area methods currently in use by the Bureau of the Census. Most of the Census Bureau's methods are included in this section.

In the 1940s the U.S. Bureau of the Census developed Census Component Methods I and II for postcensal estimates of local populations [U.S. Bureau of the Census (1949, 1966, 1969)]. Essentially, the estimate of the total population in an area at a particular postcensal date is equal to:

Civilian population in the area at the last census
+ natural increase during the intervening time (births - deaths)
+ net civilian migration during the intervening time
- those who left the area for military service
+ military personnel stationed in the area at the postcensal date of interest.

The main difference between Component Methods I and II is in the way net civilian migration is estimated from school enrollment data. Since each
term in the estimator is derived from actual counts, the success of these estimates depends largely on the quality of records kept by local authorities. Such methods that rely on elaborate bookkeeping have been labeled "symptomatic accounting techniques" by Purcell and Kish (1979, 1980). Ericksen (1973a) also described the major symptomatic accounting techniques.

More recently Starsinic (1974) tried a method of estimating postcensal population for revenue sharing purposes. The method is very similar to Component Methods I and II except that net migration is estimated from IRS records of residence.

Another symptomatic accounting technique presented by Bogue (1950) has come to be known as the vital rates method. This method uses birth and death records of small subareas and larger geographic regions called parent areas. The steps are as follows:

1. Compute crude birth and death rates for the subareas and the parent areas based on the last census.

2. Compute \( \frac{\text{subarea birth rate}}{\text{parent area birth rate}} \) and \( \frac{\text{subarea death rate}}{\text{parent area death rate}} \) for the census year. These ratios may need to be adjusted if subarea rates have changed over time relative to the parent area.

3. Compute crude birth and death rates of parent areas for the year of interest.

4. Multiply the ratios in step 2 by the rates in step 3 to get estimates of the subarea rates for the year of interest.
5. Divide the estimated birth (and death) rates in step 4 by the actual number of births (deaths) in the subarea.

6. Average the two resulting estimates of the subarea population.

7. Adjust all subarea estimates so that they sum to the parent area population.

Bogue did not claim that this method gives more accurate estimates for each subarea than the more meticulous census component methods; rather, Bogue's method is useful to determine demographic trends during intercensal periods.

The Housing Unit method discussed by Rives (1976) and the U.S. Bureau of the Census (1969) uses housing inventories, vacancy rates, and other similar symptomatic variables to estimate populations. Let

\[ H = \text{number of housing units in the area} \]

\[ = H_c + H_a - H_d, \]

where

\[ H_c = \text{number of housing units in the area at the time of the most recent census}, \]

\[ H_a = \text{number of houses added during the postcensal period}, \]
Let

\[ b = \text{average household size at the last census}, \]

\[ w = \text{vacancy rate at the last census}, \]

\[ g = \text{proportion of the area population living in group housing at the last census}. \]

Then the population \( P \) is estimated by

\[ P = \frac{b H(1 - w)}{(1 - g)}. \]

Rives also suggested a modified version in which \( H \) is kept in a current housing unit file, and \( b \) and \( w \) are estimated from a current sample of the housing units in the file.

The last symptomatic accounting technique presented here is the composite method developed by Bogue and Duncan (1959). This method estimates postcensal populations of local areas by age, sex, and race. The symptomatic variables used are births, deaths, and school enrollments. In this procedure, estimates for each age group by sex and race are computed by whatever established method is deemed best for that age group. The total population or the population of one sex or race in a local area is then estimated by adding the corresponding age group estimates. Composite method estimates have been used by the Bureau of
the Census (1969), sometimes averaged with estimates computed by other methods.

C. Regression Methods

Historically, the next method of estimating postcensal or intercensal populations to come into vogue was the ratio-correlation method which relies on multiple regression theory. The use of multiple regression for small area population estimates has a relatively long history, dating back to the work of Snow (1911). Snow's formulation is very general.

Let \( x_0 \) be the population of a district (or the increase in the population since the previous census). The population \( x_0 \) is known for a census year, and estimates are desired for a postcensal year. Let \( x_1, x_2, \ldots, x_n \) be functions of symptomatic variables that are highly correlated with \( x_0 \) as indicated by previous censuses. These variables are known for the census year and the postcensal year of interest.

Multiple regression on the census data gives an estimated relationship between \( x_0 \) and the symptomatic functions. Assuming this relationship to remain constant over time, the postcensal population estimates are obtained by substituting the postcensal symptomatic functions into the regression equations. Snow pointed out that this procedure does not attempt to give an exact value of \( x_0 \), but it gives the most probable value of \( x_0 \) based on past experience.

The ratio-correlation method as proposed by Schmitt and Crosetti (1954) is illustrated by the following example. To estimate the 1983
populations of counties, the first step is to use multiple regression to estimate the following county equations:

\[ x_{1.234} = \alpha_0 + \alpha_1 x_2 + \alpha_2 x_3 + \alpha_3 x_4, \]

where

- \( x_{1.234} \) = change in the county's share of statewide population from 1970 to 1980,
- \( x_2 = \frac{1980 \text{ county/state ratio of live births}}{1970 \text{ county/state ratio of live births}} \),
- \( x_3 = \frac{1980 \text{ county/state ratio of registered vehicles}}{1970 \text{ county/state ratio of registered vehicles}} \),
- \( x_4 = \frac{1980 \text{ county/state ratio of public school enrollment}}{1970 \text{ county/state ratio of public school enrollment}} \).

Then to get the actual estimates, the values of the ratio of 1983 to 1980 for the independent variables are substituted into the regression equations.

One advantage of this method is that it is very flexible with respect to the number and choice of auxiliary variables. Studies by
Schmitt and CroSETti (1954), CroSETti and Schmitt (1956), and Goldberg, Rao, and Namboodiri (1964) showed the ratio-correlation method to yield more accurate results for most counties than other methods in use at the time, including linear extrapolation, the vital rates method, Census Component Method II, and the composite method. The major assumption of the ratio correlation method is that the regression relationships of the census years also hold for the estimation year. Namboodiri (1972) pointed out that since this assumption is not always valid, this method is not best in all situations. In fact, the ratio-correlation method performed rather poorly in some counties with unusual characteristics.

Rosenberg (1968) improved ratio-correlation estimates of Ohio county populations by stratifying the counties by growth rates and by percentages of urban or agricultural areas. Pursell (1970) successfully introduced dummy variables for other economic and demographic characteristics in his application to West Virginia counties. Martin and Serow (1978) extended the ratio-correlation method and these variations to estimate populations in small areas by age and race, but their results were inconclusive.

O'Hare (1976) suggested the use of differences instead of ratios. Thus, in the above example,

\[ x_2 = (1980 \text{ county/state ratio of live births}) - (1970 \text{ county/state ratio of live births}) \]
with the other variables defined analogously. Swanson (1978) supported the difference method for two reasons. First, if either the numerator or denominator of the ratio is zero, the whole ratio is set to zero or is treated as a missing observation, and some information is lost. Second, differences are said to be more sensitive than ratios, resulting in a better fit of the model.

The regression method of Morrison and Relles (1975) is similar to the ratio-correlation method, but it uses a log-linear model to estimate populations of districts within cities. After estimating the parameters using data from the two previous censuses, the estimated change in population for district \( i \) one year after the census is computed from the equation

\[
\Delta \log P_t(i) = \alpha_t + \beta[\Delta \log B_t(i)] + \gamma[\Delta \log D_t(i)] + \epsilon_t,
\]

where

\( \Delta \) = one year's change,

\( P_t(i) \) = population at year \( t \),

\( B_t(i) \) = number of births during year \( t \),

\( D_t(i) \) = number of deaths during year \( t \),
\( \alpha_t = \text{indicator of city-wide change in population during year } t \),

\( \epsilon_t = \text{random error} \),

\( t = \text{year of interest} \).

An estimate for the same district \( k \) years after the census is determined by the equation

\[
\Delta^k \log P_t(i) = [\alpha_t + \alpha_{t-1} + \ldots + \alpha_{t-k+1}] \\
+ \beta [\Delta^k \log B_t(i)] + \gamma [\Delta^k \log D_t(i)] \\
+ [\epsilon_t + \epsilon_{t-1} + \ldots + \epsilon_{t-k+1}].
\]

District estimates are then adjusted to sum to the estimate for the city total. For the \( k \)-th intercensal year, one could use a weighted average of the \( k \)-th forward estimate from the preceding census with the \((10-k)\)-th backward estimate from the following census. Morrison and Relles claimed that this model better fits the assumptions of least squares. They recommended its use for making annual population estimates to determine demographic trends.

Other regression methods were inspired by Hansen, Hurwitz, and Madow's (1953) Radio Listening Survey. From a two-stage stratified
sample, interviews provided estimates for some of the PSUs (small areas). A volunteer mail survey provided auxiliary data for all PSUs. The estimated regression line of the sample data on the corresponding auxiliary data was used to estimate nonsampled PSUs from the auxiliary data.

Woodruff (1966) estimated monthly county retail trade by an extension of the Radio Listening Survey regression procedure. Let $y_i$ be the estimate of the monthly retail sales for the $i$-th sampled county, and $x_i$ be the annual retail sales figure for the county from the last census. After estimating the regression equations $y_i = \beta x_i$, an estimate for a nonsampled county is $\hat{y}_p = \beta \hat{x}_p$, where $\hat{\beta} = \sigma_{xx}^{-1} \sigma_{xy}$ and $x_p$ corresponds to the entire region. Instead of $y_i$, an estimate of the $i$-th sampled county is $y_i + \hat{\beta}(x_p - x_i)$ which adds to $y_i$ a correction term. This estimator is the survey regression estimator of Cochran (1977.)

A multivariate extension of the Radio Listening Survey method is the regression-sample data technique proposed by Ericksen (1973a, 1973b, 1974). This method takes advantage of auxiliary data without relying on past structural relationships.

To compute estimates for the ratio-correlation illustration by this method, a sample of $n$ counties is taken. Estimates of the populations of these counties are obtained, often by two-stage sampling. Auxiliary data are available for all the counties of interest. Using the sample counties only, the following model is estimated by multiple regression
\[
\begin{pmatrix}
  y_1 \\
  \vdots \\
  y_n
\end{pmatrix} = \begin{pmatrix}
  1 & x_{11} & x_{21} & \cdots & x_{p1} \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  1 & x_{1n} & x_{2n} & \cdots & x_{pn}
\end{pmatrix} \begin{pmatrix}
  \alpha_0 \\
  \alpha_1 \\
  \vdots \\
  \alpha_p
\end{pmatrix} + \begin{pmatrix}
  u_1 \\
  \vdots \\
  u_n
\end{pmatrix},
\]

where

\[
y_j = \frac{\text{estimated 1983 population}}{\text{1980 population}} \text{ for the } j\text{-th county},
\]

\[
x_{ij} = \frac{\text{1983 value of } i\text{-th symptomatic variable}}{\text{1980 value of } i\text{-th symptomatic variable}} \text{ for the } j\text{-th county}.
\]

The symptomatic ratios for all the counties are then substituted into the estimated equation to get the county population estimates.

Like ratio-correlation, this method is very flexible with respect to the number and kind of symptomatic variables. A significant portion of the error of Ericksen's estimates is due to sampling error. On the other hand, much of the error of ratio-correlation estimates is due to changes in structural relationships over time. In general, one cannot say which method is better.

D. Categorical Data Analysis Approaches

The problem of estimating populations of small areas by age, sex, and race lends itself to categorical approaches. The data are arranged
in a multi-dimensional table where the rows of the $i$-th dimension represent the categories of the $i$-th attribute. Estimates from a particular cell "borrow" information from neighboring cells. The resulting estimates often have a log-linear relationship with the main effects and interactions of the different attributes.

Bousfield (1977), Chambers and Feeney (1977), and Freeman and Koch (1976) have all worked with the algorithm of iterative proportional fitting (IPF) developed by Deming and Stephan (1940). For simplicity, assume the data are categorized by two attributes, age and sex. The goal is to estimate the number of males and females in each age group in the small area. Data about the small area from the most recent census are arranged in a two-way table where the rows represent age groups and the columns represent the two sexes. Also available are estimates of age group and sex totals based on current survey data. Row and column totals are often called the marginal totals. Each row of the census data is multiplied by a constant so that the entries in a row add up to the survey estimated row total. This process is called raking. The columns can also be raked, but then the row raking will be disturbed. The rows and columns can be raked alternately until the table converges. The resulting table contains the desired age-by-sex population estimates. This is the IPF procedure. The procedure generalizes to higher dimensions.

The main feature of categorical approaches is that the observations in a cell are assumed to be correlated with neighboring cells. The
relationship between the variable of interest, such as population, and the categorical variables within the small area is called the association structure. The relationship between the variable of interest in the small area and the categorical variables in the larger parent area is called the allocation structure. In many cases, the association structure is determined at the time of the census. The allocation structure is observable at a postcensal date only through the marginal totals, or sample data. The objective of categorical data analysis procedures is to modify the original association structure as little as possible while adjusting the data to fit the current allocation structure. The particular estimator depends on the information in the association and allocation structures. The IPF estimator is just one member in this class named "structure preserving estimators" (SPREE) by Purcell and Kish (1980). A thorough discussion of SPREE estimators can be found in their paper. In general, SPREE estimators tend to be biased, and the severity of the bias depends on the accuracy of the assumed association and allocation structures.

The best-known member of the SPREE family is the synthetic estimator. According to Purcell and Kish (1980), synthetic estimates are based on incomplete association structures, resulting in larger biases. In their barest form, they do not adequately account for local factors. Nevertheless, synthetic estimates have been used extensively during the past 15 years. The National Center for Health Statistics (1977a) used synthetic estimates with its National Health Survey to make

In synthetic estimation, the parent area population is divided into $G$ subgroups according to the associated categorical variables. The small subareas for which estimates are desired can be thought of as being crossed with the $G$ subgroups. Let $U_{j}$ be an unbiased estimate for the $j$-th subgroup (marginal data). For the $i$-th subarea, assign $G$ weights $P_{ij}$ to each subgroup such that $\sum_{j=1}^{G} P_{ij} = 1$. The synthetic estimate is $U_{i} = \sum_{j=1}^{G} P_{ij} U_{j}$. Further discussion and evaluation of synthetic estimates can be found in Gonzalez (1973), Laske (1979), Levy (1971), Namekata, Levy and O'Rourke (1975), the National Center for Health Statistics (1977b), and Schaible, Brock, and Schnack (1977).

Cohen and Kalsbeek (1977) proposed a "base unit" method which is synthetic estimation with geographic subgroups. Counties and other small areas for which estimates are desired are called target areas. Each target area is divided into base units which in turn are sampled and grouped into strata according to symptomatic information. All base units from the target areas are assigned to the strata according to symptomatic data. Multistage sampling is used to obtain strata estimates. Estimates for the target area base units are made from the
strata to which they were assigned. These estimates are then pooled to give an estimate of the target area. This method is not necessarily very accurate, but it has the advantages of being quick and cheap.

Some improvements have been suggested for synthetic estimates. Levy (1971) assumed that the percentage difference between a synthetic estimate and the true value is a linear function of an available symptomatic variable. Estimates of the linear coefficients can be found by regression. Then, a correction factor can be computed for the synthetic estimate.

Gonzalez and Hoza (1975, 1978) combined synthetic estimates with Ericksen's regression sample data method by including the synthetic estimates as an independent variable in the regression. Gonzalez and Hoza (1978) also studied the effect of excluding outliers from the estimation procedure.

E. Prediction Approaches

The prediction approach to small area estimation assumes a super-population model holds for the variable of interest and the auxiliary variables. Royall (1970) made the following assumptions.

Let $y_1, \ldots, y_N$ be realizations of the independent random variables $Y_1, \ldots, Y_N$. Let $x_1, \ldots, x_N$ be the corresponding fixed auxiliary data. $Y_i$ has mean $\beta x_i$ and variance $\sigma^2 v(x_i)$, where $v(.)$ is a known positive function, and $\beta$ and $\sigma^2$ are unknown. An estimate of $T = \sum_{i=1}^{N} y_i$ is desired, and a sample $y_1, \ldots, y_n$ is observed. Royall showed that any estimator $\hat{T}$ for $T$ can be expressed
uniquely as a linear combination of the sample $y_1$'s and a predictor for the nonsample $y_1$'s. That is,

$$\hat{T} = \sum_{i=1}^{n} \hat{y}_i + \hat{\beta} \sum_{i=n+1}^{N} x_i,$$

where the best choice of $\hat{\beta}$ is

$$\hat{\beta} = \frac{\sum_{i=1}^{n} \frac{x_i y_i}{v(x_i)}}{\sum_{i=1}^{n} \frac{x_i^2}{v(x_i)}}.$$

An advantage of prediction estimators is that their mean squared errors can be computed under the method. Actually, under the model, the MSE of other estimators can be computed and compared with the prediction estimator. Laake (1979) did this with the synthetic estimator. Holt, Smith, and Tomberlin (1979) followed Royall's work by examining some particular models and their estimators. Battese and Fuller (1981) applied the prediction approach to the estimation of crop areas, and the resulting estimator is a James-Stein estimator in nature. Battese and Fuller's work is presented in greater detail in the next chapter.

F. James-Stein Estimators

Stein (1956) showed that when at least three means are being estimated simultaneously, the conditionally unbiased maximum likelihood estimator is not admissible under the sum of mean squared errors loss.
function. For this reason, James-Stein theory [James and Stein (1961)] is applicable when means or totals of several small areas are being estimated simultaneously. James-Stein estimators are shrinkage estimators; that is, the small area estimates are shrunken toward the grand mean.

As a simple case, consider the model

\[ x_i | \mu_i \sim \text{NID}(\mu_i, 1), \quad i = 1, \ldots, k \]

and

\[ \mu_i \sim \text{NID}(0, \sigma^2), \quad i = 1, \ldots, k, \]

where \( x_i \) might be a sample mean. The conditionally unbiased maximum likelihood estimator of \( \mu_i \) is \( x_i \). The James-Stein estimator of \( \mu_i \) is

\[ k \left[ 1 - \left( \sum_{i=1}^{k} x_i^2 \right)^{-1}(k - 2) \right] x_i. \]

The shrinkage term \( 1 - \left( \sum_{i=1}^{k} x_i^2 \right)^{-1}(k - 2) \) is less than 1, and it tends to shrink the value of \( x_i \) back toward 0. In practice, if the shrinkage term is less than zero, it is replaced by zero.

Zellner and Vandeale (1974) reviewed James-Stein and similar estimators with Bayesian interpretations. Fay and Herriot (1979)
applied a modified James-Stein estimator to estimate per capita income in areas of small populations. Efron and Morris (1973) derived the James-Stein estimator from an empirical Bayes point of view. The Efron and Morris study provides the foundation for some of the work in later chapters.
III. PREVIOUS WORK IN COUNTY CROP AREA ESTIMATION

Of the small area procedures developed thus far, no single method has emerged as the best method for all small area estimation problems. The procedures were developed for different purposes with different types of data available and different assumptions in mind. Thus, the choice of a small area procedure depends on the particular problem. The Battese-Fuller predictor and its variations were developed for the specific purpose of estimating mean crop acreages for counties and other small areas. For this reason, the presentation of the Battese-Fuller predictor will be preceded by a description of the crop estimation project.

A. Prediction of Crop Areas

Each year the Statistical Reporting Service of the U.S. Department of Agriculture estimates crop areas at the national and state levels using data from the June Enumerative Survey. In each state the Statistical Reporting Service prepares the county estimates based on data from the June Enumerative Survey, mail surveys, the state census of agriculture, and other sources. The Statistical Reporting Service is attempting to improve the results of these estimates by using additional data supplied by LANDSAT satellites. Hanuschak, Allen and Wigton (1982) outlined the history of the crop estimation program with LANDSAT data. Cárdenas, Blanchard and Craig (1978) and Hanuschak, et al. (1979)
described the estimation procedure and gave details about the June Enumerative Survey and the LANDSAT satellites.

The June Enumerative Survey is an annual nationwide agricultural survey conducted in late May and early June by the Statistical Reporting Service. The sample units, or segments, typically consist of about one square mile (or 259 hectares) of land. The segments are stratified at two levels - first by state, and second by land use based on percentage of cultivated land. Approximately 16,000 segments (0.5 percent of the total U.S. land area) are sampled by stratified random sampling. Individual fields within each segment are located by aerial photography, and interviewers record field size and crop or land use. These ground data are the basis for the national and state crop estimates.

LANDSAT is the name of the Earth Resources Technology Satellites which monitor the earth from sun-synchronous orbits at an altitude of 570 miles. Each satellite is equipped with a multispectral scanner that measures the amount of radiant energy emitted or reflected from the earth's surface. A single resolution of the scanner is called a pixel (an abbreviation for picture element), and it encompasses an area of about one acre (or 0.45 hectares). For each pixel, the scanner measures radiation in four wavelength bands of the electromagnetic spectrum. Using discriminant analysis on these four variables, the crop cover in each pixel is classified. The total number of pixels classified as having a given crop cover is then determined for each segment. Since
each satellite covers almost the entire earth in an 18-day cycle, satellite data can be made available for the entire sampling frame.

The Statistical Reporting Service has conducted several studies using LANDSAT data [Sigman, et al. (1978)]. In all of these studies, LANDSAT data were used as auxiliary information in regression estimation techniques. Previously, the Statistical Reporting Service used a direct expansion estimator which is the usual estimator for totals in stratified random sampling. The estimators using LANDSAT data are compared with the direct expansion estimator by calculating the relative efficiency which is defined to be the ratio of the variance of the direct expansion estimator to the variance of the regression estimator. The regression estimator is "better" if the relative efficiency is larger than one. A 1975 Illinois study used LANDSAT data to estimate crop acreages at the county and multi-county level with mixed results. Among the better results, the multi-county relative efficiencies ranged from 1.3 to 6.3 for corn. In a 1976 Kansas study of winter wheat, estimates using LANDSAT data had relative efficiencies of 3.1 to 13.0, except where cloud cover interfered with the satellite imagery. In Kings County, California, studies in 1976 and 1977 successfully used LANDSAT data to make timely crop estimates with relative efficiencies of 5.2 to 28.0. These studies showed the regression estimators using LANDSAT data to be an improvement over the direct expansion estimators.
B. The Battese-Fuller Predictor

This section is devoted to the crop area estimation procedure recently suggested by Battese and Fuller (1981, 1982). Preliminary studies presented in their papers showed their predictor to be preferable to the direct expansion estimator previously in use. The Statistical Reporting Service has indicated that the software for the Battese-Fuller procedure will be implemented for the prediction of county crop areas. The Battese-Fuller procedure and results will be described here in general terms and without proof.

1. The nested-error model and assumptions

In this section the model assumptions of the Battese-Fuller approach to small area estimation are presented. The approach assumes that the population is composed of units grouped into clusters (i.e., small areas) and that a sample of clusters is selected. Within each selected cluster, a sample of units is obtained, and the observations are assumed to have a correlation structure that is defined by the nested-error model. Prediction of the cluster means for the units within the clusters is assumed to be of interest. For the crop estimation problem, the units are the segments of the June Enumerative Survey, and the clusters are counties.

The nested-error regression model is defined by
\[ \begin{align*}
  y_{ij} &= \xi_{ij} \beta + u_{ij} , \\
  u_{ij} &= v_i + e_{ij} \quad i = 1, 2, \ldots, t, \ j = 1, 2, \ldots, n_i ,
\end{align*} \] (3.1)

where \( i \) is the cluster identification of the units; \( n_i \) is the number of sampled units in the \( i \)-th cluster; \( y_{ij} \) is the value of the study variate in the \( j \)-th unit of the \( i \)-th cluster; \( \xi_{ij} \) is a \((1 \times k)\) vector of fixed auxiliary variates; and \( \beta \) is a \((k \times 1)\) vector of unknown parameters. The random errors, \( v_i \), \( i = 1, 2, \ldots, t \), are assumed to be \( \text{NID}(0, \sigma_v^2) \) independent of the \( e_{ij} \)'s, which are assumed to be \( \text{NID}(0, \sigma_e^2) \). Thus, the variances and covariances of the random errors, \( u_{ij} \), are

\[
E(u_{ij} u_{lk}) = \begin{cases} 
\sigma_v^2 + \sigma_e^2 & \text{if } i = l \text{ and } j = k \\
\sigma_v^2 & \text{if } i = l \text{ and } j \neq k \\
0 & \text{if } i \neq l
\end{cases} \] (3.2)

The covariance matrix of the \( u_{ij} \)'s is then a block-diagonal matrix denoted by

\[
V = \text{Diag} (V_i, i = 1, 2, \ldots, t) , \] (3.3)

where
\[ V_i = \sigma_e^2 I_{n_i} + \sigma_v^2 J_{n_i}, \]

where \( I_{n_i} \) is the identity matrix of order \( n_i \), and \( J_{n_i} \) is the \((n_i \times n_i)\) matrix having all elements equal to one.

It is assumed that predictions are required for the cluster means,

\[ u_i = \bar{x}_{i}(p) \hat{\beta} + v_i, \quad i = 1, 2, \ldots, t, \quad (3.4) \]

where \( \bar{x}_{i}(p) = \frac{1}{N_i} \sum_{j=1}^{N_i} x_{ij} \) is the \((1 \times k)\) vector of the population means of the \( x_{ij} \)'s for the \( N_i \) total units in the \( i \)-th cluster. It is assumed that \( \bar{x}_{i}(p) \) is known. Initially, it will be assumed that \( \sigma_v^2 \) and \( \sigma_e^2 \) are also known.

In the crop estimation problem, \( y \) corresponds to the June Enumerative Survey data, \( X \) contains the LANDSAT data and possibly an intercept, and \( v \) contains random county effects.

2. Predictors when \( \hat{\beta} \) is known

Usually \( \hat{\beta} \) is not known in practice. The case with \( \hat{\beta} \) known is considered because it is simpler and it motivates the unknown \( \hat{\beta} \) case.

When \( \hat{\beta} \) is known, the problem of predicting the cluster means reduces to that of making predictions for the cluster effects, \( v_i \), \( i = 1, 2, \ldots, t \). The predictor for the cluster mean is

\[ \tilde{u}_i = \bar{x}_{i}(p) \hat{\beta} + \tilde{v}_i, \quad \text{where} \quad \tilde{v}_i \quad \text{is the predictor for} \quad v_i \quad \text{that is based on the observable random errors}, \quad u_{ij} = y_{ij} - x_{ij} \hat{\beta}. \]
cluster mean predictors, \( \tilde{u}_i \), are the same as those of the cluster
effects, \( \tilde{v}_i \), since one is merely a translation of the other by the
constant \( \bar{x}_{i}(p)\).

If the sample mean of the observable random errors is represented
by \( \bar{e}_i \), then it is the sum of the \( i \)-th cluster effect, \( v_i \), and the
sample mean, \( \bar{e}_i \), of the residual errors for the \( i \)-th cluster. Now,
the variance of \( \bar{u}_i \) is \( \sigma_v^2 + \sigma_e^2/n_i \) and the covariance between \( \bar{u}_i \)
and \( v_i \) is \( \sigma_v \). The prediction of \( v_i \) given knowledge of the sample
mean, \( \bar{u}_i \), is considered for the class of linear predictors defined by

\[
\tilde{v}_i(\delta) = \delta \bar{u}_i, \quad 0 < \delta < 1. \tag{3.5}
\]

The mean squared error of predictor (3.5) is

\[
E\{(\tilde{v}_i(\delta) - v_i)^2\} = (1 - \delta)^2 \sigma_v^2 + \delta^2 \sigma_e^2/n_i. \tag{3.6}
\]

This is a quadratic function of \( \delta \) which has a minimum value when \( \delta \)
is equal to the ratio of the covariance of \( \bar{u}_i \) and \( v_i \) to the
variance of \( \bar{u}_i \). This ratio is denoted by \( \gamma_i \); i.e.,

\[
\gamma_i = \sigma_v^2(\sigma_v^2 + \sigma_e^2/n_i)^{-1}. \tag{3.7}
\]

The value of \( \delta \) which minimizes the mean squared error can be found by
the usual differentiation procedure for finding critical points.
Alternatively, since normality is assumed, it follows that $\gamma_1 \bar{u}_1$ minimizes the mean squared error because it is the conditional expectation of $v_1$ given $\bar{u}_1$. The mean squared error of this "best" predictor is $\gamma_1 \sigma^2_n$, which is equal to $(1 - \gamma_1) \sigma^2_v$.

The "zero predictor" is defined to be $\tilde{v}_i^{(0)} = 0$, and it has a mean squared error of $\sigma^2_v$. The predictor $\tilde{v}_i^{(1)} = \bar{u}_i$ is conditionally unbiased for $v_1$ [i.e., $E(\bar{u}_i | v_1) = v_1$], and it has a mean squared error of $\sigma^2_e/n_1$. The question of whether the conditionally unbiased predictor is better (i.e., has a smaller mean squared error) than the zero predictor depends on the value of $\gamma_1$. In fact, the conditionally unbiased predictor is better if and only if $\gamma_1$ exceeds 0.5.

Although the predictor $\tilde{v}_i^{(1)} = \bar{u}_i$ is conditionally unbiased for $v_1$, the general predictor, $\tilde{v}_i^{(\delta)}$, has a conditional bias, namely,

$$E(\tilde{v}_i^{(\delta)} | v_1) - v_1 = \delta_i v_1 - v_1 = -(1 - \delta_i)v_1.$$ 

The expectation of the square of the conditional bias is called the mean squared conditional bias. This is given by

$$E[(E(\tilde{v}_i^{(\delta)} | v_1) - v_1)^2] = (1 - \delta_i^2) \sigma^2_v.$$ 

This is a quadratic function of $\delta_i$, $0 < \delta_i < 1$, which has its largest value, $\sigma^2_v$, at $\delta_i = 0$ and decreases monotonically to 0 at $\delta_i = 1$. 

3. **Predictors when \( \beta \) is unknown**

When \( \beta \) is unknown, estimation of \( \beta \) is required in addition to the cluster effects \( v_i, i = 1, 2, \ldots, t \). The cluster means are predicted by the class of linear predictors defined by

\[
\tilde{\mu}_1^{(\delta)} = \bar{x}_1(p) \tilde{\beta} + \delta_1 (\bar{y}_i - \bar{x}_i. \tilde{\beta}),
\]

where \( \bar{y}_i \) and \( \bar{x}_i. \) represent the sample means of the observations for the \( i \)-th cluster, and \( \tilde{\beta} \) is the best linear unbiased estimator for \( \beta \).

Note that when \( \delta_1 = 1 \),

\[
\tilde{\mu}_1^{(1)} = \bar{x}_1(p) \tilde{\beta} + (\bar{y}_i - \bar{x}_i. \tilde{\beta}) = \bar{y}_i. + \tilde{\beta}(\bar{x}_1(p) - \bar{x}_1.).
\]

is the sample mean of the \( i \)-th cluster adjusted for the difference between \( \bar{x}_1(p) \) and \( \bar{x}_1. \). When \( \delta_1 = 0 \),

\[
\tilde{\mu}_1^{(0)} = \bar{x}_1(p) \tilde{\beta}
\]

is the classical regression estimator (based on the whole population) which takes values from the regression line that is estimated by use of all the sample data.

The mean squared error of \( \tilde{\mu}_1^{(\delta)} \) is

\[
E[(\tilde{\mu}_1^{(\delta)} - \mu_1)^2] = (\bar{x}_1(p) - \delta_1 \bar{x}_1.) \nu(\tilde{\beta}) (\bar{x}_1(p) - \delta_1 \bar{x}_1.),
\]
where $V(\hat{g})$ is the covariance matrix of $\tilde{g}$. As in the case when $g$ is known, the mean squared error is minimized when $\delta_1$ is equal to $\gamma_1 = \sigma_v^2(a_v^2 + \sigma_e^2/n_1)^{-1}$. In fact, when $\delta_1 = \gamma_1$, the predictor is the best linear unbiased predictor [Goldberger (1962)]. By substitution into (3.10), the minimum mean squared error for the best predictor is

$$E[(\hat{\mu}_1(\gamma) - \mu_1)^2] = (\bar{x}_{1}(p) - \gamma_1 \bar{x}_1,_{1})V(\hat{g})(\bar{x}_{1}(p) - \gamma_1 \bar{x}_1,_{1})' + (1 - \gamma_1)\sigma_v^2. \quad (3.11)$$

The mean squared error of the general predictor, $\hat{\mu}_1(\delta)$, can be expressed in terms of that for the best predictor by

$$E[(\hat{\mu}_1(\delta) - \mu_1)^2] = E[(\hat{\mu}_1(\gamma) - \mu_1)^2] + (\delta_1 - \gamma_1)^2[(\sigma_v^2 + \sigma_e^2/n_1)\
- \bar{x}_1,_{1}V(\hat{g})\bar{x}_1,_{1}] . \quad (3.12)$$

The general predictor $\hat{\mu}_1(\delta)$ is a conditionally biased predictor for $\mu_1$ with a mean squared conditional bias of
where \( \mathbf{y} = (y_1, y_2, \ldots, y_c)' \). This is a quadratic function of \( \delta_1 \) which has a positive second derivative. Typically, the mean squared bias attains a minimum near \( \delta_1 = 1 \). In this case, \( \hat{\mu}_1^{(1)} \) is approximately conditionally unbiased for \( \mu_1 \).

4. Predictions for nonsampled areas

One of the objectives of small-area estimation techniques is to make predictions for nonsampled areas using auxiliary information and data from sampled units. Suppose that the mean of a nonsampled cluster is denoted by \( \mu_* \), where

\[
\mu_* = \bar{X}_*(p)\bar{\beta} + v_*
\]

(3.14)

such that \( \bar{X}_*(p) \) is a known vector, and \( v_* \) has zero mean, variance \( \sigma_v^2 \), and is independent of the random effects, \( v_1, v_2, \ldots, v_c \), for the sample clusters. The best predictor for \( \mu_* \) in the class of linear predictors is \( \hat{\mu}_*^{(0)} = \bar{X}_*(p)\bar{\beta} \). The mean squared error for the predictor \( \hat{\mu}_*^{(0)} \) is
C. An Extension for Stratification

The June Enumerative Survey stratifies segments by similar land use. These strata can be thought of as being crossed with the counties. Walker and Sigman (1982) have extended the Battese-Fuller predictor to take the stratification into account. The work of Walker and Sigman is summarized in this section, but the notation and derivations have been modified to emphasize the similarities with the results of the previous section.

1. The model and assumptions

For the j-th unit in the i-th county and the h-th stratum, the following model equation is assumed:

\[ y_{hij} = \delta_h x_{hij} + v_{hi} + e_{hij} , \]

\[ h = 1, \ldots, s, \quad i = 1, \ldots, t, \]

\[ j = 1, \ldots, n_{hi} , \]

where \( \delta_h \) is a \((k \times 1)\) vector of fixed parameters for the h-th stratum, \( v_{hi} \) is a random component for the i-th county and the h-th stratum, and \( e_{hij} \) is a random error. The variance-covariance structure is given by
$E(u_{hij} u_{k\ell m}) = \begin{cases} 
\sigma_{vvhk} + \sigma_{e\ell hh} & \text{if } h = k, i = \ell, j = m \\
\sigma_{vvhk} & \text{if } h = k, i = \ell, j \neq m \\
\sigma_{vvhk} & \text{if } h \neq k, i = \ell \\
0 & \text{if } i \neq \ell 
\end{cases} \quad (3.17)$

It is assumed that the variance components in (3.17) are known.

Let the matrix

$$\bar{X}_{i(p)} = \text{block diag}(\bar{X}_{hi(p)}), \quad h = 1, \ldots, s,$$

contain the known population means of the auxiliary variable for the $s$ strata in the $i$-th county. Let

$$\bar{X}_{i.} = \text{block diag}(\bar{X}_{hi.})$$

be similarly defined for the sample data. Let

$$\beta = (\beta_1', \beta_2', \ldots, \beta_s')',$$

$$\nu_i = (\nu_{i1}', \nu_{i2}', \ldots, \nu_{is}')',$$

and
where

\[ \bar{x}_{hi} = \frac{1}{n_{hi}} \sum_{j=1}^{n_{hi}} x_{hij}, \]

Let

\[ \bar{x}_i = (w_{i1}, w_{i2}, \ldots, w_{is})', \]

be the \((s \times 1)\) vector of weights for the strata in the \(i\)-th county, where \(N_{hi}\) is the total number of units in the \(h\)-th stratum of the \(i\)-th county, and

\[ N_{i} = \sum_{h=1}^{s} N_{hi}. \]

It is desirable to estimate the county-stratum means,

\[ \mu_{hi} = \bar{x}_{hi}(p) + v_{hi}, \quad (3.18) \]

and the county mean,
If \( \xi \) is known, then only the county-stratum effects, \( v_{hi} \), must be predicted in order to estimate the county-stratum means (3.18) and the county mean (3.19). If \( \xi \) is known, then the residuals, \( u_{hij} = v_{hi} + e_{hij} \), are observable. The vector of county-stratum effects, \( \bar{v}_i \), is estimated by

\[
\bar{v}_i = (\bar{u}_{li}, \ldots, \bar{u}_{si}),
\]

where

\[
\bar{u}_{hi} = \frac{1}{n_{hi}} \sum_{j=1}^{n_{hi}} u_{hij}.
\]

The class of predictors considered for predicting the county-stratum effects \( v_{hi} \) for the \( h \)-th stratum of the \( i \)-th county is defined by

\[
\gamma(\delta) v_{hi} = \delta_{hi} \bar{v}_i.
\]
where $\delta_{hi}$ is some \((1 \times s)\) vector whose elements are between 0 and 1. The county-stratum mean for the \(h\)-th stratum of the \(i\)-th county is estimated by

$$\mu_{hi} = \bar{x}_{hi}(p)\delta_{hi} + \delta_{hi}y_i.$$  

Let

$$\mu_1 = (\mu_{11}, \ldots, \mu_{si}),$$

$$= \bar{x}_i(p)\delta + D_1y_i,$$  \hspace{1cm} (3.20)

where the \(h\)-th row of \(D_1\) is $\delta_{hi}$. Then, the county mean is estimated by

$$\mu_i = \sum_{h=1}^{s} w_{hi}\mu_{hi}$$

$$= \bar{x}_i\mu_1.$$  \hspace{1cm} (3.21)

The optimal matrix \(D\) for estimating the county mean is the matrix that minimizes the mean squared error of $\mu'_i$. Let
\[ H = E\{y'_4 y'_1\} \]

\[
\begin{bmatrix}
\sigma_{vv11} & \sigma_{vv12} & \cdots & \sigma_{vv1s} \\
\sigma_{vv12} & \sigma_{vv22} & \cdots & \sigma_{vv2s} \\
. & . & \cdots & . \\
. & . & \cdots & . \\
\sigma_{vvls} & \sigma_{vv2s} & \cdots & \sigma_{vvss}
\end{bmatrix}, \quad (3.22)
\]

and let

\[ R_1 = E\{e_{\text{e}_1} e'_{\text{e}_1}\} \]

\[ = \text{diag}(n^{-1} \sigma_{\text{ee}hh}) \]. \quad (3.23) \]

Then,

\[ A_1 = E\{\text{h}_{\text{e}_1} \text{h}'_{\text{e}_1}\} \]

\[ = H + R_1 \]. \quad (3.24) \]

The error of the estimator (3.21) is
Therefore, the mean squared error of \( \hat{\mu}_1 \) can be written as

\[
\text{MSE}(\hat{\mu}_1) = \mathbf{w}_1'((I - \mathbf{D}_1)\mathbf{H}(I - \mathbf{D}_1)' + \mathbf{D}_1\mathbf{R}_1\mathbf{D}_1')\mathbf{w}_1. \tag{3.25}
\]

The mean squared error is minimized when \( \mathbf{D}_1 = \mathbf{A}_1^{-1}\mathbf{H} \). If \( \sigma_{vvhk} = 0 \) whenever \( h \neq k \), then this optimal \( \mathbf{D}_1 \) reduces to

\[\mathbf{D}_1 = \text{diag}(\gamma_{1h}, \ldots, \gamma_{s_1}),\]

where

\[
\gamma_{hi} = \sigma_{vvh}^{-1}\sigma_{vvh}^{-1} + \sigma_{ehh}^{-1}. \tag{3.26}
\]

The mean squared conditional bias of \( \hat{\mu}_1 \) is given by

\[
\text{MSCB}(\hat{\mu}_1) = \mathbf{w}_1'\mathbb{E}\big((\mathbb{E}(\mathbf{D}_1\mathbf{y}_1, y_1|\mathbf{x}_1) - \mathbf{y}_1)^2\big)\mathbf{w}_1
\]

\[= \mathbf{w}_1'(\mathbf{D}_1 - \mathbf{I})\mathbf{y}_1\mathbf{v}_1'(\mathbf{D}_1 - \mathbf{I})'\mathbf{w}_1
\]

\[= \mathbf{w}_1'(\mathbf{I} - \mathbf{D}_1)(\mathbf{I} - \mathbf{D}_1)'\mathbf{w}_1. \tag{3.27}
\]
3. **Predictors when $\beta$ is unknown**

For the case in which $\beta$ is estimated by the generalized least squares estimator $\hat{\beta}$, the county-stratum effect $v_{hi}$ is predicted by the class of predictors

$$
\tilde{v}_{hi} = \delta_{hi} \tilde{u}_i,
$$

where

$$
\tilde{u}_i = (\tilde{u}_{li}, \ldots, \tilde{u}_{si}),
$$

and

$$
\tilde{u}_{hi} = y_{hi} - \tilde{x}_{hi} \hat{\beta}_h.
$$

The vector of county-stratum means is predicted by

$$
\tilde{\nu}_{D_{hi}} = \tilde{x}_{i(p)} \hat{\beta} + D_{hi} \tilde{u}_i,
$$

(3.28)

where the $h$-th row of $D_1$ is $\hat{\beta}_{hi}$. The county mean is predicted by

$$
\tilde{\mu}_{D_{hi}} = \tilde{\nu}'_{D_{hi}} \tilde{\nu}_{D_{hi}}.
$$

(3.29)
The mean squared error of $\hat{u}_1^{(D)}$ is given by

$$\text{MSE}\{\hat{u}_1^{(D)}\} = E\{w_1' \tilde{\mathbf{x}}_1(p) \tilde{\mathbf{y}} + D_1 \tilde{\mathbf{y}}_1. - \tilde{\mathbf{x}}_1(p) \tilde{\mathbf{y}} - \mathbf{y}_1)(\tilde{\mathbf{x}}_1(p) \tilde{\mathbf{y}}$$

$$+ D_1 \tilde{\mathbf{y}}_1. - \tilde{\mathbf{x}}_1(p) \tilde{\mathbf{y}} - \mathbf{y}_1)' \mathbf{y}_1\}$$

$$= E\{w_1' ((\tilde{\mathbf{x}}_1(p) - D_1 \tilde{\mathbf{x}}_1.)(\tilde{\mathbf{y}} - \mathbf{y}) - (I - D_1) \mathbf{y}_1$$

$$+ D_1 \tilde{\mathbf{y}}_1.)((\tilde{\mathbf{x}}_1(p) - D_1 \tilde{\mathbf{x}}_1.)(\tilde{\mathbf{y}} - \mathbf{y})

$$- (I - D_1) \mathbf{y}_1 + D_1 \tilde{\mathbf{y}}_1.)' \mathbf{y}_1\}$$

$$= w_1' ((\tilde{\mathbf{x}}_1(p) - D_1 \tilde{\mathbf{x}}_1.))E\{(\tilde{\mathbf{y}} - \mathbf{y})(\tilde{\mathbf{y}} - \mathbf{y})'\}(\tilde{\mathbf{x}}_1(p) - D_1 \tilde{\mathbf{x}}_1.)'$$

$$+ (I - D_1)E\{\mathbf{y}_1 \mathbf{y}_1'\}(I - D_1)'$$

$$+ D_1 E(\tilde{\mathbf{y}}_1. \tilde{\mathbf{y}}_1. D_1')$$

$$- (\tilde{\mathbf{x}}_1(p) - D_1 \tilde{\mathbf{x}}_1.))E\{(\tilde{\mathbf{y}} - \mathbf{y}) \mathbf{y}_1'(I - D_1)'$$

$$- (I - D_1)E\{\mathbf{y}_1 (\tilde{\mathbf{y}} - \mathbf{y})'\}(\tilde{\mathbf{x}}_1(p) - D_1 \tilde{\mathbf{x}}_1.)$$
Now, \[ E[(\tilde{\mathbf{g}} - \mathbf{g})\mathbf{v}_i^\prime] = (\mathbf{x}'\mathbf{v}^{-1}\mathbf{x})^{-1}\mathbf{x}'\mathbf{v}^{-1} E[\mathbf{u} \mathbf{v}_i^\prime] \]
\[ = \mathbf{v}(\tilde{\mathbf{g}})\mathbf{x}'\mathbf{v}^{-1} \mathbf{c}_1, \]

where \[ \mathbf{V} = E[\mathbf{u} \mathbf{u}'] = \text{block diag}(\mathbf{V}_i), \]
\[ \mathbf{V}_i = E[u_i u_i'] \]
is a block matrix whose \((h, k)\) block is given by \[
\mathbf{V}_{i_{hk}} = \begin{cases} 
\sigma_{eih}_i + \sigma_{vih} J_{n_i} x n_h, & h = k \\
\sigma_{vih} J_{n_i} x n_k, & h \neq k 
\end{cases}
\]
and
Similarly,

\[ E(\tilde{\beta} - \beta)\tilde{e}_i^2_i = V(\tilde{\beta})X_i^{-1}C_2, \]

where

\[ C_{21} = \text{block diag}(n_{hi}^{-1}, \sigma_{ehh}^{-1} n_{hi}), \]

Then,\n
\[ \text{MSE}(\hat{\mu}^{(D)}) = X_1'(I - D_1\bar{x}_i) V(\tilde{\beta})(\bar{x}_i(p) - D_1\bar{x}_i) V(\tilde{\beta})X_1^{-1}C_{11}(I - D_1)' + (I - D_1)M(I - D_1)' + D_1R_1D_1' \]

\[ - (\bar{x}_i(p) - D_1\bar{x}_i)V(\tilde{\beta})X_1^{-1}C_{11}(I - D_1)' \]

\[ - (\bar{x}_i(p) - D_1\bar{x}_i)V(\tilde{\beta})X_1^{-1}C_{11}(I - D_1)' \]
For the special case in which \( \sigma_{vvhk} = 0 \) for \( h \neq k \), then

\[
\mathbf{v}_i = \text{block diag}(\sigma_{eehh} I_{n_{hi}} + \sigma_{vvhh}^{1/2} n_{hi} \times n_{hi})
\]

and

\[
\mathbf{c}_{il} = \text{block diag}(\sigma_{vvhh}^{-1} n_{hi})
\]

So

\[
E(\mathbf{g} - \mathbf{g})_{v_i'} = \mathbf{v}(\mathbf{g}) \text{block diag}(\mathbf{x}_i'_{hi} \mathbf{v}_i^{-1} \sigma_{vvhh}^{1/2} n_{hi})
\]

\[
= \mathbf{v}(\mathbf{g}) \text{b.d.} \left\{ \mathbf{x}_i'_{hi} \sigma_{eehh}^{-1} \sigma_{vvhh} \left( \mathbf{I} - \frac{\sigma_{vvhh}}{\sigma_{eehh} + n_{hi} \sigma_{vvhh}} \mathbf{J} \right)^{-1/2} \right\}
\]
\[
= \mathbb{V}(\tilde{g}) \text{b.d.} \left\{ X_{hi}^\prime \frac{\sigma_{vvhh}}{\sigma_{eehh} + n_{hi} \sigma_{vvhh}} l_{n_{hi}} \right\}
\]

\[
= \mathbb{V}(\tilde{g}) \text{b.d.} \{ X_{hi}. \gamma_{hi} \}
\]

\[
= \mathbb{V}(\tilde{g}) X_{hi}^\prime \Gamma_i,
\]

where

\[
\Gamma_i = \text{diag}(\gamma_{i1}, \ldots, \gamma_{si})
\]

and

\[
\mathbb{E}\{(\tilde{g} - \tilde{g})X_{hi}^\prime \} = \mathbb{V}(\tilde{g}) \text{b.d.} \left\{ X_{hi} n_{hi}^{-1} \left( I - \frac{\sigma_{vvhh}}{\sigma_{eehh} + n_{hi} \sigma_{vvhh}} J \right) l_{n_{hi}} \right\}
\]

\[
= \mathbb{V}(\tilde{g}) \text{b.d.} \left\{ X_{hi} n_{hi}^{-1} \frac{\sigma_{eehh}}{\sigma_{eehh} + n_{hi} \sigma_{vvhh}} l_{n_{hi}} \right\}
\]

\[
= \mathbb{V}(\tilde{g}) \text{b.d.} \{ X_{hi}^\prime (1 - \gamma_{hi}) \}
\]

\[
= \mathbb{V}(\tilde{g}) X_{hi}^\prime (I - \Gamma_i).
\]
The mean squared error of \( \tilde{\nu}_1^{(D)} \) reduces to

\[
\text{MSE}(\tilde{\nu}_1^{(D)}) = E_1'[\langle \tilde{x}_1(p) - D_1 \tilde{x}_1, \rangle V(\tilde{\theta}) \langle \tilde{x}_1(p) - D_1 \tilde{x}_1 \rangle'] \\
+ (I - D_1)\text{diag}(\sigma_{vvh})' (I - D_1)' + D_1 R D_1' \\
- (\tilde{x}_1(p) - D_1 \tilde{x}_1, \rangle V(\tilde{\theta}) \tilde{x}_1', \Gamma_1 (I - D_1)' \\
- (I - D_1) \Gamma_1 \tilde{x}_1, \rangle V(\tilde{\theta}) \tilde{x}_1(p) - D_1 \tilde{x}_1, \rangle' \\
+ (\tilde{x}_1(p) - D_1 \tilde{x}_1, \rangle V(\tilde{\theta}) \tilde{x}_1', (I - \Gamma_1) D_1' \\
+ D_1 (I - \Gamma_1) \tilde{x}_1, \rangle V(\tilde{\theta}) \tilde{x}_1(p) - D_1 \tilde{x}_1, \rangle'] E_1. \\
(3.31)
\]

The expected bias of \( \tilde{\nu}_1^{(D)} \), conditioned on \( \chi \), is

\[
E(\tilde{\nu}_1^{(D)} - \nu_1 | \chi) = E_1[E(\tilde{x}_1(p) - D_1 \tilde{x}_1, \rangle \tilde{\theta} - \theta - (I - D_1) \chi_1 \\
+ D_1 \tilde{\epsilon}_1, | \chi)] \\
= E_1'[\langle \tilde{x}_1(p) - D_1 \tilde{x}_1, \rangle E(\tilde{\theta} - \theta | \chi) - (I - D_1) \chi_1]
\]
The mean squared conditional bias of $\hat{\nu}_i^{(D)}$ is given by

\begin{equation}
\text{MSCB}(\hat{\nu}_i^{(D)}) = \varphi_1'((\tilde{\mathbf{x}}_1(p) - D_1\tilde{\mathbf{x}}_1. )\mathbf{V}(\tilde{\mathbf{g}})\mathbf{X}'\mathbf{V}^{-1} \mathbf{E}(\mathbf{y}'\mathbf{y}' \mathbf{V}^{-1} \mathbf{X} \mathbf{V}(\tilde{\mathbf{g}})(\tilde{\mathbf{x}}_1(p))
- D_1\tilde{\mathbf{x}}_1. )' + (I - D_1)\mathbf{E}(\mathbf{x}_1.')(I - D_1)'$
\end{equation}

\begin{equation}
+ (\tilde{\mathbf{x}}_1(p) - D_1\tilde{\mathbf{x}}_1. )\mathbf{V}(\tilde{\mathbf{g}})\mathbf{X}'\mathbf{V}^{-1} \mathbf{E}(\mathbf{x}_1.')(I - D_1)'
+ (I - D_1)\mathbf{E}(\mathbf{x}_1.')(I - D_1)' \mathbf{x}' \mathbf{V}(\tilde{\mathbf{g}})(\tilde{\mathbf{x}}_1(p) - D_1\tilde{\mathbf{x}}_1. )'\mathbf{x}_1$
\end{equation}

\begin{equation}
= \varphi_1'((\tilde{\mathbf{x}}_1(p) - D_1\tilde{\mathbf{x}}_1. )\mathbf{V}(\tilde{\mathbf{g}})\mathbf{X}'\mathbf{V}^{-1} \mathbf{C}_3\mathbf{V}^{-1} \mathbf{X} \mathbf{V}(\tilde{\mathbf{g}})(\tilde{\mathbf{x}}_1(p))
- D_1\tilde{\mathbf{x}}_1. )' + (I - D_1)\mathbf{H}(I - D_1)'
+ (\tilde{\mathbf{x}}_1(p) - D_1\tilde{\mathbf{x}}_1. )\mathbf{V}(\tilde{\mathbf{g}})\mathbf{X}'\mathbf{V}^{-1} \mathbf{C}_1(\mathbf{I} - D_1)'
+ (\mathbf{I} - D_1)\mathbf{C}_1\mathbf{V}^{-1} \mathbf{x}_1 \mathbf{V}(\tilde{\mathbf{g}})(\tilde{\mathbf{x}}_1(p) - D_1\tilde{\mathbf{x}}_1. )'\mathbf{x}_1 , (3.32)
\end{equation}
where the \((h, k)\) block of \(C_{31}\) is 
\[
\sigma_{\text{vvhk}}^{1}_{h_i} x^{n_{hi}}_{hi},
\]
h, k, = 1, \ldots, s. For the special case in which \(\sigma_{\text{vvhk}} = 0\) for \(h \neq k\), then

\[
x'V^{-1}C_{31}V^{-1}x = \text{block diag}\{x'_{hi}V^{-1}\sigma_{\text{vvh}}JV^{-1}_{hi}x_{hi}\}
\]

\[
= \text{b.d.}\left\{x'_{hi}e_{\text{vvh}}\sigma_{\text{vvh}}\left(I - \frac{\sigma_{\text{vvh}}}{\sigma_{\text{e}} + n_{hi}\sigma_{\text{vvh}}}J\right)J\left(I
\right.
\]

\[
\left.- \frac{\sigma_{\text{vvh}}}{\sigma_{\text{e}} + n_{hi}\sigma_{\text{vvh}}}J\right)_{x_{hi}}\right\}
\]

\[
= \text{b.d.}\{x'_{hi}e_{\text{e}}\sigma_{\text{vvh}}(1 - \gamma_{hi})^{2}J_{x_{hi}}\}
\]

\[
= \text{b.d.}\left\{x'_{hi}e_{\text{e}}\frac{n^{2}_{hi}\sigma_{\text{vvh}}}{(\sigma_{\text{e}} + n_{hi}\sigma_{\text{vvh}})^{2}}x_{hi}\right\}
\]

\[
= \text{b.d.}\{\bar{x}^{i}_{hi}G_{1}\bar{x}^{i}_{hi}\}
\]

where

\[
G_{1} = \text{diag}(\sigma_{\text{vvh}}^{-1}, \gamma_{hi}^{2})
\]
The mean squared conditional bias reduces to

\[
\text{MSCB}\{\tilde{u}_1^{(D)}\} = \mathbf{x}_1'[\{\tilde{x}_{1}(p) - D_1 \tilde{x}_1 \} \mathbf{V} (\tilde{\mathbf{g}}) \mathbf{b.d.} (\tilde{x}_1, D_1 \tilde{x}_1) \mathbf{V} (\tilde{\mathbf{g}}) (\tilde{x}_{1}(p) \\
- D_1 \tilde{x}_1)]' + (I - D_1) \text{diag}(\sigma_{vvh})(I - D_1)'

+ (\tilde{x}_{1}(p) - D_1 \tilde{x}_1) \mathbf{V} (\tilde{\mathbf{g}}) \tilde{x}_1 \mathbf{r}_1 (I - D_1)'

+ (I - D_1) \mathbf{r}_1 \tilde{x}_1 \mathbf{V} (\tilde{\mathbf{g}}) (\tilde{x}_{1}(p) - D_1 \tilde{x}_1)'] \mathbf{x}_1.
\]

(3.33)
IV. ESTIMATION OF VARIANCE COMPONENTS

All of the results on the Battese-Fuller predictor in Section III.B are based on the assumption that $\sigma_v^2$ and $\sigma_e^2$ are known. When $\sigma_v^2$ and $\sigma_e^2$ are not known, they can be estimated by any of a number of methods. Some of the major variance component estimators are discussed by Searle (1971), Rao (1971), and Harville (1977). The estimators of the variance components can then be substituted into the formulas for the predictors, mean squared errors, and mean squared conditional biases. The actual mean squared errors of the resulting predictors will increase due to the estimation of the variance components. When prior estimates of the variance components are available, they can be combined with the sample estimates to give improved estimates of the variance components.

Some properties of the estimators and estimated predictors are given in terms of their asymptotic behavior as the number of observations increases. The order in probability concepts used to describe asymptotic convergence are defined by Fuller (1976).

A. The Fitting-of-Constants Estimators

The procedure for computing the fitting-of-constants estimators, also known as Henderson's Method 3 [Searle (1971)], for the nested-error model was given by Fuller and Battese (1973). Under normality, which is assumed here, the variances of the estimators were presented by Battese
and Fuller (1982). The procedure involves two ordinary least squares regressions.

For the first regression, the sample $X$-matrix is augmented by a set of indicator variables for the cluster effects. Let $X_\ast$ denote the augmented matrix. The ordinary least squares regression of $y$ on $X_\ast$ is equivalent to the regression of the $y$ deviations, $y_{ij} - \bar{y}_i$, on the $x$ deviations, $x_{ij} - \bar{x}_i$, which are not identically zero. The residuals from this regression, $\hat{\varepsilon}$, are used to compute an estimator of the component for variation within clusters. The estimator is

$$\hat{\sigma}_e^2 = \hat{\varepsilon}'\hat{\varepsilon}/n_e,$$

(4.1)

where $n_e$ is the degrees of freedom for the "within" component. With $n$, $k$, and $t$ defined as in Section III.B,

$$n_e = n - (k + t - \lambda),$$

where $\lambda$ is the number of $x$-variables which are linear combinations of the indicator variables. In other words, $n_e = n - \text{rank}(X_\ast)$. The estimator $\hat{\sigma}_e^2$ is unbiased for $\sigma_e^2$.

The second regression is the ordinary least squares regression of $y$ on $X$. Let $\hat{y}$ denote the residuals. The mean square for variation among clusters, MSA, is defined by
where \( t - \lambda \) is the degrees of freedom among clusters. The expected value of \( \hat{u}'\hat{u} \) is

\[
E\{\hat{u}'\hat{u}\} = (n - k)\sigma^2_e + (t - \lambda)n_\kappa\sigma^2_v,
\]

where

\[
n_\kappa = \sum_{i=1}^{t} n_i [1 - n_i \overline{x}_i \cdot (X'X)^{-1} \overline{x}_i' \cdot (t - \lambda)^{-1}].
\]

Therefore,

\[
E\{\text{MSA}\} = n_\kappa\sigma^2_v + \sigma^2_e,
\]

and an unbiased estimator of \( \sigma^2_v \) is

\[
\hat{\sigma}^2_v = (\text{MSA} - \hat{\sigma}^2_e)/n_\kappa.
\]

It is possible for \( \hat{\sigma}^2_v \) to be negative. In practice, if \( \hat{\sigma}^2_v < 0 \), one would use \( \hat{\sigma}^2_v = 0 \). The estimator would no longer be unbiased, but it would be in the closure of the parameter space.

The variance of \( \hat{\sigma}^2_v \) is
\[ V(\hat{\sigma}^2_e) = 2 \frac{\sigma^4}{n_e} = O((n - t)^{-1}) \] \hspace{1cm} (4.5)

The variance of MSA is

\[ V(\text{MSA}) = 2[\sigma^4 + 2 n_x \sigma^2_e \sigma^2_v + n_{xx} \sigma^4_v](t - \lambda)^{-1} \]
\[ = O(t^{-1}) , \] \hspace{1cm} (4.6)

where

\[ n_{xx} = \sum_{i=1}^{t} n_i^2[1 - \frac{\bar{x}_{i1}}{\bar{x}_{11}}(\mathbf{x}'\mathbf{x})^{-1}\bar{x}_{i1}](t - \lambda)^{-1} . \] \hspace{1cm} (4.7)

Also, \( \hat{\sigma}^2_e \) and MSA are independently distributed, so that

\[ V(\hat{\sigma}^2_v) = \frac{[V(\text{MSA}) + V(\hat{\sigma}^2_e)]}{n_x^2} \]
\[ = O(\max[t^{-1}, (n - t)^{-1}]) , \] \hspace{1cm} (4.8)

and

\[ C(\hat{\sigma}^2_v, \hat{\sigma}^2_e) = - \frac{V(\hat{\sigma}^2_e)}{n_x} . \] \hspace{1cm} (4.9)

These variances and covariances are estimated by substituting \( \hat{\sigma}^2_v \) and \( \hat{\sigma}^2_e \) for \( \sigma^2_v \) and \( \sigma^2_e \), respectively, into the above formulas.
B. Effects of Estimating the Variance Components

1. Effects on the estimation of \( \hat{\beta} \)

The estimated generalized least squares estimator for \( \hat{\beta} \) under the nested-error model is any solution \( \hat{\beta} \) to

\[
(X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}Y = \hat{\beta},
\]

where \( \hat{\Sigma} \) is formed from \( \Sigma \) by replacing \( \sigma_v^2 \) and \( \sigma_e^2 \) by \( \hat{\sigma}_v^2 \) and \( \hat{\sigma}_e^2 \), respectively. It is assumed that \( \hat{\Sigma} \) is invertible. In general, the properties of \( \hat{\beta} \) are difficult to obtain.

Khatri and Shah (1981) explored the properties of linear functions of \( \hat{\beta} \) for a class of models which includes the nested-error model. If \( h'\hat{\beta} \) is any estimable function of \( \hat{\beta} \), then \( h'\hat{\beta} \) is an unbiased estimator of \( h'\beta \). Furthermore, an expression for the variance of \( h'\hat{\beta} \) is given. Let

\[
L = I - X(X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1},
\]

and let

\[
z = h'(X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1} \sqrt{L(L'\Sigma L)^{-1}L'Y}
\]

\[
= h'(X'X)^{-1}X'(\sqrt{V}L(L'\Sigma L)^{-1}V - \sqrt{V}L(L'\Sigma L)^{-1}L'Y).
\]
Then,
\[ V\{h'\hat{\hat{\beta}}\} = h'(XX')^{-1}h + E\{z^2\} \]
\[ = V\{h'\tilde{\hat{\beta}}\} + E\{z^2\}, \]
where \( \tilde{\hat{\beta}} \) is the generalized least squares estimator of \( \beta \) computed with the true \( V \).

Using the fitting-of-constants estimators \( \hat{\sigma}_V^2 \) and \( \hat{\sigma}_e^2 \) defined in the previous section, Fuller and Battese (1973) showed that under certain regularity conditions, \( \hat{\hat{\beta}} \) has the same asymptotic distribution as \( \tilde{\hat{\beta}} \), and
\[ (\hat{\hat{\beta}} - \beta) = (\tilde{\hat{\beta}} - \beta) + O_p(n^{-1/2} \max (t^{-1/2}, (n - t)^{-1/2})) . \]

(4.11)

2. Effects on the Battese-Fuller predictor

Recall from Section III.B that the best predictor of the cluster mean is
\[ \tilde{\mu}_1(y) = \bar{x}_1(p)\tilde{\hat{\beta}} + \gamma_1(\bar{y}_1 - \bar{\bar{x}}_1.\tilde{\hat{\beta}}) . \]

(4.12)

When the variance components are estimated, then \( \tilde{\hat{\beta}} \) is replaced by \( \hat{\hat{\beta}} \) and \( \gamma_1 \) must be estimated. The approximate predictor is
One estimator of \( \gamma_i \) is \( \hat{\gamma}_i = \frac{\hat{\sigma}_v^2}{\hat{\sigma}_v^2 + n_i^{-1} \hat{\sigma}_e^2} \). An alternative estimator of \( \gamma_i \) is

\[
\hat{\gamma}_i = \frac{1 - \hat{\theta}}{1 + (a_i - 1)\hat{\theta}},
\]

where

\[
a_i = \frac{n_i^{-1}}{n_e},
\]

and

\[
\hat{\theta} = \frac{t - \lambda - 2}{t - \lambda} \frac{\hat{\sigma}_e^2}{\text{MSA}}.
\]

In practice, one would replace \((\text{MSA})^{-1} \hat{\sigma}_e^2\) with \(\min[1, \text{(MSA)}^{-1} \hat{\sigma}_e^2]\).

The estimator \(\hat{\theta}\) is approximately unbiased for

\[
\theta = (n_e \sigma_v^2 + \sigma_e^2)^{-1} \sigma_e^2.
\]

When all the \(n_i\)'s are equal, then \(a_i = 1\) for all \(i\), and \(\hat{\gamma}_i\) is unbiased for \(\gamma_i\). The estimator \(\hat{\theta}\) also corresponds to the ratio used
in constructing the James-Stein estimator [Efron and Morris (1973)] for certain special cases.

The relationship between the mean squared errors of the predictors \( \hat{\mu}_1 \) and \( \tilde{\mu}_1 \) is considered in Theorem 4.1. First, however, a lemma is introduced and some assumptions are stated concerning the model and its associated matrices.

**Lemma 4.1** Let

\[
U_1 \sim \chi^2_a \sim \Gamma\left(\frac{a}{2}, \frac{1}{2}\right) \quad \text{and}
\]

\[
U_2 \sim \chi^2_b \sim \Gamma\left(\frac{b}{2}, \frac{1}{2}\right),
\]

where \( U_2 \) is independent of \( U_1 \). Let \( h \) and \( m \) be nonnegative integers such that \( a + b + 2h - 2m > 1 \). Then,

\[
E\left\{ \frac{U_1^h}{(U_1 + U_2)^m} \right\} = \frac{\Gamma\left(\frac{a}{2} + h\right)\Gamma\left(\frac{a+b}{2} + h - m\right)}{\left(\frac{1}{2}\right)^{h-m} \Gamma\left(\frac{a}{2}\right)\Gamma\left(\frac{a+b}{2} + h\right)}.
\]

**Proof.**

\[
E\left\{ \frac{U_1^h}{(U_1 + U_2)^m} \right\} = \int_0^\infty \int_0^\infty \frac{U_1^h}{\Gamma\left(\frac{a}{2}\right)\Gamma\left(\frac{b}{2}\right)} \frac{e^{-\frac{U_1}{2}}}{U_1} \frac{e^{-\frac{U_2}{2}}}{U_2} \frac{e^{-\frac{a-1}{2}U_1 - \frac{b-1}{2}U_2}}{(U_1 + U_2)^m} dU_1 dU_2
\]
\[
\begin{align*}
    &\frac{\Gamma(a/2 + h)}{(1/2)^h \Gamma(a/2)} \frac{1}{\Gamma(a/2 + h) \Gamma(b/2)} \int_0^\infty \int_0^\infty (U_1 + U_2)^{-m} U_1^{a/2} U_2^{b/2} - 1 - (U_1 + U_2) U_1^{-m} dU_1 dU_2 \\
    &= \frac{\Gamma(a/2 + h)}{(1/2)^h \Gamma(a/2)} E\{(U_1 + U_2)^{-m}\},
\end{align*}
\]

where

\[
U_1 \sim \Gamma(a/2 + h, 1/2)
\]

and

\[
U_2 \sim \Gamma(b/2, 1/2), \text{ independently of } U_1. \text{ Then,}
\]

\[
E\left\{\frac{U_1^h}{(U_1 + U_2)^m}\right\} = \frac{\Gamma(a/2 + h)}{(1/2)^h \Gamma(a/2)} E\{U_3^{-m}\},
\]

where

\[
U_3 \sim \Gamma(a+b/2 + h, 1/2).
\]
Therefore,

\[
E\left\{ \frac{U_1^h}{(U_1 + U_2)^m} \right\} = \frac{\Gamma(a + h)}{(1)^h \Gamma(a/2)} \frac{\Gamma(a/h + h - m)}{(1)^{-m} \Gamma(a/2 + h)}
\]

\[
= \frac{\Gamma(a + h)\Gamma(a/b + h - m)}{(1)^{-m} \Gamma(a/2)\Gamma(a/b + h)} .
\]

using basic results for the gamma distribution.

The following assumptions are used in Theorem 4.1. Assumptions (1) - (3) correspond to the regularity conditions of Theorem 3 of Fuller and Battese (1973). Assumption (1) is a verifiable property of the covariance matrix of the nested-error model.

(1) For n observations, the elements of the covariance matrix \( V_n = E\{u_n u_n'\} \) and its inverse \( V_n^{-1} \) are functions of \( \sigma_v^2 \) and \( \sigma_e^2 \). The partial derivatives of the elements of \( V_n^{-1} \) with respect to \( \sigma_v^2 \) and \( \sigma_e^2 \) are continuous functions for all \( \sigma_v^2 > 0 \) and \( \sigma_e^2 > 0 \). When \( \sigma_v^2 > 0 \) and \( \sigma_e^2 > 0 \) are used in place of \( \hat{\sigma}_v^2 \) and \( \hat{\sigma}_e^2 \) in \( \hat{V}_n \), then \( \hat{V}_n^{-1} \) exists for all n.

(2) The elements of the sequences of matrices \( \{X_n\} \) and \( \{V_n\} \) are bounded sequences such that
where $G$ is a $(k \times k)$ matrix of fixed constants and $G^{-1}$ exists for all $n > k$, $\sigma^2 > 0$, and $\sigma^2 > 0$.

(3) The sequences of matrices $\{X_n\}$ and $\{V_n\}$ are such that

$$\lim_{n \to \infty} n^{-1} X_n' n^{-1} X_n = G,$$

where $G$ is a $(k \times k)$ matrix of fixed constants and $G^{-1}$ exists for all $n > k$, $\sigma^2 > 0$, and $\sigma^2 > 0$.

$$\frac{\partial V}{\partial \sigma^2_V} X_n = H_v$$

and

$$\frac{\partial V}{\partial \sigma^2_e} X_n = H_e,$$

where the elements of $H_v$ and $H_e$ are continuous functions of $\sigma^2_V$ and $\sigma^2_e$.

(4) The sequence $\{n_i\}$ is such that

$$n^{-1} \sum_{i=1}^{t} n_i^2 = O(1),$$

and

$$\sum_{i=1}^{t} |n_i - n_*| = O(1),$$

where $n_*$ is defined by (4.3).
Note that assumption (4) implies that
\[ \sum_{i=1}^{t} (n_i - n_\ast)^2 < (\sum_{i=1}^{t} |n_i - n_\ast|^2) = o(1). \] \hspace{1cm} (4.17a)

The statement of Theorem 4.1 below expresses the squared error of \( \hat{\gamma}_1 \) in terms of the squared error of \( \tilde{\gamma}_1 \) and a penalty term for estimating the variance components. The penalty term can be expressed in different forms to the order in probability of the theorem. The expectation of the penalty term in the statement of the theorem agrees with a result by Peixoto (1982) for a special case of the model. This will be discussed in more detail following the theorem.

Kackar and Harville (1980) approximated the mean squared error of the predictor with estimated variance components for the general mixed linear model. Kackar and Harville's work is similar to Theorem 4.1, but in a much more general setting.

**Theorem 4.1** Assume model (3.1) - (3.2), and let assumptions (2), (3), and (4) hold. Let
\[ \hat{\gamma}_1 = \tilde{\gamma}_1(p) \hat{\beta} + \gamma_1(\tilde{\gamma}_1, \tilde{\beta}), \]
where \( \hat{\beta} \) is defined by (4.10) and \( \hat{\gamma}_1 \) is defined by (4.14). Then,
\[ (\hat{\gamma}_1 - \gamma_1)^2 = (\tilde{\gamma}_1 - \gamma_1)^2 + \frac{a_1^2(\hat{\theta} - \theta)^2}{[1 + (a_1 - 1)\theta]^4} (\tilde{\gamma}_1 - \tilde{\gamma}_1)^2 \]
- 2[(1 - \gamma_1)v_1 - \gamma_1 e_i.\] \frac{a_1(\tilde{\theta} - \theta)}{[1 + (a_1 - 1)\theta]^2} (\bar{u}_i. - \bar{u}_.)

+ O_p(\max\{t^{-1}, (n-t)^{-1}\})

where

\[
\tilde{\theta} = \frac{t - \lambda - 2}{t - \lambda} \frac{\sigma^2}{\tilde{\omega}},
\]

\[
\tilde{\omega} = \left(\frac{n_1}{\sigma_v^2} + \frac{\sigma_e^2}{\tilde{\omega}_{\text{MSA}}}\right)^{1/2} \Sigma_{i=1}^t (\bar{u}_i. - \bar{u}_.)^2 (V(\bar{u}_i. - \bar{u}_.))^{-1},
\]

\[
V(\bar{u}_i. - \bar{u}_.) = \frac{\sigma^2}{\sigma_v} (1 - 2n^{-1} n_1 + n^{-2} \Sigma_{i=1}^t n_i^2)
+ \frac{\sigma^2}{\sigma_e} (n_1^{-1} - n^{-1}),
\]

\[
\bar{u}_. = n^{-1} \Sigma_{i=1}^t n_i \bar{u}_i.
\]

\[
a_i \text{ and } \theta \text{ are defined by (4.15) and (4.16), respectively, and } \tilde{\omega}(\gamma)
\]

is defined by (4.12). Furthermore,

\[
E \left\{ (\bar{\omega}(\gamma) - \mu_i)^2 + \frac{a_i^2(\tilde{\theta} - \theta)^2}{[1 + (a_i - 1)\theta]^4} (\bar{u}_i. - \bar{u}_.)^2 \right\}
\]
\[
- 2[(1 - \gamma_1)v_1 - \gamma_1 \bar{e}_1] \frac{a_1(\tilde{\theta} - \theta)}{[1 + (a_1 - 1)\theta]^2} (\tilde{u}_1 - \tilde{u}_*)
\]

\[
= E(\hat{\mu}_1 - \mu_1)^2 + 2 n_1^{-1} \sigma^2(1 - \gamma_1)(\sigma^2 + n_1^{-1} \sigma^2)^2
\]

\[
\times V(\tilde{u}_1 - \tilde{u}_*)(v_1 + v_2 - 2)((\sigma^2 + n_1^{-1} \sigma^2)^3 v_1 v_2)^{-1},
\]

where \( v_1 = n - k - t + \lambda \) and \( v_2 = t - \lambda \). (4.19)

**Proof.** Write

\[
\tilde{\mu}_1 - \mu_1 = \tilde{x}_1(p)\hat{\theta} + \gamma_1(\tilde{x}_1. \theta + v_1 + \bar{e}_1 - \tilde{x}_1. \theta)
\]

\[- \tilde{x}_1(p)\theta - v_1
\]

\[
= (\tilde{x}_1(p) - \gamma_1 \tilde{x}_1.)(\hat{\theta} - \theta) - (1 - \gamma_1)v_1 + \gamma_1 \bar{e}_1.
\]

\[
= (\tilde{x}_1(p) - \gamma_1 \tilde{x}_1.)(\hat{\theta} - \theta) - (\hat{\gamma}_1 - \gamma_1)\tilde{x}_1. (\hat{\theta} - \theta)
\]

\[- (1 - \gamma_1)v_1 + (\hat{\gamma}_1 - \gamma_1)(v_1 + \bar{e}_1.)
\]

\[+ \gamma_1 \bar{e}_1.
\]

\[
= (\tilde{x}_1(p) - \gamma_1 \tilde{x}_1.)(\hat{\theta} - \theta) + (\tilde{x}_1(p) - \gamma_1 \tilde{x}_1.)(\hat{\theta} - \tilde{\theta})
\]
As stated in (4.11) and proved by Theorem 3 of Fuller and Battese (1973), $\hat{\beta}$ has the same asymptotic distribution as $\tilde{\beta}$,

$$\hat{\beta} - \beta = O_p\left(n^{-1/2}\right),$$

and

$$\hat{\beta} - \tilde{\beta} = O_p\left(n^{-1/2} \max\left[t^{-1/2}, (n-t)^{-1/2}\right]\right)$$

$$= O_p\left(\max\left[t^{-1}, (n-t)^{-1}\right]\right).$$

Expanding $\hat{\gamma}_1$ in a Taylor series gives
\[ \gamma_i = \frac{1-\theta}{1+(a_i-1)\theta} - \frac{a_i}{[1+(a_i-1)\theta]^2} (\hat{\theta} - \theta) + \frac{2a_i(a_i-1)}{[1+(a_i-1)\theta]^4} (\hat{\theta} - \theta)^2 \]

+ remainder.

Expanding \( \hat{\theta} \) as a function of \( \hat{\sigma}_e^2 \) and MSA gives

\[
\hat{\theta} = \frac{t - \lambda - 2}{t - \lambda} \theta + \frac{t - \lambda - 2}{t - \lambda} \frac{\hat{\sigma}_e^2 - \sigma_e^2}{n_* \sigma_v^2 + \sigma_e^2} \\
- \frac{t - \lambda - 2}{t - \lambda} \frac{\sigma_e^2(MSA - n_* \sigma_v^2 - \sigma_e^2)}{(n_* \sigma_v^2 + \sigma_e^2)^2} \\
+ \frac{t - \lambda - 2}{t - \lambda} \frac{\sigma_e^2(MSA - n_* \sigma_v^2 - \sigma_e^2)^2}{(n_* \sigma_v^2 + \sigma_e^2)^3} \\
- \frac{t - \lambda - 2}{t - \lambda} \frac{(\sigma_e^2 - \sigma_e^2)(MSA - n_* \sigma_v^2 - \sigma_e^2)}{(n_* \sigma_v^2 + \sigma_e^2)^2} \\
+ \text{remainder},
\]

where \( n_* \) is defined in (4.3). It follows that
\[ \hat{\theta} - \theta = O_p(\max[t^{-1/2}, (n-t)^{-1/2}]) . \]  

Therefore,

\[ \hat{\gamma}_1 - \gamma_1 = O_p(\max[t^{-1/2}, (n-t)^{-1/2}]) . \]  

Since

\[ \bar{u}_{..} = n^{-1} \sum_{i=1}^{n} u_{1i} \sim N\left(0, \sigma^2 \frac{1}{n} + \frac{\sigma^2}{n} \right), \]

it follows from assumption (4) that

\[ \bar{u}_{..} = O_p(n^{-1/2}) . \]

Using the results of (4.11), (4.21), and (4.22),

\[ (\hat{\mu}_i - \mu_i) = (\hat{\mu}_i - \mu_i) + (\hat{\gamma}_1 - \gamma_1)(\bar{u}_{1i} - \bar{u}_{..}) + O_p(\max[t^{-1}, (n-t)^{-1}]) . \]

Then,

\[ (\hat{\mu}_i - \mu_i)^2 = (\hat{\mu}_i - \mu_i)^2 + (\hat{\gamma}_1 - \gamma_1)^2(\bar{u}_{1i} - \bar{u}_{..})^2 \]
$$- 2[(1 - \gamma_i)\nu_1 - \gamma_i \bar{e}_1, \nu_1 - \bar{e}_1,] (\nu_1 - \gamma_i)(\bar{u}_1, - \bar{u}_1)$$

$$+ o_p[\max[t^{-1}, (n-t)^{-1}]]$$

$$= (\nu_1 - \mu_1)^2 + \frac{a_1^2(\hat{\theta} - \theta)^2}{[1 + (a_1 - 1)\theta]^h} (\bar{u}_1, - \bar{u}_1)^2$$

$$+ 2[(1 - \gamma_i)\nu_1 - \gamma_i \bar{e}_1, \nu_1 - \gamma_i \bar{e}_1,] \frac{a_1(\hat{\theta} - \theta)}{[1 + (a_1 - 1)\theta]^2} (\bar{u}_1, - \bar{u}_1)$$

$$+ o_p[\max[t^{-1}, (n-t)^{-1}]] .$$

(4.23)

Now $\hat{\theta}$ is a function of MSA, but an explicit expression for MSA is not available. Here the MSA will be replaced by an approximation that is a function of chi-square random variables. By definition,

$$MSA = (t - \lambda)^{-1} \frac{\sum_{i=1}^{n_1} \sum_{j=1}^{1} u_{ij}^2 - \sum_{i=1}^{n_1} \sum_{j=1}^{1} e_{ij}^2}{\sum_{i=1}^{n_1} \sum_{j=1}^{1} u_{ij} - \sum_{i=1}^{n_1} \sum_{j=1}^{1} e_{ij}} .$$

Note that

$$\hat{u}_{1j} = u_{1j} - \bar{e}_{1j}(\hat{\bar{e}} - \bar{e})$$

$$= u_{1j} + o_p(n^{-1/2}) ,$$
where \( \hat{\beta} \) denotes the ordinary least squares estimator from the regression of \( y \) on \( X \). Then,

\[
(t - \lambda)^{-1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} \hat{u}_{ij}^2 = (t - \lambda)^{-1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} [u_{ij}^2 - 2 \hat{u}_{ij} (\hat{\beta} - \beta) + (\hat{x}_{ij} (\hat{\beta} - \beta))^2].
\]

Since \((\hat{\beta} - \beta)^2 = O_p(n^{-1})\), it follows that

\[
(t - \lambda)^{-1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} [\hat{x}_{ij} (\hat{\beta} - \beta)]^2 = O_p(n^{-1}) .
\]

Also, \( u_{ij} \hat{x}_{ij} \) has mean \( \beta' \) and a bounded covariance matrix, so

\[
(t - \lambda)^{-1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} u_{ij} \hat{x}_{ij} = O_p(t^{-1/2}) ,
\]

and

\[
(t - \lambda)^{-1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} u_{ij} \hat{x}_{ij} (\hat{\beta} - \beta) = O_p(n^{-1/2} t^{-1/2}) .
\]

Therefore,

\[
(t - \lambda)^{-1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} \hat{u}_{ij}^2 = (t - \lambda)^{-1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} u_{ij}^2 + O_p(n^{-1/2} t^{-1/2}) .
\]
Similarly,

\[
(t - \lambda)^{-1} \sum_{i=1}^{t} \sum_{j=1}^{n_i} \frac{e^2_{ij}}{e_{ij} - e_{i*}} = (t - \lambda)^{-1} \sum_{i=1}^{t} \sum_{j=1}^{n_i} \left( e_{ij} - e_{i*} \right)^2 \\
+ O_p \left( n^{-1/2} t^{-1/2} \right).
\]

Therefore,

\[
\text{MSA} = (t - \lambda)^{-1} \left[ \sum_{i=1}^{t} \sum_{j=1}^{n_i} u^2_{ij} - \sum_{i=1}^{t} \sum_{j=1}^{n_i} \left( e_{ij} - e_{i*} \right)^2 \right] \\
+ O_p \left( n^{-1/2} t^{-1/2} \right)
\]

\[
= (t - \lambda)^{-1} \sum_{i=1}^{t} \sum_{j=1}^{n_i} [(v^2_i + 2v_i e_{ij} - e^2_{ij}) - (e^2_{ij} - 2e_{ij} e_{i*}) - e_{i*}^2] + O_p \left( n^{-1/2} t^{-1/2} \right)
\]

\[
= (t - \lambda)^{-1} \sum_{i=1}^{t} (n_i v^2_i - 2n_i v_i e_{i*} + n_i e_{i*}^2) + O_p \left( n^{-1/2} t^{-1/2} \right)
\]

\[
= (t - \lambda)^{-1} \sum_{i=1}^{t} n_i \overline{u}_{i*}^2 + O_p \left( n^{-1/2} t^{-1/2} \right).
\]

Next, since \( \overline{u}_{i*} = O_p(n^{-1/2}) \) and

\[
(t - \lambda)^{-1} \sum_{i=1}^{t} n_i \overline{u}_{i*} = O_p(t^{-1/2}),
\]
it follows that

\[
(t - \lambda)^{-1} \sum_{i=1}^{t} n_i \overline{u_i} \overline{u_i} = O_p(n^{-1/2} t^{-1/2})
\]

and

\[
(t - \lambda)^{-1} \sum_{i=1}^{t} n_i \overline{u_i} = O_p(t^{-1}) + O_p(t^{-1}).
\]

Therefore,

\[
\text{MSA} = (t - \lambda)^{-1} \sum_{i=1}^{t} n_i (\overline{u_i} - \overline{u_i})^2 + O_p(t^{-1}).
\]

Now consider the difference

\[
(t - \lambda)^{-1} \sum_{i=1}^{t} n_i \overline{u_i} \overline{u_i} - (t - \lambda)^{-1} \sum_{i=1}^{t} n_i \overline{u_i} (n_i V(\overline{u_i}) - \overline{u_i}) (n_i \sigma^2_v + \sigma^2_e)
\]

\[
= (t - \lambda)^{-1} \sum_{i=1}^{t} \frac{\overline{u_i} \overline{u_i}}{V(\overline{u_i})} \frac{V(\overline{u_i})}{V(\overline{u_i} - \overline{u_i})} \left[ n_i V(\overline{u_i} - \overline{u_i}) \right] - (n_i \sigma^2_v + \sigma^2_e),
\]

(4.24)
where \( V(\bar{u}_i - \bar{u}) \) is defined in (4.18). The quantities 
\( \bar{u}_i^2 V(\bar{u}_i)^{-1} \) are independent chi-square random variables. The 
multiplier of the \( i \)-th chi-square variable in (4.24) is

\[
c_i = \frac{c_i (\sigma_v^2 + n_i^{-1} \sigma_e^2)}{V(\bar{u}_i - \bar{u})}
\]

where

\[
c_i = n_i V(\bar{u}_i - \bar{u}) - (n_i \sigma_v^2 + \sigma_e^2).
\]  

(4.25)

Then,

\[
V \left\{ (t - \lambda)^{-1} \sum_{i=1}^{t} \frac{\bar{u}_i^2}{V(\bar{u}_i)} \right\}
\]

\[
= 2(t - \lambda)^{-2} \sum_{i=1}^{t} \frac{c_i^2 (\sigma_v^2 + n_i^{-1} \sigma_e^2)^2}{(V(\bar{u}_i - \bar{u}))^2}
\]

\[
< 2(t - \lambda)^{-1} \sigma_v^{-4} (\sigma_v^2 + \sigma_e^2)^2 \sum_{i=1}^{t} c_i^2.
\]  

(4.25a)

The last inequality of (4.25a) follows from the fact that

\[
\sigma_v^2 + n_i^{-1} \sigma_e^2 < \sigma_v^2 + \sigma_e^2
\]

(4.25b)
and the fact that

\[ V(\bar{u}_{i}, - \bar{u}_{i}) > 3^{-1} \sigma^2 \]  

(4.25c)

for large \( t \). Inequality (4.25c) is verified for large \( t \) using assumption (4) and (4.17a). First, suppose \( n_1 > 3^{-1} n \). Then,

\[
\begin{align*}
\sum_{i=1}^{t} (n_i - \bar{n})^2 &= \sum_{i=1}^{t} (n_i^2 - t \bar{n}^2) \\
&= n_1^2 \sum_{j \neq i} (n_j^2 - t^{-1} \bar{n}^2) \\
&> n^2(3^{-2} - t^{-1}) \\
&= O(n^2).
\end{align*}
\]

But

\[
\begin{align*}
\sum_{i=1}^{t} (n_i - n_\star)^2 &= O(1),
\end{align*}
\]

and

\[
\begin{align*}
\sum_{i=1}^{t} (n_i - n_\star)^2 > \sum_{i=1}^{t} (n_i - \bar{n})^2.
\end{align*}
\]

because
\[ \bar{n} = t^{-1} \sum_{i=1}^{t} n_i \]

minimizes the sum of squared deviations. Therefore, there exists an \( M \) such that for \( t > M \), \( n_i < 3^{-1} n \) for all \( i = 1, 2, \ldots, t \).

Therefore,

\[ \text{Var}(\bar{u}_i - \bar{u} \ldots) = \sigma_v^2 \left( 1 - 2n^{-1} N_i + n^{-2} \sum_{i=1}^{t} n_i^2 \right) \]

\[ + \sigma_e^2 (n_i^{-1} - n^{-1}) \]

\[ > \sigma_v^2 (1 - 2n^{-1} N_i) \]

\[ > 3^{-1} \sigma_v^2 , \]

and (4.25c) is verified.

It is now shown that \( \sum_{i=1}^{t} c_i^2 = O(1) \). Write

\[ \sum_{i=1}^{t} c_i^2 = \sum_{i=1}^{t} \left[ n_i \text{Var}(\bar{u}_i - \bar{u} \ldots) - (n_i \sigma_v^2 + \sigma_e^2)^2 \right] \]

\[ = \sum_{i=1}^{t} \left[ \sigma_v^2 (n_i - 2n^{-1} N_i + n^{-2} \sum_{i=1}^{t} N_i^2) + \sigma_e^2 (1 - n^{-1} N_i) \right. \]

\[ \left. - (n_i \sigma_v^2 + \sigma_e^2)^2 \right] \]

\[ = \sum_{i=1}^{t} \left[ \sigma_v^2 (n_i - n_i - 2n^{-1} N_i + n^{-2} \sum_{i=1}^{t} N_i^2) - \sigma_v^2 n^{-1} N_i \right]^2 \]
\[\begin{align*}
&= \sigma^4 \left( \sum_{i=1}^t (n_i^2 - n_\star^2) + 4 n^{-2} \sum_{i=1}^t n_i^4 + n^{-4} \left( \sum_{i=1}^t n_i^2 \right)^2 \\
&\quad - 4 n^{-1} \sum_{i=1}^t n_i^2 (n_i - n_\star) - 4 n^{-3} \left( \sum_{i=1}^t n_i^3 \right) (\sum_{i=1}^t n_i^2) \\
&\quad - 2 n^{-2} \sum_{i=1}^t n_i (n_i - n_\star) (\sum_{i=1}^t n_i^2) + \sigma^2 e^{-n^{-2} \sum_{i=1}^t n_i^2} \\
&\quad - 2 \sigma^2 \sigma^2 [n^{-1} \sum_{i=1}^t n_i (n_i - n_\star) - 2 n^{-2} \sum_{i=1}^t n_i^3] \\
&\quad + n^{-3} (\sum_{i=1}^t n_i^2)^2 \right) \\
&= \sigma^4 [n^{-2} \sum_{i=1}^t n_i^4 - 4 n^{-1} \sum_{i=1}^t n_i^2 (n_i - n_\star) - 4 n^{-2} \sum_{i=1}^t n_i^3 \\
&\quad + 2 n^{-1} \sum_{i=1}^t n_i (n_i - n_\star)] - 2 \sigma^2 \sigma^2 [n^{-1} \sum_{i=1}^t n_i (n_i - n_\star) \\
&\quad - 2 n^{-2} \sum_{i=1}^t n_i^3] + O(1)
\end{align*}\]

by assumption (4). Now

\[\begin{align*}
n^{-2} \sum_{i=1}^t n_i^3 &< n^{-2} \left( \sum_{i=1}^t n_i^2 \right) \left( \sum_{i=1}^t n_i \right) \\
&= n^{-1} \sum_{i=1}^t n_i^2 \\
&= O(1)
\end{align*}\]
Similarly,

\[
\sum_{i=1}^{t} n_i^4 < \sum_{i=1}^{t} (n_i^2)^2
\]

\[
= O(1) .
\]

Also,

\[
\left| \sum_{i=1}^{t} n_i^{-1} (n_i - n^*) \right| < \sum_{i=1}^{t} \max[n_i^{-2}, (n_i - n^*)^2]
\]

\[
< \sum_{i=1}^{t} n_i^{-2} n_i^4 + \sum_{i=1}^{t} (n_i - n^*)^2
\]

\[
= O(1) .
\]

Similarly,

\[
\left| \sum_{i=1}^{t} n_i^{-1} (n_i - n^*) \right| < \sum_{i=1}^{t} n_i^{-2} n_i^2 + \sum_{i=1}^{t} (n_i - n^*)^2
\]

\[
= O(1) .
\]

Therefore,

\[
\sum_{i=1}^{t} c_i^2 = O(1)
\]

(4.25d)
by assumption (4), and

\[ 2(t - \lambda)^{-2} \sigma_v^{-2} (\sigma_v^2 + \sigma_e^2) \sum_{i=1}^{t} c_i^2 = O(t^{-2}) . \]

Again using assumption (4), it can be shown that the expectation of (4.24) is \( O(t^{-1}) \). Then, by Corollary 5.1.1.2 of Fuller (1976), it follows that the difference of (4.24) is

\[ O_p(t^{-1}) . \]  

(4.25e)

Next, using the results of (4.25b) and (4.25c),

\[
\sqrt{V \left\{ (t - \lambda)^{-1} \sum_{i=1}^{t} \frac{c_i u_i}{\sqrt{V(\bar{u}_i - \bar{u}_.)}} \right\}}
\]

\[
< (t - \lambda)^{-2} 9 \sigma_v^{-4} (\sigma_v^2 + \sigma_e^2) \sum_{i=1}^{t} c_i^2
\]

\[ = O(t^{-2}) . \]

Therefore,

\[
(t - \lambda)^{-1} \sum_{i=1}^{t} \frac{c_i u_i}{\sqrt{V(\bar{u}_i - \bar{u}_.)}} = O_p(t^{-1}) ,
\]
and

\[ 2(t - \lambda)^{-1} \sum_{i=1}^{t} \frac{c_i \bar{u}_i \bar{u}_i}{\sqrt{\sum_{i=1}^{t} \bar{u}_i - \bar{u}_i}} = O(p^{-1/2} t^{-1}). \quad (4.25f) \]

Also,

\[ \sqrt{V(\bar{u}_i^2)} \sigma^{-4}(t - \lambda)^{-2} \sum_{i=1}^{t} \frac{c_i^2}{\sqrt{\sum_{i=1}^{t} \bar{u}_i - \bar{u}_i}} \]

\[ = O(t^{-2}), \]

so

\[ (t - \lambda)^{-1} \sum_{i=1}^{t} \frac{c_i \bar{u}_i^2}{\sqrt{\sum_{i=1}^{t} \bar{u}_i - \bar{u}_i}} = O(t^{-1}). \quad (4.25g) \]

Combining the results of (4.25e), (4.25f) and (4.25g), it follows that
Let

\[ \cdot (I - c)^d_0 = \]

\[ (\varepsilon^0 + \Lambda^0 \xi u)_{I-} (\{ n - \Gamma n \} \Lambda I =^T) \]

\[ \cdot \frac{\{ n - \Gamma n \} _{I =^T} I =^T}{z (n - \Gamma n)_{I =^T} I =^T (\gamma - c)} = \text{MSA} \]

and let

\[ \cdot \frac{\{ n - \Gamma n \} _{I =^T}}{z (n - \Gamma n)_{I =^T} I =^T (\gamma - c)} = \text{MSA} \]
Since \( \tilde{\theta} \) is a continuous function of \( \text{MSA} > 0 \), then by Theorem 5.1.4 of Fuller (1976),

\[
\hat{\theta} - \tilde{\theta} = O_p(t^{-1}).
\]

From (4.20),

\[
\tilde{\theta} - \theta = O_p(\max\{t^{-1/2}, (n-t)^{-1/2}\}).
\]

Therefore, equation (4.23) becomes

\[
(\hat{\mu}(\gamma) - \mu) = (\tilde{\mu}(\gamma) - \mu) + \frac{a_1^2(\tilde{\theta} - \theta)^2}{[1 + (a_1 - 1)\theta]^4} (\tilde{u}_1. - \bar{u}.)^2
\]

\[+ 2[(1 - \gamma_1)\nu_1 - \gamma_1 e_1.] \frac{a_1(\tilde{\theta} - \theta)}{[1 + (a_1 - 1)\theta]^2} (\tilde{u}_1. - \bar{u}.)
\]

\[+ O_p(\max\{t^{-1}, (n-t)^{-1}\}) .
\]

(4.28)

The quantities \( \hat{\sigma}_e^2 \) and \( \text{MSA} \) are now shown to be independent. Recall that \( \hat{\sigma}_e^2 \) is a multiple of \( \hat{e}_e^2 \), and
\[ \hat{\varepsilon}'\hat{\varepsilon} = \chi'N\chi, \]

where

\[ M = N - NX(X'NX)^{-1}X'N \]

and

\[ N = \text{block diag}(I_{n_1} - n_1^{-1}J_{n_1}, \ldots, J_{n_1}, \ldots, n_1) \cdot \]

Then,

\[ \hat{\varepsilon}'\hat{\varepsilon} = (X\bar{\xi} + \mu)'M(X\bar{\xi} + \mu) \]

\[ = \mu'M\mu, \]

since \( N \) is an idempotent matrix, and

\[ [N - N \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \times \time
\[ Q = \text{block diag}\{n^{-1}_1 J_1, \ldots\} \]

Then, \( \hat{\varepsilon}'\hat{\varepsilon} \) and \( \bar{u}_i \) are independent if

\[ M V Q' = N - N X(X' M X)^{-1} X' M V Q' = 0. \]

Now

\[ N V Q' = \text{block diag}\{(X_{n_1} - n^{-1}_1 J_{n_1} x n_1)\sigma^2 e_{n_1} + \sigma^2 J_{n_1} x n_1\} \]

\[ = \text{block diag}\{(\sigma^2 I - n^{-1}_1 \sigma^2 J + \sigma^2 J - \sigma^2 J)n^{-1}_1 J\} \]

\[ = \text{block diag}\{n^{-1}_1 \sigma^2 J - n^{-1}_1 \sigma^2 J\} \]

\[ = 0. \]

Therefore, \( \hat{\sigma}^2_e \) and MSA are independent.

Consider the expectation
The expectation of the middle term of (4.29) is

\[
E \left\{ \frac{a_i^2(\bar{\theta} - \theta)^2}{[1 + (a_i - 1)\theta]^4} (\bar{u}_{i.} - \bar{u}..)^2 \right. \\
+ 2[(1 - \gamma_i)\nu_{i1} - \gamma_i \bar{e}_{i1.}] \frac{a_i(\bar{\theta} - \theta)}{[1 + (a_i - 1)\theta]^2} (\bar{u}_{i.} - \bar{u}..) \left. \right\}.
\]

(4.29)

Using the independence of \( \bar{\sigma}_e^2 \) and \( \bar{u}_{i.} \), \( i = 1, 2, \ldots, t \),

\[
E\{[(1 - \gamma_i)\nu_{i1} - \gamma_i \bar{e}_{i1.}] (\bar{\theta} - \theta)(\bar{u}_{i.} - \bar{u}..)\} = E\{E\{[(1 - \gamma_i)\nu_{i1} - \gamma_i \bar{e}_{i1.}] (\bar{\theta} - \theta)(\bar{u}_{i.} - \bar{u}..) | \bar{u}_{i.}, i = 1, \ldots, t)\}\}
\]

\[
= E\{[(1 - \gamma_i)\nu_{i1} - \gamma_i \bar{e}_{i1.}] (\bar{\theta} - \theta)(\bar{u}_{i.} - \bar{u}..) | \bar{u}_{i.}, i = 1, \ldots, t)\}
\]

\[
= 0 . \quad (4.30)
\]

The expectation of the middle term of (4.29) is

\[
E \left\{ \frac{a_i^2(\bar{\theta} - \theta)^2}{[1 + (a_i - 1)\theta]^4} (\bar{u}_{i.} - \bar{u}..)^2 \right. \\
\left. \right\}.
\]
Without loss of generality, let \( i = 1 \).

The quantities \((\bar{u}_{1*} - \bar{u}_{..})(\sqrt{\nu_{1*} - \bar{u}_{..}})^{-1/2}, i = 1, \ldots, t\), are standard normal random variables, but they are not independent. There is a triangular transformation of the vector with elements

\[(\bar{u}_{1*} - \bar{u}_{..})(\sqrt{\nu_{1*} - \bar{u}_{..}})^{-1/2}
\]
to a vector \( \mathbf{w} \) such that the elements of \( \mathbf{w} \) are independent, \( w_i = (\bar{u}_{1*} - \bar{u}_{..})(\sqrt{\nu_{1*} - \bar{u}_{..}})^{-1/2} \), and the first \((t - \lambda)\) elements of \( \mathbf{w} \) are standard normal random variables.

Furthermore, the transformation is such that

\[
\frac{t}{\Sigma} \frac{(\bar{u}_{1*} - \bar{u}_{..})^2}{\sqrt{\nu_{1*} - \bar{u}_{..}}} = \Sigma w_j^2, \quad j = 1
\]

so

\[
\tilde{MSA} = (n_x \sigma^2_y + \sigma^2_e)(t - \lambda)^{-1} \Sigma w_j^2.
\]

Note that

\[
w_1^2 \sim \chi^2_1,
\]
\[
\sum_{t=1}^{t-\lambda} w_j^2 \sim \chi^2_{t-\lambda},
\]

and

\[
\sum_{j=2}^{t-\lambda} w_j^2 \sim \chi^2_{t-\lambda-1}.
\]

Then,

\[
E \left\{ \left( \frac{\tilde{\theta} - \theta}{\theta} \right)^2 \left( \tilde{v}_{11} - \tilde{u}_{11} \right)^2 \left( \tilde{v}(\tilde{u}_{11} - \tilde{u}_{11}) \right)^{-1/2} \right\}
\]

\[
= E \left\{ \left( \frac{\tilde{\theta} - \theta}{\theta} \right)^2 w_1^2 \right\}
\]

\[
= E(\theta^{-2} \tilde{\theta}^2 w_1^2) - 2E(\theta^{-1} \tilde{\theta} w_1^2) + E(w_1^2).
\] (4.33)

Consider the individual terms in expression (4.33). Clearly the third term on the right of (4.33) is

\[E(w_1^2) = 1.\]

The second term on the right of (4.33) is
Because $\tilde{\sigma}^2_{e}$ is independent of the elements of \(\w\),

\[
E(\theta^{-1} \tilde{\theta} w_1^2) = \left( t-\lambda-2 \right) \frac{E\left( \tilde{\theta}^2 \right)}{E\left( \tilde{\theta}^2_{t-\lambda-2} \right)} \left( \sum_{j=1}^{t-\lambda} w_j^2 \right)^{-1} \left( w_1^2 \right) . \tag{4.34}
\]

Because $\tilde{\sigma}^2_{e}$ is independent of the elements of \(\w\),

\[
E(\theta^{-1} \tilde{\theta} w_1^2) = (t-\lambda-2)E\left( \sum_{j=1}^{t-\lambda} \left( w_j^2 \right)^{-1} \right) w_1^2
\]

\[
= (t-\lambda-2) E\left( \left( U_1 + U_2 \right)^{-1} U_1 \right) ,
\]

where $U_1 \sim \chi^2_{t-\lambda-1}$ and $U_2 \sim \chi^2_{t-\lambda-1}$ independently of $U_1$. Therefore, using Lemma 4.1,

\[
E(\theta^{-1} \tilde{\theta} w^2) = \frac{(t-\lambda-2) \Gamma\left( \frac{t-\lambda}{2} \right) \Gamma\left( \frac{t-\lambda}{2} + 1 \right)}{\Gamma\left( \frac{1}{2} \right) \Gamma\left( \frac{1}{2} \right) \Gamma\left( \frac{t-\lambda}{2} + 1 \right)}
\]

\[
= \frac{(t-\lambda-2) \frac{1}{2} \Gamma\left( \frac{1}{2} \right) \Gamma\left( \frac{t-\lambda}{2} \right)}{\Gamma\left( \frac{1}{2} \right) \Gamma\left( \frac{t-\lambda}{2} \right) \Gamma\left( \frac{t-\lambda}{2} \right)}
\]

\[
= (t-\lambda)^{-1} (t-\lambda-2)
\]
Similarly, the first term on the right of (4.33) is

\[
E\{(\theta^{-1} \hat{\theta})^2 \omega_1^2\} = \sigma_e^{-4} \ E\{g^4\} (t-\lambda-2)^2 \ E\{(U_1 + U_2)^{-2} U_1\} \\
= \sigma_e^{-4} \ (V[\sigma^2_e] + \sigma^4_e) (t-\lambda-2)^2 \ \frac{\Gamma\left(\frac{3}{2}\right) \Gamma\left(\frac{t-\lambda}{2} - 1\right)}{\left(\frac{1}{2}\right)^{-1} \ \Gamma\left(\frac{1}{2}\right) \Gamma\left(\frac{t-\lambda}{2} + 1\right)} \\
= \sigma_e^{-4} \ \left( \frac{2\sigma^4_e}{\nu_1} + \sigma^4_e \right) \ \frac{(t-\lambda-2)^2}{4(t-\lambda)(t-\lambda-2)} \\
= \left( \frac{2}{\nu_1} + 1 \right) \ \frac{(\nu_2 - 2)}{\nu_2}. \quad (4.36)
\]

Therefore, expression (4.33) becomes

\[
E\left\{ \left( \frac{\hat{\theta} - \bar{\theta}}{\theta} \right)^2 \ \frac{(\hat{U}_1 - \bar{U}_2)^2}{V(\hat{U}_1 - \bar{U}_2)} \right\} = \frac{(2+\nu_1)(\nu_2-2)}{\nu_1 \nu_2} - \frac{2(\nu_2-2)}{\nu_2} + 1 \\
= \frac{\nu_1 \nu_2 + 2\nu_2 - 2\nu_1 - 4 + 2\nu_1 \nu_2 + 4\nu_1 + \nu_1 \nu_2}{\nu_1 \nu_2} \\
= \frac{2(\nu_1 + \nu_2 - 2)}{\nu_1 \nu_2}. \quad (4.37)
\]
The constant multiplier of (4.37) in expression (4.31) is

\[
\frac{a_1^2 \theta^2 \sigma^2_e}{[1 + (a_i - 1)\theta]^4} = \frac{a_1^2 \frac{\sigma^4_e}{(n_\star \sigma^2_v + \sigma^2_e)^2} \psi(\bar{u}_{i, -} - \bar{u}_{..})}{\left[1 + (a_i - 1) \frac{\sigma^2_e}{(n_\star \sigma^2_v + \sigma^2_e)}\right]^4}
\]

\[
= \frac{n_i^2 (n_i - 1) \sigma^2_v (n_i \sigma^2_v + \sigma^2_e)^2 \psi(\bar{u}_{i, -} - \bar{u}_{..})}{n_i^4 (\sigma^2_v + n_i^{-1} \sigma^2_e)^4}
\]

\[
= \frac{n_i^{-1} \sigma^2_e (1 - \gamma_i) (\sigma^2_v + n_i^{-1} \sigma^2_e)^2 \psi(\bar{u}_{i, -} - \bar{u}_{..})}{(\sigma^2_v + n_i^{-1} \sigma^2_e)^3}
\]

Therefore,

\[
E\left(\frac{a_1^2 (\bar{u} - \bar{v})^2}{[1 + (a_i - 1)\theta]^4} (\bar{u}_{i, -} - \bar{u}_{..})^2\right)
\]

\[
= \frac{2 n_i^{-1} \sigma^2_e (1 - \gamma_i) (\sigma^2_v + n_i^{-1} \sigma^2_e)^2 \psi(\bar{u}_{i, -} - \bar{u}_{..})(v_1 + v_2 - 2)}{(\sigma^2_v + n_i^{-1} \sigma^2_e)^3 v_1 v_2}
\]

(4.38)
Recall from (4.18) that

\[ V(u_i - \bar{u}) = \alpha^2 (1 - 2n_i^{-1} n + n^{-2} \sum_{i=1}^{t} n_i^2) + o^2(n_i^{-1} - n^{-1}) \, . \]

Substituting the results of (4.30) and (4.38) into (4.29) gives the final result

\[
E\left(\left(\hat{\mu}_i - \mu_i\right)^2 + \frac{a_i^2(\hat{\theta} - \theta)}{[1 + (a_i - 1)\theta]^4} \, (\bar{u}_{i.} - \bar{u}_{..})^2 \right)
+ 2[(1 - \gamma_i)\bar{v}_{i.} - \gamma_i\bar{e}_{i.}] \, \frac{a_i(\hat{\theta} - \theta)}{[1 + (a_i - 1)\theta]^4} \, (\bar{u}_{i.} - \bar{u}_{..})
= E\{\hat{\mu}_i - \mu_i\}^2
+ 2 \left[ n_i^{-1} \alpha^2 (1 - \gamma_i) (\sigma^2 + n_i^{-1} \sigma_e^2) \right] V(\bar{u}_{i.} - \bar{u}_{..})
(\bar{v}_1 + \bar{v}_2 - 2) [(\sigma^2 + n_i^{-1} \sigma_e^2)^3 \bar{v}_1 \bar{v}_2]^{-1} \, . \]

If it is assumed that the \( n_i \)'s are bounded and their average is greater than 1, then all the order statements can be simplified to involve only \( t \) . (Replace \( \delta \) or \( n-t \) by \( t \) .)

For the special case in which all sample sizes are equal \( (n_i = r, i = 1, \ldots, t) \) and only a scalar parameter is being estimated
(\(x_{ij} = 1\) for all \(i, j\)), Peixoto (1982) has given an exact expression for
\(\text{MSE}\{\hat{\mu}_i^{(\gamma)}\}\). Peixoto's result is

\[
E\{(\hat{\mu}_i^{(\gamma)} - \mu_i)^2\} = E\{(\hat{\mu}_i^{(\gamma)} - \mu_i)^2\} + 2 r^{-1} \sigma_e^2 (1-\gamma) \frac{rt-3}{t^2(t-1)},
\]

where

\[
\gamma_i = \gamma = \frac{\sigma_v^2}{\sigma_v^2 + r^{-1} \sigma_e^2}, \quad i = 1, \ldots, t .
\]

It can be verified that in this special case the penalty term added to \(\text{MSE}\{\hat{\mu}_i^{(\gamma)}\}\) given in Theorem 4.1 is identical to the increase given by Peixoto. In the special case of \(n_i = r\) and \(x_{ij} \equiv 1\),

\[
V\{\bar{\mu}_i, \bar{\mu}_j\} = (\sigma_v^2 + r^{-1} \sigma_e^2) \frac{(rt-1)}{rt} + \sigma_v^2 \left( \frac{tr^2}{(rt)^2} - \frac{r}{rt} \right)
\]

\[
= (\sigma_v^2 + r^{-1} \sigma_e^2) t^{-1}(t-1),
\]

and

\[
n_x = \sum_{i=1}^{t} \frac{r(1 - \frac{r}{rt})}{t-1} = \frac{tr^2-r}{rt} \frac{rt-r}{rt} = r(t-1) = r.
\]
For $n = r$ and $x_{ij} = 1$, the penalty term given in Theorem 4.1 reduces to

$$2 \frac{r^{-1} \sigma^2(1-\gamma) (rt-3)}{t^2(r-1)},$$

which matches Peixoto's result.

C. Estimation with Prior Estimates

Sometimes prior values of the variance components are available. In the crop estimation problem, for example, estimated variance components from the previous year or from neighboring areas can be used as prior information. Under fairly general conditions, it is possible to construct estimators of the variance components by combining the sample and the prior estimates. First, the general theory will be presented, followed by a presentation of the formulas for three particular cases of prior information for the nested-error model.

1. General theory

Durbin (1953) suggested a method of combining sample and prior data by generalized least squares. Let $\mathbf{a}_1$ be an $(h \times 1)$ vector of unbiased (prior) estimates for $\mathbf{a}$, where $\mathbf{a}$ is an $(h \times 1)$ vector of unknown parameters. Let $\mathbf{a}_2$ be an independent $(k \times 1)$ vector of unbiased (sample) estimates for $\mathbf{a}$, where $\mathbf{a}$ is a $(k \times 1)$ vector of parameters in which the first $h$ components are $\mathbf{a}_1$. Assume that
\( V(\hat{\alpha}_1) \) and \( V(\hat{\alpha}_2) \) are known and of full rank. The estimators can be expressed as
\[
\begin{pmatrix}
\hat{\alpha}_1 \\
\hat{\alpha}_2
\end{pmatrix} = \begin{pmatrix}
I_h & 0 \\
I_k & 0
\end{pmatrix} \alpha + \begin{pmatrix}
\xi_1 \\
\xi_2
\end{pmatrix},
\]
where \( \xi_1 \) and \( \xi_2 \) are independently distributed with means 0 and covariance matrices \( V(\hat{\alpha}_1) \) and \( V(\hat{\alpha}_2) \), respectively. The best linear (in terms of \( \hat{\alpha}_1 \) and \( \hat{\alpha}_2 \)) unbiased estimator of \( \alpha \) is the generalized least squares estimator
\[
\hat{\alpha} = \left( \begin{pmatrix}
V(\hat{\alpha}_1)^{-1} & 0 \\
0 & V(\hat{\alpha}_2)^{-1}
\end{pmatrix} + V(\hat{\alpha}_2)^{-1} \right)^{-1} \begin{pmatrix}
V(\hat{\alpha}_1)^{-1} \hat{\alpha}_1 \\
0
\end{pmatrix}
\]
\[
+ V(\hat{\alpha}_2)^{-1} \hat{\alpha}_2
\]
\[
= \hat{\alpha}_2 + \left( \begin{pmatrix}
V(\alpha)^{-1} & 0 \\
0 & V(\hat{\alpha}_2)^{-1}
\end{pmatrix} + V(\hat{\alpha}_2)^{-1} \right)^{-1} \begin{pmatrix}
V(\hat{\alpha}_1)^{-1} 0 \\
0 0
\end{pmatrix} \begin{pmatrix}
\hat{\alpha}_1 \\
0
\end{pmatrix} - \hat{\alpha}_2.
\]
\hspace{1cm} (4.40)

The variance of \( \hat{\alpha} \) is
In most applications, \( V(\mathbf{a}_1) \) and \( V(\mathbf{a}_2) \) are not known. Estimates of these covariance matrices can be substituted into the above formulas to estimate the generalized least squares estimator of \( \mathbf{a} \) and its variance.

Generally, the sample data would be expected to yield estimates that are close to the prior values. However, it may be that the prior values are unsuitable. This could happen, for example, if the prior estimates came from a sample taken previously, and the model relationships changed over time. Theil (1963) proposed a test statistic for testing whether the sample and prior estimates are consistent.

Theil assumed that \( V(\mathbf{a}_2) = \sigma^2 I_k \) and that both \( \sigma^2 \) and \( V(\mathbf{a}_1) \) are known. Let \( Q_{(h \times k)} = (I_h | 0) \). The test statistic is

\[
\eta = (\mathbf{a}_1 - Q \mathbf{a}_2)' \left( V(\mathbf{a}_1) + \sigma^2 Q Q' \right)^{-1} (\mathbf{a}_1 - Q \mathbf{a}_2). \tag{4.42}
\]

Theil showed that when \( \mathbf{a}_1 \) and \( \mathbf{a}_2 \) are normally distributed, then \( \eta \) is distributed as a chi-square random variable with \( k \) degrees of freedom.
2. Prior estimates of $\sigma^2_v$ and $\sigma^2_e$

Returning to the nested-error model, suppose prior values $\hat{\sigma}^2_v$ and $\hat{\sigma}^2_e$ exist for $\sigma^2_v$ and $\sigma^2_e$, respectively. Let the covariance matrix

$$
V = \begin{pmatrix}
\hat{\sigma}^2_v \\
\hat{\sigma}^2_e
\end{pmatrix} = \begin{pmatrix}
\hat{\sigma}^2_v \\
\hat{\sigma}^2_e
\end{pmatrix}
$$

also be given. Denote the estimated covariance matrix of the sample estimates by

$$
\hat{V} = \begin{pmatrix}
\hat{\sigma}^2_v \\
\hat{\sigma}^2_e
\end{pmatrix} = \begin{pmatrix}
w_{11} & w_{12} \\
w_{12} & w_{22}
\end{pmatrix}.
$$

By (4.40) with $h = k$, the best estimators of $\sigma^2_v$ and $\sigma^2_e$ are given by

$$
\begin{pmatrix}
\hat{\sigma}^2_v \\
\hat{\sigma}^2_e
\end{pmatrix} = \begin{pmatrix}
\hat{\sigma}^2_v \\
\hat{\sigma}^2_e
\end{pmatrix} + \left(\hat{W}^{-1} + \hat{W}^{-1}\right)^{-1} \hat{W}^{-1} \begin{pmatrix}
\sigma^2_v - \hat{\sigma}^2_v \\
\sigma^2_e - \hat{\sigma}^2_e
\end{pmatrix}
$$

with

$$
\begin{pmatrix}
\sigma^2_v - \hat{\sigma}^2_v \\
\sigma^2_e - \hat{\sigma}^2_e
\end{pmatrix} = \begin{pmatrix}
\sigma^2_v \\
\sigma^2_e
\end{pmatrix} - \begin{pmatrix}
\hat{\sigma}^2_v \\
\hat{\sigma}^2_e
\end{pmatrix}.$$
\[ V \left\{ \begin{array}{c} \tilde{\sigma}_V^2 \\ \tilde{\sigma}_e^2 \end{array} \right\} = (W^{-1} + W^{-1})^{-1}. \]

These formulas are not in the best computing form because of the matrix inverse involved, but they can be written in a form that is easier to compute, though not necessarily neater. Let

\[ S = \begin{pmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{pmatrix} \]

\[ = (W^{-1} + W^{-1})^{-1} \]

\[ = \begin{pmatrix} \frac{w_{11}}{d(W)} + \frac{w_{11}}{d(W)} & \frac{w_{12}}{d(W)} + \frac{w_{12}}{d(W)} \\ \frac{w_{12}}{d(W)} + \frac{w_{12}}{d(W)} & \frac{w_{22}}{d(W)} + \frac{w_{22}}{d(W)} \end{pmatrix} d(W^{-1} + W^{-1})^{-1}, \]

where \( d(.) \) refers to the determinant. So \( d(W) = w_{11}w_{22} - w_{12}^2 \), \( d(W) = w_{11}w_{22} - w_{12}^2 \) and

\[ d(W^{-1} + W^{-1}) = \left( \frac{w_{11}}{d(W)} + \frac{w_{11}}{d(W)} \right) \left( \frac{w_{22}}{d(W)} + \frac{w_{22}}{d(W)} \right) - \left( \frac{w_{12}}{d(W)} + \frac{w_{12}}{d(W)} \right)^2. \]
With this notation (4.43) becomes

\[
\begin{pmatrix}
\tilde{\sigma}_v^2 \\
\tilde{\sigma}_e^2
\end{pmatrix}
= \begin{pmatrix}
\hat{\sigma}_v^2 \\
\hat{\sigma}_e^2
\end{pmatrix}
+ \begin{pmatrix}
\tilde{s}_{11}^{W}_{22} - \tilde{s}_{12}^{W}_{12} & \tilde{s}_{12}^{W}_{11} - \tilde{s}_{11}^{W}_{12} \\
\tilde{s}_{21}^{W}_{22} - \tilde{s}_{22}^{W}_{12} & \tilde{s}_{22}^{W}_{11} - \tilde{s}_{21}^{W}_{12}
\end{pmatrix}
\begin{pmatrix}
\tilde{\sigma}_v^2 - \hat{\sigma}_v^2 \\
\tilde{\sigma}_e^2 - \hat{\sigma}_e^2
\end{pmatrix}
\mathbf{w}^{-1}.
\]

The test statistic for testing the consistency of the sample and prior values is

\[
F_{\alpha}^2 = 2 \left( \begin{pmatrix}
\tilde{\sigma}_v^2 - \hat{\sigma}_v^2 \\
\tilde{\sigma}_e^2 - \hat{\sigma}_e^2
\end{pmatrix}
\right)^\prime \left( \mathbf{w} + \mathbf{w} \right)^{-1}
\left( \begin{pmatrix}
\tilde{\sigma}_v^2 - \hat{\sigma}_v^2 \\
\tilde{\sigma}_e^2 - \hat{\sigma}_e^2
\end{pmatrix}
\right).
\] (4.44)

Although the statistic has the same form as Theil's test statistic, the assumptions of normality and known covariance matrices are not satisfied. Nevertheless, if both the sample and prior estimators are consistent and asymptotically normally distributed, then asymptotically the test statistic (4.44) has the same distribution as Theil's test statistic.

If the prior values are known exactly, i.e., \( \mathbf{w} = 0 \), then the logical estimators are
with covariance matrix $0$. The test statistic (4.44) for consistency is still valid.

3. Prior estimate of $\sigma_v^2$

Suppose a prior estimator of $\sigma_v^2$ and the variance of the prior estimator are available. Denote the prior value by $\sigma_v^2$ and the variance of $\sigma_v^2$ by $w$. Then, the generalized least squares estimator of the variance components is

$$\begin{pmatrix} \hat{\sigma}_v^2 \\ \hat{\sigma}_e^2 \end{pmatrix} = \begin{pmatrix} \hat{\sigma}_v^2 \\ \hat{\sigma}_e^2 \end{pmatrix} + \begin{pmatrix} w^{-1} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \sigma_v^2 - \hat{\sigma}_v^2 \\ \sigma_e^2 - \hat{\sigma}_e^2 \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_v^2 \\ \sigma_e^2 \end{pmatrix} + \begin{pmatrix} w_{11} & (\sigma_v^2 - \hat{\sigma}_v^2) \\ w_{21} & (w + w_{11}) \end{pmatrix} \frac{\sigma_v^2}{w} \cdot \frac{\sigma_e^2}{w} .$$

The covariance matrix is
The suggested test statistic for consistency is

\[ t_\omega = (\hat{w} + w_{11})^{-1/2} \left| \hat{\sigma}_v^2 - \sigma_v^2 \right| . \]  

(4.47)

The square of the statistic has the same form as (4.44) and asymptotically the same form as Theil's test statistic.

Note that (4.45), (4.46), and (4.47) can be used even if \( \hat{w} = 0 \).

The resulting estimator of \( \sigma_v^2 \) will be identically equal to \( \hat{\sigma}_v^2 \).

4. Prior estimate of \( \sigma_v^2/\sigma_e^2 \)

In some instances the prior information may be an estimate of the ratio \( R = \sigma_v^2/\sigma_e^2 \). Let \( \hat{R} \) denote the prior estimator of \( R \) and \( V(\hat{R}) \) the variance of \( \hat{R} \). Let \( \hat{\sigma}_v^2/\hat{\sigma}_e^2 \). The estimator \( \hat{R} \) is biased, but here it will be assumed that the bias is negligible. Once \( \hat{R} \) and \( \hat{\sigma}_e^2 \) are obtained, \( \hat{\sigma}_v^2 \) can be recovered from the product \( \hat{R} \hat{\sigma}_e^2 \).
The formulas in this case are very much like the formulas of Section IV.C.3. The estimators of $R$ and $\sigma^2_e$ are given by

\[
\begin{pmatrix}
\hat{R} \\
\hat{\sigma}^2_e
\end{pmatrix} = \begin{pmatrix}
\hat{R} \\
\hat{\sigma}^2_e
\end{pmatrix} + \begin{pmatrix}
\hat{V}_R \\
\hat{C}[R, \hat{\sigma}^2_e]
\end{pmatrix} \frac{(\hat{R} - \bar{R})}{\hat{V}_R + \hat{V}[R]} ,
\]

where

\[
\hat{V}_R = (\hat{\sigma}^2_e)^{-2}[w_{11} - 2 \bar{R} w_{12} + \bar{R}^2 w_{22}] ,
\]

and

\[
\hat{C}[R, \hat{\sigma}^2_e] = \hat{\sigma}^{-2}_e[w_{12} - \bar{R} w_{22}] .
\]

The estimated covariance matrix of the estimates is

\[
\hat{V} \begin{pmatrix}
\hat{R} \\
\hat{\sigma}^2_e
\end{pmatrix} = \begin{pmatrix}
(\hat{V}_R)^{-1} & 0 \\
0 & 0
\end{pmatrix} + \begin{pmatrix}
\hat{V}_R & \hat{C}[R, \hat{\sigma}^2_e]
\end{pmatrix}^{-1} \begin{pmatrix}
\hat{C}[R, \hat{\sigma}^2_e] & \hat{V}[\hat{\sigma}_e]
\end{pmatrix}^{-1} .
\]

The estimator of $\sigma^2_v$ is
\[ \tilde{\sigma}_v^2 = \tilde{\sigma}_e^2 \]

\[ = R \sigma_e^2 + \sigma_e^2(\tilde{R} - R) + R(\tilde{\sigma}_e^2 - \sigma_e^2) + (\tilde{\sigma}_e^2 - \sigma_e^2)(\tilde{R} - R) \]

\[ = R \sigma_e^2 + \sigma_e^2(\tilde{R} - R) + R(\tilde{\sigma}_e^2 - \sigma_e^2) \]

using the first order approximation. The variance of \( \tilde{\sigma}_v^2 \) is estimated by

\[ V(\tilde{\sigma}_v^2) = \sigma_v^2 V(\tilde{R}) + R^2 V(\tilde{\sigma}_e^2) + 2 \sigma_e^2 R C(\tilde{R}, \tilde{\sigma}_e^2) . \]  \hspace{1cm} (4.50)

The test statistic for consistency of prior and sample data is

\[ t_\infty = (V(\tilde{R}) + V(\hat{R}))^{-1/2} |\tilde{R} - \hat{R}| . \]  \hspace{1cm} (4.51)
V. PREDICTION UNDER THE HETEROGENEOUS NESTED-ERROR MODEL

One of the assumptions of the nested-error model of Section III.B is that the within-cluster variance component, $\sigma_e^2$, is the same in every cluster. For this reason, the model defined by (3.1) - (3.2) will be referred to as the homogeneous nested-error model. The assumption of homogeneous random error components is not always reasonable. For example, in the crop area prediction problem, sample counties might be quite diverse, especially if the number of counties in the sample is large. This chapter deals with small area prediction under the heterogeneous nested-error model in which the within-cluster variance components, $\sigma_1^2$, or more simply $\sigma_i^2$, differ from cluster to cluster.

A. The Model and Predictor

Let the heterogeneous nested-error model be defined by

$$
\begin{align*}
  y_{ij} &= x_{ij} \beta + u_{ij}, \\
  u_{ij} &= v_i + e_{ij}, & i = 1,2,\ldots, t, & j = 1,2,\ldots, n_i,
\end{align*}
$$

(5.1)

where $y_{ij}$, $x_{ij}$, and $\beta$ are defined as before. The random errors, $e_{ij}$, are assumed to be NID(0, $\sigma_i^2$), independent of the cluster effects, $v_i$, which are assumed to be NID(0, $\sigma_v^2$). Thus, the covariance structure is defined by
\[
E(u_{ij}, u_{km}) = \begin{cases} 
\sigma^2_{ij} + \sigma^2_{km}, & \text{if } i \neq j \text{ and } j = m \\
\sigma^2_{ij}, & \text{if } i = j \text{ and } j \neq m \\
0, & \text{if } i = k
\end{cases}
\]  
(5.2)

The only difference between this model and (3.1) - (3.2) is that in (5.2) the variance \( \sigma^2_i \) of the \( e_{ij} \)'s is different for each cluster.

Under the assumption that the variance components are known, the best predictor of the cluster mean,

\[
u_i = \bar{x}_i(p)\hat{\beta} + v_i, \quad i = 1, 2, \ldots, t,
\]
can be obtained by the same procedure as that used by Battese and Fuller (1981). That is, a class of estimators can be defined as

\[
\bar{x}_i(p)\tilde{\beta} + \delta_i(y_i - \bar{x}_i, \tilde{\beta}), \quad 0 < \delta_i < 1,
\]
where \( \tilde{\beta} \) is the generalized least squares estimator of \( \beta \). The mean squared error can be computed for this class, and the best predictor is found by determining the value of \( \delta_i \) which minimizes the mean squared error. Using this procedure, the best predictor is found to be

\[
\bar{v}_i^{(\gamma)} = \bar{x}_i(p)\tilde{\beta} + \gamma(y_i - \bar{x}_i, \tilde{\beta}),
\]  
(5.3)

where
Alternatively, the best predictor can be obtained through linear model theory. The heterogeneous nested-error model can be expressed as a mixed linear model

$$\chi = X \beta + Z \gamma + \varepsilon ,$$

where $Z$ is a matrix of indicator variables. For ease of notation, let

$$H = E\{\gamma \gamma'\} = \sigma^2_v I_t ,$$

$$R = E\{\varepsilon \varepsilon'\} = \text{block diag}(\sigma^2_i I_{n_i}) ,$$

and

$$V = \text{Var}\{\chi\} = Z H Z' + R .$$

Note that

$$V = \text{block diag}(V_i) ,$$

where

$$\gamma_1 = \sigma^2(v \sigma^2 + \sigma^2/n_i)^{-1} .$$
When $V$ is known, the best linear unbiased estimator of $\theta$ is any solution $\tilde{\theta}$ to the generalized least squares equations

$$(X'V^{-1}X)\tilde{\theta} = X'V^{-1}Y.$$

The best linear unbiased predictor of

$$\nu_i = \tilde{x}_1(p)\tilde{\theta} + v_i$$

is

$$\tilde{x}_1(p)\tilde{\theta} + \tilde{v}_i,$$

where

$$\tilde{\chi} = H Z'V^{-1}(\chi - X\tilde{\theta})$$

$$= \sigma^2_v \text{block diag}(L_{n_1}^{-1}v_i^{-1})(\chi - X\tilde{\theta})$$
\[
\begin{align*}
\sigma_v^2 \text{ block diag} & \left( \sigma_1^{-2} \left[ \lambda_n^\prime - n_1 \sigma_1^2 (\sigma_1^2 + n_1 \sigma_2^2)^{-1} \lambda_n^\prime \right] \right) (y - \tilde{X} \tilde{\theta}) \\
= \sigma_v^2 \text{ block diag} & \left( (\sigma_1^2 + n_1 \sigma_2^2)^{-1} \lambda_n^\prime \right) (y - \tilde{X} \tilde{\theta}) ,
\end{align*}
\]

so

\[
\begin{align*}
\tilde{\nu}_1 &= \sigma_v^2 (\sigma_1^2 + n_1 \sigma_2^2)^{-1} \sum_{j=1}^{n_1} (y_{1j} - \tilde{x}_{1j} \tilde{\theta}) \\
&= n_1 \sigma_v^2 (\sigma_1^2 + n_1 \sigma_2^2)^{-1} (\tilde{y}_1 - \tilde{x}_1 \tilde{\theta}) \\
&= \gamma_1 (\tilde{y}_1 - \tilde{x}_1 \tilde{\theta}) .
\end{align*}
\]

Hence, the best linear unbiased predictor agrees with \( \hat{\mu}_1^{(Y)} \) of (5.3), and it is completely analogous to the best predictor for the homogeneous nested-error model.

The mean squared error of the best predictor is

\[
E \left\{ (\hat{\mu}_1^{(Y)} - \mu_1)^2 \right\} = (\tilde{x}_1^{(p)} - \gamma_1 \tilde{x}_1) V(\tilde{\theta})(\tilde{x}_1^{(p)} - \gamma_1 \tilde{x}_1)' + (1 - \gamma_1) \sigma_v^2 .
\] (5.9)

The mean squared conditional bias of \( \hat{\mu}_1^{(Y)} \) is

\[
E \left\{ E \left\{ \hat{\mu}_1^{(Y)} \mid y \right\} - \mu_1 \right\}^2 \right\}
\]
\[
\sum_{j=1}^{t} [(\bar{X}_i(p) - \gamma_1 \bar{X}_1) \gamma_1 \bar{X}_1]^{2} \gamma_1^2 / \sigma_v^2 \\
+ (1 - \gamma_1)^2 \sigma_v^2 - 2(1 - \gamma_1) \gamma_1 \bar{X}_1(p) - \gamma_1 \bar{X}_1 \gamma_1 \bar{X}_1 \gamma_1 \bar{X}_1 \gamma_1 \bar{X}_1
\]

(5.10)

The derivations of (5.9) and (5.10) are identical to the algebraic derivations of (3.11) and (3.13) with \( \sigma_e^2 \) and \( \gamma_1 \) of (5.4) replacing \( \sigma_e^2 \) and \( \gamma_1 \) of (3.7), respectively.

B. Estimation of Variance Components

When the variance components are not known, they must be estimated. Estimates of the variance components can be substituted into the formulas of the previous section. The penalty for estimating the variance components is an increase in the mean squared error of the predictor.

One assumption is that the number of observations within a cluster is bounded, but the number of clusters increases. Thus, any statement of order in probability will be in terms of \( t \). This assumption is reasonable in the crop area estimation problem because the number of segments within a county is finite. Specifically, the assumptions are as follows:

(1) The sequence \( \{n_i\} \) satisfies \( 8 < n_i < n_U < \infty \) for all \( i \);
(2) There are positive constants \( \sigma_L^2 \) and \( \sigma_U^2 \) such that
\[
\sigma_L^2 < \sigma_v^2 < \sigma_i^2 < \sigma_U^2 \quad \text{for all } i;
\]

(3) The rows of \( X \), \((x_{ij1}, \ldots, x_{ijk})\), form a fixed sequence with
\[
|x_{ijh}| < x_U < \infty \quad \text{for all } i, j, \text{ and } h, \text{ and one column of } X \text{ is a column of ones};
\]

(4) As \( t \to \infty \), the limits of \( n^{-1} X'BX \), \( n^{-1} X'G_1X \),
\( n^{-1} X'G_2X \), \( n^{-1} X'N X \) and \( n^{-1} X'KX \) exist and are positive definite, where

\[
B = \text{block diag}\{n_1 \sigma_i^{-2}[(n_1 - 3)^{-1} I - \sigma_v^2 \omega_i J]\},
\]

\[
\omega_i = E\{(U_1 + n_1^2 \sigma_v^2 \sigma_i^{-2})^{-1}\}, \quad U_1 \sim \chi^2_{n_1-1},
\]

\[
G_1 = E\{\tilde{V}^{-1} y y'\tilde{V}^{-1}\},
\]

\[
G_2 = \text{block diag}\{\sigma_i^2(I_n - n_1^{-1} J_{n_1} x n_1)\},
\]

\[
\tilde{V} = \hat{\sigma}_v^2 I_{n_1} + \sigma_v^2 J_{n_1} x n_1,
\]

\[
N = \text{block diag}\{I_{n_1} - n_1^{-1} J_{n_1} x n_1\},
\]

and

\[
K = \text{block diag}\{n_1(n_1 - 3)^{-1} \sigma_i^{-2} I_{n_1} \}.
\]

1. The variance component estimators

The variance component estimators that are considered are analogous to the fitting-of-constants estimators of Section IV.A. The first step
is the ordinary least squares regression of the $y$-deviations, 

$$y_{ij} - \bar{y}_i,$$ 

on the $x$-deviations, $x_{ij} - \bar{x}_i$. Let $\hat{\beta}$ denote the estimated vector obtained in the regression of $y_{ij} - \bar{y}_i$ on $x_{ij} - \bar{x}_i$. The estimator $\hat{\beta}$ is unbiased for $\beta$, and

$$\hat{\beta} - \beta = O_p((n - t)^{-1/2})$$

$$= O_p(t^{-1/2}). \quad (5.11)$$

Let $\hat{\epsilon}_{ij} - \hat{\epsilon}_i$ denote the estimated residuals of this first regression. An estimator of $\sigma_i^2$ is

$$\hat{\sigma}_i^2 = (n_i - 1)^{-1} \sum_{j=1}^{n_i} (\hat{\epsilon}_{ij} - \hat{\epsilon}_i)^2. \quad (5.12)$$

In general, $\hat{\sigma}_i^2$ is not unbiased. Neither is it consistent since $n_i$ is bounded. That is,

$$\hat{\sigma}_i^2 - \sigma_i^2 = O_p(1). \quad (5.13)$$

However, $\hat{\sigma}_i^2$ is converging in probability to an unbiased estimator of $\sigma_i^2$. Note that

$$\hat{\epsilon}_{ij} - \hat{\epsilon}_i = e_{ij} - \hat{\epsilon}_i - (x_{ij} - \bar{x}_i)(\hat{\beta} - \beta).$$
Since $\bar{\bar{g}} - \bar{g} = 0_p(t^{-1/2})$, it follows that

\[
\frac{1}{n_1} \sum_{j=1}^{n_1} \frac{(\bar{e}_{ij} - \bar{e}_{i.})^2}{n_1 - 1} = \frac{1}{n_1} \sum_{j=1}^{n_1} \frac{(e_{ij} - \bar{e}_{i.})^2}{n_1 - 1}
\]

\[
-2 \sum_{j=1}^{n_1} \frac{(e_{ij} - \bar{e}_{i.})(\bar{x}_{ij} - \bar{x}_{.i})(\bar{g} - \bar{g})}{(n_1 - 1)}
\]

\[
+ \sum_{j=1}^{n_1} \frac{[(\bar{x}_{ij} - \bar{x}_{.i})(\bar{g} - \bar{g})]^2}{n_1 - 1}
\]

\[
= \sum_{j=1}^{n_1} \frac{(e_{ij} - \bar{e}_{i.})^2}{n_1 - 1} = 0_p(t^{-1/2}), \quad (5.13a)
\]

and

\[
\sigma_1^2 = \sum_{i=1}^{n_1} \frac{(e_{ij} - \bar{e}_{i.})^2}{n_1 - 1}
\]
is an unbiased estimator of $\sigma_1^2$.

The $\tilde{\sigma}_i^2$ are independent, and the variance of $\hat{\sigma}_1^2$ is approximated by the variance of $\tilde{\sigma}_1^2$, where

$$V(\tilde{\sigma}_1^2) = \frac{2 \sigma_i^4}{n_i - 1}.$$  

The second step in the estimation of the variance components is the ordinary least squares regression of $\chi$ on $X$. Let $\hat{u}$ denote the vector of estimated residuals of this regression. The among mean square is defined by

$$MSA = \frac{1}{t - \lambda} \sum_{i=1}^{t} \sum_{j=1}^{n_i} \hat{u}_{ij}^2 - \sum_{i=1}^{t} \sum_{j=1}^{n_i} (\hat{e}_{ij} - \bar{e}_i.)^2.$$  

Then, $\sigma_v^2$ is estimated by

$$\hat{\sigma}_v^2 = n_\times^{-1} [MSA - (t - \lambda)^{-1} \sum_{i=1}^{t} (1 - n_i^{-1} n_\times \hat{\sigma}_1^2)], \quad (5.14)$$

where $n_\times$ and $\lambda$ are defined as in Section IV.A.

The estimator $\hat{\sigma}_v^2$ is demonstrated to be consistent for $\sigma_v^2$ under assumptions (1) - (4). Recall that

$$\hat{u}_{ij} = u_{ij} - \bar{x}_{ij}(\bar{e} - \bar{e}).$$
where \( \hat{\beta} \) denotes the estimated vector obtained in the regression of \( \chi \) on \( X \). Then,

\[
(t - \lambda)^{-1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} \hat{u}^2_{ij} = (t - \lambda)^{-1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} u^2_{ij} \\
- 2(t - \lambda)^{-1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} u_{ij} x_{ij}(\hat{\beta} - \beta) \\
+ (t - \lambda)^{-1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} (x_{ij}(\hat{\beta} - \beta))^2
\]

\[
= (t - \lambda)^{-1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} u^2_{ij} + o_p(t^{-1})
\]

since \( (\hat{\beta} - \beta) = o_p(n^{-1/2}) \), and

\[
(t - \lambda)^{-1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} u_{ij} x_{ij} = o_p(t^{-1/2})
\]

Using expansion (5.13a) and the fact that

\[
(t - \lambda)^{-1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} (e_{ij} - \bar{e}_i)(x_{ij} - \bar{x}_i) = o_p(t^{-1/2})
\]

it follows that

\[
(t - \lambda)^{-1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} \hat{u}^2_{ij} = (t - \lambda)^{-1} \left[ \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} (e_{ij} - \bar{e}_i)^2 - \sum_{i=1}^{n_1} (1 - n^{-1} n_1)\sigma^2_i \right]
\]

\[
= (t - \lambda)^{-1} \left[ \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} (e_{ij} - \bar{e}_i)^2 - \sum_{i=1}^{n_1} (1 - n^{-1} n_1)\sigma^2_i \right]
\]
Next, using assumptions (1) and (3),

\[
n - n^*_\lambda (t - \lambda) = n - \sum_{i=1}^{t} n_i (1 - n_i \bar{x}_{i1} (x'x)^{-1} \bar{x}'_{i1})
\]

\[
= n_i \bar{x}_{i1} \left( \sum_{i=1}^{t} (n_i \bar{x}_{i1})' (n_i \bar{x}_{i1}) \right) - n_i \bar{x}_{i1}^t
\]

\[
= o(t^{-1})
\]

Therefore,

\[
n^{-1}_\lambda (t - \lambda)^{-1} - n^{-1} = o(t^{-1})
\]

and

\[
\sum_{i=1}^{t} n_i \left[ \sum_{j=1}^{n_i} u_{ij}^2 - \sum_{j=1}^{n_i} (e_{ij} - \bar{e}_{i1})^2 \right]
\]

\[
= o(t^{-1})
\]

Using assumption (2) and the fact that \( \bar{\omega}_i^2 = o_p(1) \),

\[
n^{-2} \sum_{i=1}^{t} n_i \bar{\omega}_i^2 = o_p(t^{-1})
\]
Putting these steps together,

\[
\hat{\sigma}_v^2 = n_*^{-1} \left[ \text{MSA} - (t - \lambda)^{-1} \sum_{i=1}^{t} (1 - n_*^{-1} n_i) \hat{\sigma}_i^2 \right]
\]

\[
= \frac{\sum_{i=1}^{t} \sum_{j=1}^{n_i} \hat{u}_{ij}^2 - \sum_{i=1}^{t} \sum_{j=1}^{n_i} (\hat{e}_{ij} - \bar{e}_i)^2 - \sum_{i=1}^{t} (1 - n_*^{-1} n_i) \hat{\sigma}_i^2}{n_* (t - \lambda)}
\]

\[
+ o_p(t^{-1})
\]

\[
= n_*^{-1} \left[ \sum_{i=1}^{t} \sum_{j=1}^{n_i} \hat{u}_{ij}^2 - \sum_{i=1}^{t} \sum_{j=1}^{n_i} (\hat{e}_{ij} - \bar{e}_i)^2 \right]
\]

\[
- \sum_{i=1}^{t} (1 - n_*^{-1} n_i) \hat{\sigma}_i^2 + o_p(t^{-1})
\]

\[
= \tilde{\sigma}_v^2 + o_p(t^{-1})
\]

where

\[
\tilde{\sigma}_v^2 = n_*^{-1} \left[ \sum_{i=1}^{t} \sum_{j=1}^{n_i} \hat{u}_{ij}^2 - \sum_{i=1}^{t} \sum_{j=1}^{n_i} (\hat{e}_{ij} - \bar{e}_i)^2 - \sum_{i=1}^{t} \hat{\sigma}_i^2 \right]
\]
Now $\tilde{\sigma}_v^2$ is unbiased for $\sigma_v^2$ because
\[
E(\tilde{\sigma}_v^2) = n^{-1} \sum_{t=1}^{t} \sum_{j=1}^{n_i} (\sigma_v^2 + \sigma_1^2) - \sum_{i=1}^{t} (n_i - 1)\sigma_1^2 - \sum_{i=1}^{t} \sigma_i^2
\]
\[= \sigma_v^2.\]

It is now shown that
\[
\sum_{i=1}^{t} \sum_{j=1}^{n_i} u_{ij}^2 - \sum_{i=1}^{t} \sum_{j=1}^{n_i} (e_{ij} - \tilde{e}_i)^2 = u'(I - N)u \tag{5.14a}
\]
and $\tilde{\sigma}_v^2$ are independent. The estimator $\tilde{\sigma}_v^2$ is a multiple of
$\mu' \text{ block diag}(0, \ldots, 0, N_i, 0, \ldots, 0) \mu$, where
\[
N_i = I_{n_i} - n_i^{-1} J_{n_i \times n_i}.
\]

Also,
\[
\sum_{i=1}^{t} \sum_{j=1}^{n_i} u_{ij}^2 - \sum_{i=1}^{t} \sum_{j=1}^{n_i} (e_{ij} - \tilde{e}_i)^2 = \mu'(I - N)\mu,
\]
where
\[
N = \text{block diag}(N_i).
\]

Then, $\mu'(I - N)\mu$ and
$$\tilde{\sigma}_i^2 = (n_i - 1)^{-1} [y_i' \text{ block diag}(0 \ldots, n_i, \ldots, 0) y_i]$$

are independent if \((I - N)\psi \text{ block diag}(0, \ldots, n_i, \ldots, 0) = 0\). The \(i\)-th block of this product is

\[
(I - N_i)\psi N_i = n_i^{-1} J(\sigma_i^2 I + \sigma_v^2 J)(I - n_i^{-1} J)
\]

\[= (n_i^{-1} \sigma_i^2 + \sigma_v^2)J(I - n_i^{-1} J)\]

\[= (n_i^{-1} \sigma_i^2 + \sigma_v^2)(J - J)\]

\[= 0.\]

The other blocks of the product are clearly 0. Therefore, \(\tilde{\sigma}_i^2\) and \(y_i'(I - N)\psi\) are independent.

Using this independence, the variance of \(\tilde{\sigma}_v^2\) can be computed as follows:

\[
V(\tilde{\sigma}_v^2) = n^{-2} [V(y_i'(I - N)\psi) + \sum_{i=1}^{t} V(\tilde{\sigma}_i^2)]
\]

\[= n^{-2} [2 \text{ tr}((I - N)\psi(I - N)\psi) + \sum_{i=1}^{t} 2(n_i - 1)^{-1} \sigma_i^4]\]

\[= 2 n^{-2} [\sum_{i=1}^{t} \text{ tr}(n_i^{-1} \sigma_i^2 + \sigma_v^2) J^2] + \sum_{i=1}^{t} (n_i - 1)^{-1} \sigma_i^4]\]
\[\sum_{i=1}^{t} \left[ (\sigma^2 + n \cdot \sigma^2) + (n_1 - 1)^{-1} \sigma^4 \right] \] (5.15)

\[= 0(t^{-1}) .\]

Then, by Corollary 5.1.1.1 of Fuller (1976),

\[\tilde{\sigma}^2 - \sigma^2 = O_p(t^{-1/2}) ,\]

and

\[\hat{\sigma}^2 - \sigma^2 = O_p(t^{-1/2}). \quad (5.16)\]

Thus, \(\hat{\sigma}^2\) is a consistent estimator of \(\sigma^2\). The variance of \(\hat{\sigma}^2\) may be approximated by the variance of \(\tilde{\sigma}^2\) given by (5.15), and

\[C(\hat{\sigma}^2, \hat{\sigma}^2) = C(\tilde{\sigma}^2, \tilde{\sigma}^2)\]

\[= \frac{-2 \cdot \sigma^4}{n(n_1 - 1)} .\]

In practice, one would use the nonnegative estimator \(\max[0, \hat{\sigma}^2]\) to estimate \(\sigma^2\).

For the special case in which \(X\) consists of only an intercept, the estimators of the variance components reduce to
\[ \hat{\sigma}_i^2 = (n_i - 1)^{-1} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{ij})^2 \quad (5.17) \]

and

\[ \hat{\sigma}_v^2 = \frac{\sum_{i=1}^{t} n_i (\bar{y}_{i.} - \bar{y}_{..})^2 - \sum_{i=1}^{t} (1 - n_i^{-1}) n_i \hat{\sigma}_i^2}{n - \sum n_i^2 / n} \quad . \quad (5.18) \]

These estimators are identical to the ANOVA estimators discussed by Rao, Kaplan and Cochran (1981). The estimators are unbiased, and their variances are given by

\[ \text{Var}(\hat{\sigma}_i^2) = 2 \sigma_i^4 / (n_i - 1) \quad (5.19) \]

and

\[ \text{Var}(\hat{\sigma}_v^2) = 2(n - \sum n_i^2 / n)^{-2} \left[ \sum_{i=1}^{t} (1 - 2 n_i^{-1}) n_i \sigma_i^2 + n_i^{-1} \sigma_i^2 \right]^2 + \sum_{i=1}^{t} n_i \left( \sigma_v^2 + n_i^{-1} \sigma_i^2 \right)^2 \]

\[ + \sum_{i=1}^{t} \left(1 - n_i^{-1} \sigma_i^4 \right) \left( n_i^2 - 1 \right)^{-1} \sigma_i^4 \right] . \quad (5.20) \]
2. **Effects on the estimation of \( \beta \)**

The parameter vector \( \beta \) is estimated by the estimated generalized least squares estimator

\[
\hat{\beta} = (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}y,
\]

(5.21)

where \( \hat{\Sigma} \) estimates \( \Sigma \) with the estimated variance components of the previous section. The properties of \( \hat{\beta} \) are given in Theorems 5.1 and 5.2. The proof of Theorem 5.1 is similar to the proof given for \( \hat{\beta} \) of the homogeneous nested-error model in Theorem 2 of Fuller and Battese (1973). The series of lemmas and proof of Theorem 5.2 follow the proof of a similar result by Fuller and Rao (1978).

**Theorem 5.1.** Let model (5.1) - (5.2) hold with assumptions (1) - (3). The estimated generalized least squares estimator \( \hat{\beta} \) given by (5.21) is an unbiased estimator of \( \beta \).

**Proof.** The error vector \( y \) is symmetrically distributed about \( \mu \), and \( \sigma_i^2 \) and the \( \hat{\sigma_i}^2 \)'s are even functions of \( y \), so

\[
(X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}y
\]

is an even function of \( y \). By the result of Kakwani (1967),

\[
\hat{\beta} - \beta = (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}u
\]

has mean zero if its expectation exists.

Let \( h \) be any \( n \)-dimensional real vector. Then,

\[
|h'X(\hat{\beta} - \beta)| = |h'\hat{\Sigma}^{-1/2}u(\hat{\Sigma}^{-1/2}X)^{-1}X'\hat{\Sigma}^{-1}y|
\]
< (h'V h) \frac{1}{2} (y'V^{-1} y) \frac{1}{2}.

by the Cauchy–Schwarz inequality,

< (h'V h) \frac{1}{2} (y'V^{-1} y) \frac{1}{2}.

The last inequality follows from a result covered by Mardia, Kent and Bibby (1979, appendix) which states that for any real vector \( \chi \) and any symmetric matrices \( A \) and \( B \) of corresponding dimensions with \( B \) positive definite, then \( (\chi'B\chi)^{-1}(\chi'A\chi) < \lambda_1 \), where \( \lambda_1 \) is the largest eigenvalue of \( B^{-1}A \). In this case, \( \chi = V^{-1/2}y \), \( B = I \), and \( A = V^{-1/2}X(V^{-1/2}X)^{-1}X^{-1/2} \) which is idempotent, so \( \lambda_1 = 1 \). To use the same result again, the largest eigenvalues of \( \hat{V} \) and \( \hat{V}^{-1} \) are needed. Since \( \hat{V} \) is block diagonal,

\[ \det(\hat{V} - \lambda I) = \prod_{i=1}^{t} \det(\hat{V}_i - \lambda I). \]

The determinant of a matrix of the form \( a I + b I_n \times n_1 \) is \( a^{n_1-1} (a + n_1 b) \), so the characteristic equation of \( \hat{V} \) is

\[ \prod_{i=1}^{t} (\sigma_i^2 - \lambda)^{\frac{n_i-1}{n_1}} (\sigma_i^2 + \frac{n_i}{n_1} \sigma_i^2 - \lambda). \]
The largest eigenvalue of \( \hat{V} \) is then \( \hat{\sigma}_m^2 + n \hat{\sigma}_m^2 \), where \( m \) indicates the largest value of \( \hat{\sigma}_1^2 + n \hat{\sigma}_1^2 \), \( i = 1, 2, \ldots, t \). The largest eigenvalue of \( \hat{V}^{-1} \) is \( \hat{\sigma}_k^2 \), where \( \hat{\sigma}_k^2 \) is the smallest \( \hat{\sigma}_1^2 \), \( i = 1, 2, \ldots, t \). Then,

\[
(h'\hat{V} h)^{1/2} \left( y'\hat{V}^{-1} y \right)^{1/2} \leq \hat{\sigma}_k^{-2} (\sigma_m^2 + n \hat{\sigma}_m^2)(h'h)^{1/2} (y'y)^{1/2} .
\]

Furthermore, since \( \hat{\sigma}_m^2 + n \hat{\sigma}_m^2 < C y'y \) for some constant \( C \) which depends on \( X \),

\[
|h'X(\hat{g} - g)| < \hat{\sigma}_k^{-2} C(h'h)^{1/2} (y'y)^{3/2} .
\]

By assumption, \( y \) is normally distributed and \( n_i > 3 \), \( i = 1, 2, \ldots, t \), so the expectation of the right hand side exists. Then,

\[
E\{|h'X(\hat{g} - g)|\} < \infty ,
\]

and \( E(\hat{g} - g) \) exists. Therefore, \( \hat{g} \) is unbiased for \( g \). \( \square \)

The following lemmas are needed to determine the asymptotic distribution of \( \hat{g}_1 \), which is given in Theorem 5.2.

**Lemma 5.1.** Let \( \{Z_i\} \) be a sequence of random variables with distribution functions \( \{F_i\} \). Let \( S_t = Z_{i=1}^t \). Let \( \{b_t\} \) be an increasing sequence of positive numbers. Let
If

\[ a_t = \sum_{i=1}^{t} \int |z| < b_t \cdot z \, d F_i(z) \, \]

and

\[ (i) \sum_{i=1}^{t} \int |z| > b_t \cdot \int |z| < b_t \cdot z^2 \, d F_i(z) = o(1) \]

then

\[ b_t^{-1}(s_t - a_t) \xrightarrow{p} 0 \, . \]

Proof. The proof of Lemma 5.1 is given on page 111 of Chung (1974).

Lemma 5.2. If \( U_3 \) and \( U_6 \) are chi-square random variables with 3 and 6 degrees of freedom, respectively, and \( c \) is a positive number, then

\[ P(U_3 < c) < c^{3/2} \]

and
Proof.

\[ P(U_6 < c) < c^3. \]

Also,

\[ P(U_6 < c) = \left[ 2^{3/2} \Gamma(3/2) \right]^{-1} \int_{0}^{c} u^{1/2} e^{-\frac{u}{2}} \, du \]

\[ < \left[ 2^{3/2} \Gamma(3/2) \right]^{-1} \int_{0}^{c} u^{1/2} \, du \]

\[ = \left[ 3 \Gamma(3/2) \right]^{1/2} \int_{0}^{c} u^{1/2} \, du \]

\[ < c^{3/2}. \]

Lemma 5.3. Let model (5.1) - (5.2) and assumptions (1) - (2) hold. Let \( \{b_i\} \) be such that \( \|b_i\| < b_U < \infty \) for all \( i \). Let
where $A$ is a real number. Then,

$$
\lim_{t \to \infty} t^{-1} \sum_{i=1}^{t} b_i \tilde{\sigma}_i^{-2} = A,
$$

where

$$
\tilde{\sigma}_i^{-2} = (n_1 - 1)^{-1} \sum_{j=1}^{n_1} (e_{ij} - \bar{e}_i)^2.
$$

(5.22)

**Proof.** Let $Z_i = b_i \tilde{\sigma}_i^{-2} = b_i n_1 (\sigma_i^2 U)^{-1}$, where $U$ is a chi-square random variable with $n_1 - 1$ degrees of freedom. Then for $0 < \delta < 1$,

$$
E[|Z_i|^{1+\delta}] = (|b_i n_1 \sigma_i^{-2}|^{1+\delta} E[U^{-(1+\delta)}])
$$

$$
= c_1 \int_0^{\infty} \left( \frac{n_1 - 1}{2} - 1 - \delta \right) - 1 \left( 1 - \frac{U}{2} \right) e^{-U/2} dU
$$

$$
= c_2 \int_0^{\infty} \left( \frac{n_1 - 3}{2} - \delta \right) - 1 e^{-w} dw
$$
\[ c_2 \Gamma \left( \frac{n_i - 3}{2} - \delta \right) = 0(1), \]

so

\[ t^{-1} \sum_{i=1}^{t} \left( E[Z_i^{1+\delta}] = o(1) \right). \]

By Exercise 4.5 on page 146 of Rao (1973),

\[ t^{-1} \left( \sum_{i=1}^{t} Z_i - \sum_{i=1}^{t} E[Z_i] \right) \xrightarrow{p} 0. \]

But

\[ t^{-1} \sum_{i=1}^{t} E[Z_i] = t^{-1} \sum_{i=1}^{t} E[b_i n_i \sigma_i^2 U^{-1}] = t^{-1} \sum_{i=1}^{t} \frac{b_i n_i}{\sigma_i^2 (n_i - 3)}, \]

which converges to \( A \) by assumption. Therefore,

\[ t^{-1} \sum_{i=1}^{t} Z_i \xrightarrow{p} A. \]
Lemma 5.4. Let model (5.1) - (5.2) and assumptions (1) - (2) hold. Then for $3^{-1} < \alpha < 2^{-1}$,

$$t^{-3\alpha} \sum_{i=1}^{n_1} \left( \sum_{j=1}^{n_1} (e_{ij} - \bar{e}_{ij})^2 \right)^{2/3} |e_{i1} - \bar{e}_{i1}| \xrightarrow{P} 0.$$  

Proof. Let

$$Z_1 = \left( \sum_{j=1}^{4} (w_{1j} - \bar{w}_{1j})^2 \right)^{2/3} |w_{11} - \bar{w}_{11}|,$$

where the $w_{ij}$'s are NID(0, 1). Then,

$$P\{|Z_1| > t^{3\alpha}\} = P\{((w_{11} - \bar{w}_{11})^2)^{1/2} > t^{3\alpha} \sum_{j=1}^{4} (w_{1j} - \bar{w}_{1j})^2\}$$

$$< P\{((\sum_{j=1}^{4} (w_{1j} - \bar{w}_{1j})^2)^{1/2} > t^{3\alpha} \sum_{j=1}^{4} (w_{1j} - \bar{w}_{1j})^2\}$$

$$= P\{((\sum_{j=1}^{4} (w_{1j} - \bar{w}_{1j})^2)^{3/2} > t^{3\alpha}\}$$

$$< P\{\frac{4}{3} U < t^{-2\alpha}\},$$

where $U \sim \chi^2_3$,

$$= O(t^{-3\alpha})$$

by Lemma 5.2. So
\[
\sum_{i=1}^{t} \mathbb{P}(|Z_i| > t^{3\alpha}) < t^{1-3\alpha} = o(1),
\]

and condition (i) of Lemma (5.1) is satisfied. Similarly,

\[
\mathbb{P}(Z > z) < \mathbb{P}(U < z^{-2/3}),
\]

where \( U \sim \chi_3^2 \). For \( z > 0 \), let \( G(z) \) be a cumulative distribution function defined by

\[
G(z) = \begin{cases} 
0 & , z < 1 \\
1 - \left[ 3 \Gamma\left(\frac{3}{2}\right) \Gamma\left(\frac{1}{2}\right) z \right]^{-1} & , z > 1 
\end{cases}
\]

(5.23)

For \( z > 1 \),

\[
1 - F_1(z) = \mathbb{P}(Z > z)
\]

\[
< \mathbb{P}\left\{ U < z^{-\frac{2}{3}} \right\} = 2^{-3/2} \int_{0}^{z^{-\frac{2}{3}}} \Gamma\left(\frac{3}{2}\right)^{-1} \left( \frac{U}{2} \right) e^{-\frac{U}{2}} dU
\]

Hence,

\[ F(z) > G(z) \, . \]

By problem 11 in Chapter 3 of Lehmann (1959),

\[ \int \psi(z) d F_{1}(z) < \int \psi(z) d G(z) \]

for any nondecreasing function \( \psi \). Let

\[ \psi(z) = \begin{cases} 
  z^2 & , z < t^{3\alpha} \\
  t^{6\alpha} & , z > t^{3\alpha} 
\end{cases} \quad \text{(5.24)} \]

Then,

\[ \int_{|z| < t^{3\alpha}} z^2 d F_{1}(z) < \int \psi(z) d F_{1}(z) < \int \psi(z) d F_{1}(z) \]
< \int \psi(z) dG(z)

= G(1^+) - G(1^-) + \int_1^t 3^\alpha z^2 dG(z)

+ \int_t^\infty t^{6\alpha} dG(z)

= 1 - [3\Gamma(3/2)2^{1/2}]^{-1} + \int_1^t [3\Gamma(3/2)2^{1/2}]^{-1} z^2 dG(z)

+ \int_t^\infty t^{6\alpha} [3\Gamma(3/2)2^{1/2}]^{-1} z^2 dG(z)

= 1 - [3\Gamma(3/2)2^{1/2}]^{-1} + [3\Gamma(3/2)2^{1/2}]^{-1}(t^{3\alpha} - 1)

+ t^{3\alpha} [3\Gamma(3/2)2^{1/2}]^{-1}

= 1 - 2[3\Gamma(3/2)2^{1/2}]^{-1}(1 - t^{3\alpha})

= o(t^{3\alpha})

So

\int_{i=1}^t -6\alpha \sum \int_{|z| < t^{3\alpha}} z^2 dF_1(z) = o(1),

and condition (ii) of Lemma 5.1 is satisfied.

Using the Cauchy-Schwarz inequality,
\[ E(Z_i) = E \left( \frac{|w_{ij} - \overline{w}_i|}{\Sigma (w_{ij} - \overline{w}_i)^2} \cdot \frac{1}{\Sigma (w_{ij} - \overline{w}_i)^2} \right) \]

\[ < (E((w_{ij} - \overline{w}_i)^2(\Sigma (w_{ij} - \overline{w}_i)^2)^{-2})E((\Sigma (w_{ij} - \overline{w}_i)^2)^{-2}))^{1/2} \]

\[ < E((\Sigma (w_{ij} - \overline{w}_i)^2)^{-1})E((\Sigma (w_{ij} - \overline{w}_i)^2)^{-2})^{1/2} \]

\[ = c(E(U^{-1})E(U^{-2}))^{1/2} \]

where \( U \sim \chi^2_{n_1} - 1 \), which is finite for \( n_1 > 6 \). Then,

\[ t^{-3a} \sum_{i=1}^{t} Z_i \xrightarrow{p} 0. \]

When \( e_{ij}'s \) are used in place of the \( w_{ij}'s \), then \( Z_i \) is altered by the standard deviation of \( e_{ij} - \overline{e}_i \), which is bounded, so the asymptotic result is the same. Furthermore, since \( n_1 > 4 \) for every \( i \),

\[ \frac{|e_{ij} - \overline{e}_i|}{(\sum_{i=1}^{n_1} (e_{ij} - \overline{e}_i)^2)^{2}} < \frac{|e_{ij} - \overline{e}_i|}{(\sum_{i=1}^{n_1} (e_{ij} - \overline{e}_i)^2)^{2}} \]

and the result still holds. \( \square \)
Lemma 5.5. Let model (5.1) - (5.2) and assumptions (1) - (3) hold. Then,

$$t^{-1} \sum_{i=1}^{t} \left| \tilde{\sigma}_i - \hat{\sigma}_i \right|^2 \xrightarrow{p} 0.$$  

Proof. Using the definitions,

$$(n_i - 1)^{-1} (\tilde{\sigma}_i - \hat{\sigma}_i)^2 = \sum_{l=1}^{n_i} \left( e_{ij} - \hat{e}_{ij} \right)^2 \left( e_{ij} - \hat{e}_{ij} \right)^2,$$

$$= \frac{\sum_{l=1}^{n_i} \left( e_{ij} - \hat{e}_{ij} \right)^2 \sum_{l=1}^{n_i} \left( e_{ij} - \hat{e}_{ij} \right)^2}{\sum_{l=1}^{n_i} \left( e_{ij} - \hat{e}_{ij} \right)^2 \sum_{l=1}^{n_i} \left( e_{ij} - \hat{e}_{ij} \right)^2},$$

$$(5.25)$$

where

$$g_{ij} = (e_{ij} - \hat{e}_{ij}) \cdot \left( e_{ij} - \hat{e}_{ij} \right).$$
and \( \hat{\beta} \) is the estimated vector obtained in the ordinary least squares regression of \( y_{ij} - \bar{y}_i \) on \( x_{ij} - \bar{x}_i \).

Let \( \delta > 0 \) and \( \epsilon > 0 \) be given. Let \( \alpha \) be defined as in Lemma 5.4. Let \( A_{1t} \) be the event that \( \max_j |e_{ij} - \bar{e}_{i1} | < t^{-\alpha} \) for at least one \( i \). Let \( A_{2t} \) be the event that \( \max_i |g_{ij} | < (2t)^{-\alpha} \). Let \( A_t = A_{1t} \cup A_{2t} \). Then,

\[
P(t^{-1} \sum_{i=1}^{t} |\hat{\sigma}_i^2 - \bar{\sigma}_i^2| > \delta) \leq P(A_t) + P(t^{-1} \sum_{i=1}^{t} |\hat{\sigma}_i^2 - \bar{\sigma}_i^2| > \delta | \bar{A}_t) + P(\sum_{i=1}^{t} q_i > \delta)
\]

for some positive constant \( b_2 \), where

\[
q_i = \frac{\sum_j |e_{ij} - \bar{e}_{i1} |}{(\sum_j (e_{ij} - \bar{e}_{i1})^2)^2}.
\quad (5.25a)
\]

Now
\[ P(A_{1t}) = P(\max_{i,j} |e_{ij} - \bar{e}_{i,j}| < t^{-\alpha} \text{ for some } i) \]

\[ = P(\max_{i,j} |e_{ij} - \bar{e}_{i,i}| < t^{-\alpha} \text{ for all } j \text{ for some } i) \]

\[ < t P(\max_{i,j} |e_{ij} - \bar{e}_{i,i}| < t^{-\alpha}, j = 1, 2, 3) \]

\[ = t(P(\max_{i,j} |e_{ij} - \bar{e}_{i,i}| < t^{-\alpha}))^3 \]

\[ = t P(|w| < t^{-\alpha} \sigma_i^{-1} n_i^{-1} (n_i - 1))^3, \]

where \( w \sim N(0, 1) \). So,

\[ P(A_{1t}) < t(c t^{-\alpha} \sigma_i^{-1} n_i^{-1} (n_i - 1))^3 \]

\[ = O(t^{1-3\alpha}). \]

Hence,

\[ \lim_{t \to \infty} P(A_{1t}) = 0. \]

Next,

\[ P(A_{2t}) = P(\max_{i,j} |g_{ij}| > (2t)^{-\alpha}) \]
by assumption (3) and the fact that $(\bar{g} - \bar{g}) = O(t^{-1/2})$. Therefore,

$P(A_{1t} \cup A_{2t}) \rightarrow 0$. There is a $\tau_1$ such that $P(A_t) < \frac{\varepsilon}{2}$ for $t > \tau_1$.

When $A_t$ is true, then $|g_{ij}| < (2t)^{-\alpha}$ for all $i$ and $j$, and $|e_{im} - \bar{e}_i| = \max_j |e_{ij} - \bar{e}_i| > t^{-\alpha}$ for all $i$. Working with the denominator of (5.25),

$$\sum_j (e_{ij} - \bar{e}_i - g_{ij})^2 > (e_{im} - \bar{e}_i - g_{im})^2.$$ 

If $(e_{im} - \bar{e}_i)$ and $g_{im}$ are of opposite signs, then

$$(e_{im} - \bar{e}_i - g_{im})^2 > (e_{im} - \bar{e}_i)^2$$

$$> (.04)(e_{im} - \bar{e}_i)^2.$$ 

If $(e_{im} - \bar{e}_i)$ and $g_{im}$ are of the same sign (without loss of generality suppose both are positive), then

$$(e_{im} - \bar{e}_i - g_{im})^2 > [(e_{im} - \bar{e}_i) - 2t^{-\alpha}]^2$$
Either way,

\[(0.04)(e_{im} - \bar{e}_i.)^2 > (0.04)n^{-1} \sum_{j=1}^{n_l} (e_{ij} - \bar{e}_i.)^2,\]

so

\[\sum_{j=1}^{n_l} (e_{ij} - \bar{e}_i. - g_{ij})^2 > (0.04)n^{-1} \sum_{j=1}^{n_l} (e_{ij} - \bar{e}_i.)^2.\]

(5.26)

Working with the numerator of (5.25),

\[|\sum_{j} [2(e_{ij} - \bar{e}_i.) - g_{ij}]| < 2 \sum_{j} |e_{ij} - \bar{e}_i.| + \sum_{j} |g_{ij}|\]

\[< 2 \sum_{j} |e_{ij} - \bar{e}_i.| + n_i^{-2^{\alpha}} \tau^{-\alpha}\]

\[< 2 \sum_{j} |e_{ij} - \bar{e}_i.| + n_i^{-2^{\alpha}} |e_{im} - \bar{e}_i.|\]
for some positive constant $c$.

Combining (5.26) and (5.27), it follows that there is a positive number $b_1$ such that

$$\frac{\left| \Sigma_j [2(e_{ij} - \bar{e}_{i.}) - g_{ij}] \right|}{\Sigma_j (e_{ij} - \bar{e}_{i.})^2 \Sigma_j (e_{ij} - \bar{e}_{i.} - g_{ij})^2} < b_1 \frac{\Sigma_j |e_{ij} - \bar{e}_{i.}|}{(\Sigma_j (e_{ij} - \bar{e}_{i.})^2)^2}$$

(5.28)

$$= b_1 q_i,$$

where $q_i$ is defined in (5.25a). Then

$$P(t^{-1} \sum_{i=1}^{t} L_t \hat{\sigma}_i - \sigma_i^2 > b_2 t^{-3a} \sum_{i=1}^{t} q_i |\hat{\Lambda}_t|)$$

$$< P \left\{ t^{-1} \sum_{i=1}^{t} \left| \frac{\Sigma_j [2(e_{ij} - \bar{e}_{i.}) - g_{ij}] g_{ij}}{\Sigma_j (e_{ij} - \bar{e}_{i.})^2 \Sigma_j (e_{ij} - \bar{e}_{i.} - g_{ij})^2} \right| > b_2 t^{-3a} \sum_{i=1}^{t} q_i |\hat{\Lambda}_t| \right\}$$

$$< b_2 t^{-3a} \sum_{i=1}^{t} q_i |\hat{\Lambda}_t|\right\}$$
< P \left\{ b_1 (2t)^{-\alpha} t^{-1} \sum_{i=1}^{t} \frac{\sum |e_{ij} - \bar{e}_{ii}|}{(\sum (e_{ij} - \bar{e}_{ii})^2)^2} \right\} \\
> b_2 t^{-3\alpha} \sum_{i=1}^{t} q_i |\bar{A}_t| \\
= p \left\{ 2^{-\alpha} b_1 t^{-(1+\alpha)} \sum_{i=1}^{t} q_i > b_2 t^{-3\alpha} \sum_{i=1}^{t} q_i |\bar{A}_t| \right\} \\
= 0

for some constant \( b_2 \) since \( t^{-(1+\alpha)} < t^{-3\alpha} \).

Finally, by Lemma 5.4 there is a \( \tau_2 \) such that for \( t > \tau_2 \),

\[ P \left\{ b_2 t^{-3\alpha} \sum_{i=1}^{t} q_i > \delta \right\} < \frac{\varepsilon}{2}. \]

Let \( \tau = \max[\tau_1, \tau_2] \). Then for \( t > \tau \),

\[ P \left\{ t^{-1} \sum_{i=1}^{t} \left| \hat{\sigma}_i^{-2} - \check{\sigma}_i^{-2} \right| > \delta \right\} < \varepsilon. \quad \square \]

**Lemma 5.6.** Let model (5.1) - (5.2) and assumptions (1) - (2) hold, and let \( \{b_i\} \) and \( \{c_i\} \) be sequences such that \( |b_i| < b_U < \infty \) and \( 0 < c_i < c_U < \infty \) for all \( i \). Let
where $\omega_i = E((U + c_i)^{-1})$ and $U \sim \chi^2_{n_i} - 1$. Then,

$$t^{-1} \sum b_i n_i \sigma_i^{-2} \omega_i = A$$

Proof. Let $Z_i = b_i (\tilde{\sigma}_i^2 + c_i n_i^{-1} \sigma_i^2)^{-1} = b_i n_i \sigma_i^{-2} (U + c_i)^{-1}$, where

$U \sim \chi^2_{n_i} - 1$. Then for $0 < \delta < 1$,

$$E\{|Z_i|^{1+\delta}\} = \left(|b_i| n_i \sigma_i^{-2}\right)^{1+\delta} E\{(U + c_i)^{-(1+\delta)}\}$$

$$< \left(|b_i| n_i \sigma_i^{-2}\right)^{1+\delta} E\{U^{-(1+\delta)}\}$$

$$= o(1)$$

as in Lemma 5.3. Then,

$$t^{-(1+\delta)} \sum E\{|Z_i|^{1+\delta}\} = o(1)$$

By exercise 4.5 on page 146 of Rao (1973),
\[ t^{-1} \sum_{i=1}^{t} Z_i - \sum_{i=1}^{t} E(Z_i) \xrightarrow{p} 0. \]

That is,

\[ \text{plim } t^{-1} \sum_{i=1}^{t} Z_i = \lim_{t \to \infty} t^{-1} \sum_{i=1}^{t} b_{i} n_{i} \sigma_{i}^{-2} \omega_{i} = A. \]

It can be verified that

\[ (n_{i} - 3 + c_{i})^{-1} < \omega_{i} < (n_{i} - 3)^{-1}. \]

To establish the upper bound,

\[ E\{(U_{i} + c_{i})^{-1}\} < E\{U_{i}^{-1}\} = (n_{i} - 3)^{-1}. \]

On the other hand,

\[ E\{(U_{i} + c_{i})^{-1}\} = E\{U_{i}^{-1}\}[(U_{i} + c_{i}) - c_{i}](U_{i} + c_{i})^{-1}\]

\[ = E\{U_{i}^{-1}\} - c_{i} E\{U_{i}^{-1}(U_{i} + c_{i})^{-1}\}\]
where \( U_2 \sim \chi^2_{n_i - 3} \). So,

\[
E\{(U_1 + c_i)^{-1}\} > (n_i - 3)^{-1}[1 - c_i E\{(U_1 + c_i)^{-1}\}] \quad (5.29a)
\]

Inequality (5.29a) follows because

\[
E\{U_1^{-1}\} = (n_i - 3)^{-1} < (n_i - 5)^{-1} = E\{U_2^{-1}\} ,
\]

which, in turn, implies that

\[
E\{(U_1 + c_i)^{-1}\} < E\{(U_2 + c_i)^{-1}\} .
\]

Then,

\[
[1 + c_i(n_i - 3)^{-1}]E\{(U_1 + c_i)^{-1}\} > (n_i - 3)^{-1} ,
\]

which implies that
Lemma 5.7. Let model (5.1) - (5.2) and assumptions (1) - (3) hold. Then,

\[
E\{(U_1 + c_i)^{-1}\} > (n_i - 3 + c_i)^{-1}.
\]

Proof. Let \( \delta > 0 \) be given. Then,

\[
P\left\{ \sum_{i=1}^{t-1} \left[ \frac{\hat{\sigma}_v^2}{\sigma_i^2 + n_1 \hat{\sigma}_v^2} - \frac{\sigma_v^2}{\sigma_i^2 + n_1 \sigma_v^2} \right] > \delta \right\}
\]

\[
< P\left\{ \sum_{i=1}^{t-1} \left[ \frac{\hat{\sigma}_v^2}{\sigma_i^2 + n_1 \hat{\sigma}_v^2} - \frac{\sigma_v^2}{\sigma_i^2 + n_1 \sigma_v^2} + \sum_{i=1}^{t-1} \left[ \frac{\hat{\sigma}_v^2}{\sigma_i^2 + n_1 \hat{\sigma}_v^2} - \frac{\sigma_v^2}{\sigma_i^2 + n_1 \sigma_v^2} \right] > \frac{\delta}{2} \right\}
\]

\[
+ P\left\{ \sum_{i=1}^{t-1} \left[ \frac{\hat{\sigma}_v^2}{\sigma_i^2 + n_1 \hat{\sigma}_v^2} - \frac{\sigma_v^2}{\sigma_i^2 + n_1 \sigma_v^2} \right] > \frac{\delta}{2} \right\}.
\]

Now

\[
P\left\{ \sum_{i=1}^{t-1} \left[ \frac{\hat{\sigma}_v^2}{\sigma_i^2 + n_1 \hat{\sigma}_v^2} - \frac{\sigma_v^2}{\sigma_i^2 + n_1 \sigma_v^2} \right] > \frac{\delta}{2} \right\}
\]
\begin{align*}
&= \mathbb{P} \left\{ t^{-1} \sum_{i=1}^{\infty} \frac{\hat{\sigma}_i^2 (\sigma_i^2 - \sigma^2_v)}{(\sigma_i^2 + n_1 \sigma_i^2)(\sigma_i^2 + n_1 \sigma_i^2)} > \frac{\delta}{2} \right\} \\
&< \mathbb{P} \left\{ (8\sigma_v^2)^{-1} t^{-1} \sum_{i=1}^{\infty} \frac{\hat{\sigma}_i^2 - \sigma^2_v}{(\sigma_i^2 + n_1 \sigma_i^2)(\sigma_i^2 + n_1 \sigma_i^2)} > \frac{\delta}{2} \right\} \\
&= \mathbb{P} \left\{ (8\sigma_v^2)^{-1} |\hat{\sigma}_v^2 - \sigma_v^2| > \frac{\delta}{2} \right\} \\
&\rightarrow 0 \text{ by (5.16).}
\end{align*}

Also,
\begin{align*}
&= \mathbb{P} \left\{ t^{-1} \sum_{i=1}^{\infty} \frac{\sigma^2_v}{\sigma_i^2 + n_1 \sigma_i^2} - \frac{\sigma^2_v}{\sigma_i^2 + n_1 \sigma_i^2} > \frac{\delta}{2} \right\} \\
&< \mathbb{P} \left\{ t^{-1} \sum_{i=1}^{\infty} \frac{n_1 \sigma^2_v}{\sigma_i^2 + n_1 \sigma_i^2} - \frac{n_1 \sigma^2_v}{\sigma_i^2 + n_1 \sigma_i^2} > \frac{\delta}{2} \right\} \\
&= \mathbb{P} \left\{ t^{-1} \sum_{i=1}^{\infty} \frac{n_1 \sigma^2_v (\hat{\sigma}_i^2 - \hat{\sigma}_i^2)}{(\sigma_i^2 + n_1 \sigma_i^2)(\sigma_i^2 + n_1 \sigma_i^2)} > \frac{\delta}{2} \right\} \\
&< \mathbb{P} \left\{ \frac{n_1 \sigma^2_v}{\sigma_i^2} t^{-1} \sum_{i=1}^{\infty} \frac{\hat{\sigma}_i^2 - \hat{\sigma}_i^2}{\sigma_i^2 + \sigma_i^2} > \frac{\delta}{2} \right\} \\
&\rightarrow 0 \text{ by (5.16).}
\end{align*}
\[ = P \left( n \sigma_v^2 t^{-1} \sum_{i=1}^{\hat{n}} \left| \frac{\hat{\sigma}_i^2}{\hat{\sigma}_i^2 + n_i \sigma_v^2} - \frac{\sigma_v^2}{\sigma_v^2 + n_i \sigma_v^2} \right| > \frac{\delta_i}{2} \right) \rightarrow 0 \text{ by Lemma 5.5.} \]

Therefore,

\[ t^{-1} \sum_{i=1}^{\hat{n}} \left| \frac{\hat{\sigma}_i^2}{\hat{\sigma}_i^2 + n_i \sigma_v^2} - \frac{\sigma_v^2}{\sigma_v^2 + n_i \sigma_v^2} \right| \rightarrow 0. \]

\[ \square \]

Lemma 5.8. Let model (5.1) - (5.2) and assumptions (1) - (2) hold, and let \( 3^{-1} < \alpha < 2^{-1} \). Then,

\[ t^{-3\alpha} \sum_{i=1}^{n_i} \left[ \sum_{j=1}^{n_i} (e_{i, j} - \bar{e}_{i, j})^2 \right]^{-2} |\nu_i| \rightarrow 0. \]

Proof. Let

\[ z_i = \frac{|\nu_i|}{\sqrt{\sum_{j=1}^{n_i} (w_{i, j} - \bar{w}_{i, j})^2}}, \]

where the \( w_{i, j} \)'s are normally and independently distributed with variance 1, independently of \( \nu_i \) which is \( N(0, 1) \). Following the proof of Lemma 5.4,
\[ P(\eta_{1} > t^{3\alpha}) = P(\eta_{1} > t^{3\alpha} (\Sigma \varepsilon_{w_{i} z} - \bar{w}_{1}.)^{2}) \]

\[ = P(\eta_{1} > t^{3\alpha} (\Sigma \varepsilon_{w_{i} z} - \bar{w}_{1}.)^{2} | (\Sigma \varepsilon_{w_{i} z} - \bar{w}_{1}.)^{2} < t^{-\alpha}) \]

\[ < t^{-\alpha} P((\Sigma \varepsilon_{w_{i} z} - \bar{w}_{1}.)^{2} < t^{-\alpha}) \]

\[ + P(\eta_{1} > t^{3\alpha} (\Sigma \varepsilon_{w_{i} z} - \bar{w}_{1}.)^{2} | (\Sigma \varepsilon_{w_{i} z} - \bar{w}_{1}.)^{2} > t^{-\alpha}) \]

\[ > t^{-\alpha} P((\Sigma \varepsilon_{w_{i} z} - \bar{w}_{1}.)^{2} > t^{-\alpha}) \]

\[ < P(\sum_{j=1}^{7} (\varepsilon_{w_{i} z} - \bar{w}_{1}.)^{2} < t^{-\alpha}) + P((\eta_{1}^{2})^{1/2} > t^{3\alpha}) \]

\[ = P(U_{6} < t^{-\alpha}) + P(U_{1} > t^{2\alpha}) \]

where \( U_{6} \sim \chi_{6}^{2} \) and \( U_{1} \sim \chi_{1}^{2} \). Then,

\[ P(\eta_{1} > t^{3\alpha}) < O(t^{-3\alpha}) + P(U_{1} > t^{2\alpha}) \]

by Lemma 5.2. So,

\[ P(\eta_{1} > t^{3\alpha}) < O(t^{-3\alpha}) + \left[ 2 \Gamma(\frac{1}{2}) \right]^{-1} \int_{t^{2\alpha}}^{\infty} U \cdot \frac{U}{\sqrt{2}} e^{-\frac{U}{2}} dU \]

\[ < O(t^{-3\alpha}) + \left[ 2 \Gamma(\frac{1}{2}) \right]^{-1} \int_{t^{2\alpha}}^{\infty} e^{-\frac{U}{2}} dU \]
So,\\
\[ E_{\gamma}(z) = z \prod_{i=1}^{\infty} P\{z > z\} = o(1) \, , \]
and condition (i) of Lemma 5.1 is satisfied.

Similarly,
\[ P\{Z_i > z\} < P\{U_6 < z^{\frac{1}{3}}\} + P\{U_1 > z^3\} \, . \]

Let \( G(z) \) be a cumulative distribution function defined by

\[
G(z) = \begin{cases} 
0, & z < 1 \\
1 - [3(2^{5/2}z)^{-1} - [\Gamma(\frac{1}{2})\exp(z^{2/3})]^{-1}], & z > 1.
\end{cases}
\]

For \( z > 1 \),
\[ 1 - F(z) < P(U_6 < z) + P(U_1 > z) \]

\[ = \left[ \Gamma(3)2^z \right] \int_0^z u^2 e^{-\frac{U}{2}} dU \]

\[ + \left[ 2\Gamma\left(\frac{1}{2}\right) \right]^{-1} \int_{\frac{z}{2}}^\infty u^{-\frac{1}{2}} e^{-\frac{U}{2}} dU \]

\[ < 2 \left[ \frac{5}{2} \int_0^z u^2 dU + \left[ 2\Gamma\left(\frac{1}{2}\right) \right]^{-1} \int_{\frac{z}{2}}^\infty e^{-\frac{U}{2}} dU \right] \]

\[ = [3(2^z)z]^{-1} + \Gamma\left(\frac{1}{2}\right)^{-1} e^{-\frac{z}{2}} \]

\[ = 1 - G(z) . \]

So \( F(z) > G(z) \). Let \( \psi(z) \) be defined as in (5.24). Then,

\[ \int |z| < t^{3\alpha} \sum \psi(z) dF_1(z) < \int \psi(z) dF_1(z) \]

\[ < \int \psi(z) dG(z) \]

\[ = G(1 +) - G(1 -) + \int_1^t z^2 dG(z) + \int_t^\infty t^{6\alpha} dG(z) \]

\[ = 1 - [3(2^z)]^{-1} - [\Gamma\left(\frac{1}{2}\right)e^x]^{-1} \]
\begin{align*}
+ \int_{1}^{\infty} \frac{t^{3\alpha}}{[3(2^2)]^{-1}} \, dz \\
+ \int_{1}^{\infty} & \frac{5}{z^3} - \frac{z^{2/3}}{2} \, dz \\
+ \int_{1}^{\infty} & \frac{5}{2} e^{-\frac{1}{3}z} - \frac{z^{2/3}}{2} \, dz \\
+ \int_{1}^{\infty} & \frac{5}{2} t^{6\alpha} e^{-\frac{1}{3}z} - \frac{z^{2/3}}{2} \, dz
\end{align*}

It can be verified that

\[
\frac{5}{3} - \frac{z^{2/3}}{2}
\]

is maximized in the interval \([1, t^{3\alpha}]\) by \(z = \frac{3}{2}\). Thus,

\[
\int_{1}^{\infty} |z|^{t^{3\alpha}} z^{2} \, dz \leq 1 - \frac{5}{2} [3(2^2)]^{-1} - [3\Gamma(\frac{1}{2})]^{-1}
\]

\[
+ \int_{1}^{\infty} \frac{t^{3\alpha}}{[3(2^2)]^{-1}} \, dz \\
+ \int_{1}^{\infty} & \frac{5}{z^3} - \frac{1}{2} \, dz \\
+ \int_{1}^{\infty} & \frac{5}{2} e^{-\frac{1}{3}z} - \frac{1}{2} \, dz \\
+ \int_{1}^{\infty} & \frac{5}{2} t^{6\alpha} e^{-\frac{1}{3}z} - \frac{1}{2} \, dz
\]
\begin{equation*}
+ \int_{t^{3\alpha}}^{\infty} t^{6\alpha} [3\Gamma(\frac{1}{2})]^{-1} \ z \ - \frac{1}{3} \ e^{-\frac{z^{2/3}}{2}} \ d \ z
= 1 - [3(2^{2\alpha})]^{-1} - [\Gamma(\frac{1}{2})e^{\alpha}]^{-1}
\end{equation*}

\begin{equation*}
+ [3(2^{2\alpha})]^{-1}(t^{3\alpha} - 1) + [3\Gamma(\frac{1}{2})e^{\alpha}]^{-1}(t^{3\alpha} - 1)
\end{equation*}

\begin{equation*}
+ t^{3\alpha}[3(2^{2\alpha})] + t^{6\alpha}[\Gamma(\frac{1}{2})]^{-1} e^{-\frac{t^{3\alpha}}{2}}
= 0(t^{3\alpha}).
\end{equation*}

So, \( t^{6\alpha} \sum_{i=1}^{\infty} \int_{\left|z\right| < t^{3\alpha} z^{2} \ d F_{i}(z) = o(1) \), and condition (ii) of Lemma 5.1 is satisfied.

Now

\begin{equation*}
E\left(\frac{\left|v_{1}\right|}{\left(\sum_{1}^{7} (w_{ij} - \bar{w}_{i.})^{2}\right)^{2}}\right) = c \ E\{U_{1}^{2}\} E\{\bar{U}_{i.}^{-2}\}
= c \left(\int_{0}^{\infty} e^{-\frac{u}{2}} \ d \ u\right)^{2}
= 0(1).
\end{equation*}

Then,
\[ t^{-3\alpha} \sum_{i=1}^{t} E[Z_i] = o(1), \]

and

\[ t^{-3\alpha} \sum_{i=1}^{t} Z_i \xrightarrow{p} 0. \]

As in Lemma 5.4, using \( n_1 > 8 \), \( v_i \)'s and \( e_{ij} \)'s does not change the result. □

**Lemma 5.9.** Let model (5.1) - (5.2) and assumptions (1) - (2) hold, and \( 3^{-1} < \alpha < 2^{-1} \). Then,

\[ t^{-3\alpha} \sum_{i=1}^{t} \sum_{j=1}^{n_i} (e_{ij} - \bar{e}_{i.})^2 \frac{1}{|e_{i1}|} \xrightarrow{p} 0. \]

**Proof.** Note that

\[ t^{-3\alpha} \sum_{i=1}^{t} \sum_{j=1}^{n_i} (e_{ij} - \bar{e}_{i.})^2 \frac{1}{|e_{i1}|} \]

\[ < t^{-3\alpha} \sum_{i=1}^{t} \sum_{j=1}^{n_i} (e_{ij} - \bar{e}_{i.})^2 \frac{1}{|e_{i1}|} \]

\[ + t^{-3\alpha} \sum_{i=1}^{t} \sum_{j=1}^{n_i} (e_{ij} - \bar{e}_{i.})^2 \frac{1}{|\bar{e}_{i.}|} \]

Using the result of Lemma 5.4, the problem becomes that of proving the second term converges to zero in probability. Since \( \bar{e}_{i.} \) is normally
distributed and independent of \( \sum_{j=1}^{n_i} (e_{ij} - \bar{e}_{ij})^2 \), the proof is like that of Lemma 5.8.

**Lemma 5.10.** Let model (5.1) - (5.2) and assumptions (1) - (3) hold. Then,

\[
- \frac{1}{2} \left[ X'^\hat{\Sigma}^{-1} \hat{\Sigma} - X'^\hat{\Sigma}^{-1} \hat{\Sigma} + 2 X' \hat{K} \hat{\Sigma} (\hat{\xi} - \bar{\xi}) \right] \rightarrow 0,
\]

where

\[
\hat{\Sigma} = \text{block diag}(\hat{\sigma}_1^2 I_{n_1}, \hat{\sigma}_2^2 J_{n_1} \times n_1),
\]

\[
\hat{\Sigma} = \text{block diag}(\hat{\sigma}_1^2 I_{n_1}, \hat{\sigma}_2^2 J_{n_1} \times n_1),
\]

\[
\hat{\xi} = (X'^\hat{\Sigma}^{-1} X)^{-1} X'^\hat{\Sigma}^{-1} \xi,
\]

and

\[
K = \text{block diag}(n_1(n_1 - 3)^{-1} \hat{\sigma}_1^2 I_{n_1}).
\]

**Proof.** The r-th element of \(- \frac{1}{2} X'^\hat{\Sigma}^{-1} \hat{\Sigma} \) is

\[
- \frac{1}{2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} \left( \xi \sum_{i=1}^{n_1} u_{ij} \hat{\sigma}_{ij} \right)^2 - \frac{1}{2} \sum_{j=1}^{n_1} \sum_{i=1}^{n_1} u_{ij} \hat{\sigma}_{ij} \hat{\xi} \hat{\sigma}_{ij} (\hat{\sigma}_1^2 + n_1 \hat{\sigma}_2^2)^{-1}.
\]
\[
\begin{align*}
\Delta_t &= \frac{-1}{2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} x_{ijr} u_{ij} \sigma_{ij}^{-2} - p_t \\
&= \frac{-1}{2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} x_{ijr} u_{ij} \sigma_{ij}^{-2} - \frac{1}{2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} x_{ijr} \sum_{j=1}^{n_1} u_{ij} \sigma_{ij} \sigma_{ij}^{-2} (\sigma_{ij}^2 + n_1 \sigma_{ij}^2)^{-1},
\end{align*}
\]

where

\[
P_t = \frac{-1}{2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} x_{ijr} u_{ij} \left( \frac{\sigma_{ij}^2}{\sigma_{ij}^2 + n_1 \sigma_{ij}^2} \right) - \frac{\sigma_{ij}^2}{\sigma_{ij}^2 + n_1 \sigma_{ij}^2} \right)
\]

(5.29b)

Now

\[
\begin{align*}
-\frac{1}{2} t \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} x_{ijr} u_{ij} \sigma_{ij}^{-2} \\
&= \frac{1}{2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} x_{ijr} u_{ij} [\sigma_{ij}^{-2} - \bar{\sigma}_{ij}^{-4} (\sigma_{ij}^2 - \bar{\sigma}_{ij}^2)] \\
&+ \bar{\sigma}_{ij}^{-4} \bar{\sigma}_{ij}^{-2} (\sigma_{ij}^2 - \bar{\sigma}_{ij}^2)^2) \\
&= \frac{1}{2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} x_{ijr} u_{ij} \sigma_{ij}^{-2} \\
&+ (n_1 - 1) \sum_{i=1}^{n_1} \sum_{h=1}^{n_1} (2(e_{ih} - \bar{e}_i) - g_{ih}) g_{ih} \\
&+ (n_1 - 1)^2 \sum_{i=1}^{n_1} \sum_{h=1}^{n_1} (2(e_{ih} - \bar{e}_i) - g_{ih}) g_{ih}
\end{align*}
\]
\begin{align*}
&- g_{ih} \{ g_{ih} \}^2 \\
&= t \frac{1}{2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} x_{ijr_{ij}} [\sigma_i^{\prime -2} \\
&\quad + 2(n_i - 1)^{-1} \sigma_i^{\prime -4} \sum_{h=1}^{n_i} g_{ih}(e_{ih} - \bar{e}_i) \\
&\quad - (n_i - 1)^{-1} \sigma_i^{\prime -4} \sum_{h=1}^{n_i} g_{ih}^2 \\
&\quad + (n_i - 1)^{-2} \sigma_i^{\prime -4} \sum_{i=1}^{n_i} \sum_{h=1}^{n_i} g_{ih}(e_{ih} - \bar{e}_i) \\
&\quad - g_{ih} \} \{ g_{ih} \}^2 \\
&= t \frac{1}{2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} x_{ijr_{ij}} [\sigma_i^{\prime -2} \\
&\quad + 2(n_i - 1)^{-1} \sigma_i^{\prime -4} \sum_{h=1}^{n_i} x_{ihs}(\beta_s - \beta_i)(e_{ih} - \bar{e}_i) ] \\
&\quad - Q_t + R_t , \\
\end{align*}

where

\begin{align*}
Q_t = t \frac{1}{2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} x_{ijr_{ij}} (n_i - 1)^{-1} \sigma_i^{\prime -4} \sum_{h=1}^{n_i} g_{ih}^2 \\
\end{align*}
and

\[ R_t = -\frac{1}{2} t \sum_{i=1}^{n_i} \sum_{j=1}^{n_i} x_{ijr} u_{ij} (n_i - 1)^{-2} g_i^{-4} \sum_{h=1}^{n_i} \left[ 2(e_{ih} - \bar{e}_i) + g_{ih} \right] g_{ih} \]  \quad (5.30b)

Similarly,

\[ -\frac{1}{2} t \sum_{i=1}^{n_i} x_{ijr} \sum_{j=1}^{n_i} u_{ij} \sigma_v^2 (\sigma_i^2 + n_i \sigma_v^2)^{-1} \]

\[ = -\frac{1}{2} t \sum_{i=1}^{n_i} x_{ijr} u_{ij} \sigma_v^2 (\sigma_i^2 + n_i \sigma_v^2)^{-1} \]

\[ + 2(n_i - 1)^{-1}(\sigma_i^2 + n_i \sigma_v^2)^{-2} \sum_{h=1}^{n_i} \sum_{s=1}^{k} x_{ih} \beta_s \]

\[ - \beta_s (e_{ih} - \bar{e}_i) \]

\[ - T_t + W_t \] \quad (5.31)

where

\[ T_t = -\frac{1}{2} t \sum_{i=1}^{n_i} \sum_{j=1}^{n_i} x_{ijr} u_{ij} \sigma_v^2 (n_i - 1)^{-1}(\sigma_i^2 + n_i \sigma_v^2)^{-2} \sum_{h=1}^{n_i} g_{ih}^2 \] \quad (5.31a)
and

\[ W_t = t \frac{1}{2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} x_{ijr} \sum_{j=1}^{n_1} u_{ij} \sigma_v^2 (n_1 - 1)^{-2} \sigma_i^2 \]

\[ + n_1 \sigma_v^2 \sigma_i^{-1} \sigma_i^2 (n_1 \sigma_v^2 \sigma_i^{-1} (\sum_{h=1}^{n_1} 2(e_{ih} - \bar{e}_i) - g_{ih}) g_{ih})^2. \]

(5.31b)

The terms \( P_t, Q_t, R_t, T_t \) and \( W_t \) all converge to zero in probability. First, consider \( P_t \), which is defined by (5.29b). Let

\[ \theta_i = n_1 \sum_{j=1}^{n_1} x_{ijr} \bar{y}_i. \]

The \( \theta_i \)'s are independently and normally distributed with mean \( 0 \) and variance

\[ \sigma_{\theta_i}^2 = n_1^2 (\sum_{j=1}^{n_1} x_{ijr})^2 (\sigma_v^2 + n_1^{-1} \sigma_i^2). \]

By assumptions (1) - (3), \( \sigma_{\theta_i}^2 \) is bounded for all \( i \), and

\[ \text{Var}(|\theta_i|) < \sigma_{\theta_i}^2. \]
Then,

$$\text{Var}(t \sum_{i=1}^{t} |\theta_i|) < t^{-1} \sum_{i=1}^{t} \sigma_{\theta i}^2 = O(1),$$

so

$$t^{-1} \sum_{i=1}^{t} |\theta_i| = O_p(1).$$

Rewrite $|P_t|$ as

$$|P_t| = \left| t^{-1/2} \sum_{i=1}^{t} \theta_i \left[ \frac{\hat{\sigma}_v^2}{\hat{\sigma}_1^2 + n_1 \sigma_v^2} - \frac{\sigma_v^2}{\hat{\sigma}_1^2 + n_1 \sigma_v^2} \right] \right|$$

$$< t^{-1/2} \sum_{i=1}^{t} |\theta_i| \left| \frac{\hat{\sigma}_v^2(\sigma_v^2 - \sigma_v^2)}{(\hat{\sigma}_1^2 + n_1 \sigma_v^2)(\sigma_v^2 + n_1 \sigma_v^2)} \right|$$

$$< t^{-1/2} \sum_{i=1}^{t} |\theta_i| (8 \sigma_v^2)^{-1} |\hat{\sigma}_v^2 - \sigma_v^2|$$

$$= o(1) \text{ by (5.16)}.$$  

Next, $Q_t$ is considered. From (5.30a),

$$|Q_t| = \left| t^{-1/2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} x_{i1} u_{i1} (n_i - 1)^{-1} \sigma_{i1}^{-4} \sum_{h=1}^{n_1} g_{ih}^2 \right|$$
\[- \frac{1}{2} \sum_{i=1}^{n_i} \sum_{j=1}^{n_i} x_{ij} u_{ij} \sum_{h=1}^{k} (\beta_h - \beta_s) x_{ihs}^2 (n_i - 1)^{-1} \sigma_i^{-4} \]

\[- \sum_{i=1}^{n_i} \sum_{j=1}^{n_i} v_i + e_{ij} \sum_{h=1}^{k} \left( \sum_{h=1}^{n} (\beta_h - \beta_s) x_{ihs} \right)^2 (n_i - 1)^{-1} \sigma_i^{-4} \]

\[- n_u x_u t^{-3} \sum_{i=1}^{n_i} v_i \left( \sum_{j=1}^{n_i} (e_{ij} - \bar{e}_i)^2 \right)^{-2} \sum_{h=1}^{k} \left( \sum_{s=1}^{n} (\beta_s - \beta_h) x_{ihs} \right)^2 \]

\[- \bar{e}_i \sum_{h=1}^{n} x_{ihs} \]

\[\rightarrow 0\]

by Lemmas 5.8, 5.9, and the fact that \( t^2 (\beta_s - \beta_h) = O(1) \).

Similarly for \( T_t \) of (5.31a),

\[ |T_t| = \left| \sum_{i=1}^{n_i} \sum_{j=1}^{n_i} x_{ij} \sum_{h=1}^{k} u_{ij} \sigma_v^2 (n_i - 1)^{-1} \sigma_i^{-2} \right| \]

\[ + n_o a_v^2 \sum_{h=1}^{n} g_{ih}^2 |\]
\[ -\frac{3}{2} t - 3 \varepsilon \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} \left| v_{1j} + e_{1j} \right| \sigma_v^2 (n_1 - 1)^{-1} \sigma_1^{-4} \]

\[ k \sum_{s=1}^{n_1} t^2 (\beta_s - \bar{\beta_s}) x_{1hs}^2 \]

\[ \frac{Q_t}{p} \rightarrow 0 \text{ just as } Q_r \frac{Q_t}{p} \rightarrow 0. \]

Next, consider \( R_t \), which is defined by (5.30b). Let \( \delta > 0 \) be given. Then as in Lemma 5.5,

\[ P(\left| R_t \right| > \delta) < P(A_t) + P(\left| R_t \right| > \delta | A_t |). \]

As before, \( P(A_t) \rightarrow 0 \). Also,

\[ P(\left| R_t \right| > \delta | A_t |) < P\left( t^{-1/2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} \left| x_{1ij} \right| u_{ij} (n_1 - 1)^{-1} \sigma_1^{-4} \sigma_1^{-2} \right. \]

\[ \left. \left( \sum_{h=1}^{n_1} \left( 2(e_{1h} - \bar{e}_{1h}) - g_{1h} g_{1h} \right)^2 > \delta \right) \right) \]

\[ < P \left( b_x t^{-1/2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} \left| u_{ij} \right| \frac{(\sum_{h=1}^{n_1} (e_{1h} - \bar{e}_{1h}))^2 (2t)^{-2\alpha}}{(\sum_{h=1}^{n_1} (e_{1h} - \bar{e}_{1h})^2)^3} \right) \]
using (5.28). For some constant c,

\[ P\left( |R_t| > \delta \right) < P \left\{ c t^{-3\alpha} \sum_{i=1}^{n_t} \frac{|u_{ij}|(e_{im} - \bar{e}_{i.})^2}{(\Sigma(e_{ih} - \bar{e}_{i.})^2)(e_{im} - \bar{e}_{i.})^2} > \delta \right\}, \]

\[ < P\left( c t^{-3\alpha} \sum_{i=1}^{n_t} |v_i|((e_{ih} - \bar{e}_{i.})^2)^{-2} > \frac{\delta}{\epsilon} \right) \]

\[ + P\left( c t^{-3\alpha} \sum_{i=1}^{n_t} \sum_{j=1}^{n_t} |e_{ij}|((e_{ih} - \bar{e}_{i.})^2)^{-2} > \frac{\delta}{\epsilon} \right) \]

\[ \rightarrow 0 \]

by Lemmas 5.8 and 5.9. Hence, \( R_t \rightarrow 0 \).

Similarly for \( W_t \) of (5.31b),

\[ |W_t| < t^{-1/2} \sum_{i=1}^{n_t} \sum_{j=1}^{n_t} |x_{ijr}| \sum_{j=1}^{n_t} |u_{ij}|(\sigma_v(n_t - 1)^{-2}(\sigma_v^2 + n_t \sigma_v^2)^{-2} \]

\[ \times (\sigma_1^2 + n_t \sigma_v^2)^{-1} \left( \sum_{h=1}^{n_t} [2(e_{ih} - \bar{e}_{i.}) - g_{ih}]g_{ih} \right)^2 \]

\[ < n_u x_t t^{-1/2} \sum_{i=1}^{n_t} \sum_{j=1}^{n_t} |u_{ij}|(\sigma_v(n_t - 1)^{-2} \sigma_1^{-4} \sigma_1^{-2} \sum_{h=1}^{n_t} [2(e_{ih} - \bar{e}_{i.}) \]

\[ - g_{ih}g_{ih}]^2 \]

\[ \rightarrow 0 \] just as \( R_t \rightarrow 0 \).
The next step is to find the probability limit of some of the remaining terms in the $r$-th element of $t^{-1/2}x' \mathbf{W}^{-1} \mathbf{y}$. This step follows the proof of Lemma 5.3. For $0 < \delta < 1$,

\[
E\left[|\left(\frac{n_1}{n_1 - 1}\right)^{-\delta} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} x_{ijr} u_{ij} \sum_{h=1}^{n_1} x_{ihs}(e_{ih} - \bar{e}_{1i})|^{1+\delta}\right]
\]

exists if the expectation exists when $\delta = 0$ and $\delta = 1$.

When $\delta = 0$, (5.32) is less than or equal to

\[
(n_1 - 1) \sum_{j=1}^{n_1} |x_{ijr}| \sum_{h=1}^{n_1} x_{ihs} \left\{ \frac{|v_i| |e_{ih} - \bar{e}_{1i}|}{(e_{ij} - \bar{e}_{1i})^2} \right\}
\]

\[
+ (n_1 - 1) \sum_{j=1}^{n_1} |x_{ijr}| \sum_{h=1}^{n_1} x_{ihs} \left\{ \frac{|e_{ij}| |e_{ih} - \bar{e}_{1i}|}{(e_{ij} - \bar{e}_{1i})^2} \right\}.
\]

Using the Cauchy-Schwarz inequality and the independence of $v_i$ and the $e_{ij}$'s,
\[ E\left( \frac{v_i |e_{1h} - \bar{e}_{1i}|}{(\Sigma(e_{ij} - \bar{e}_{1i})^2)^2} \right) < \left[ E\left( v_i^2 (\Sigma(e_{ij} - \bar{e}_{1i})^2)^{-2} \right) \right] E\left( (e_{1h} - \bar{e}_{1i})^2 (\Sigma(e_{ij} - \bar{e}_{1i})^2)^{-2} \right) \]

\[ < \left[ E\left( v_i^2 \right) E\left( (\Sigma(e_{ij} - \bar{e}_{1i})^2)^{-2} \right) \right] E\left( (\Sigma(e_{ij} - \bar{e}_{1i})^2)^{-1} \right) \]

\[ = (c \sigma_v^2 \sqrt{E(U^{-2})E(U^{-1})})^{1/2}, \]

where \( U \sim \chi^2_{n_1 - 1} \), which exists if \( n_1 > 6 \). Similarly,
which is finite. So the expectation of (5.32) exists when $\delta = 0$.

When $\delta = 1$, (5.32) is equal to

$$(n_1 - 1)^2(\sum_{j=1}^{n_1} x_{ijr})^2 E \left\{ \frac{\sum_{h=1}^{n_1} x_{ih} (e_{ih} - \bar{e}_i)^2}{(\sum (e_{ij} - \bar{e}_i)^2)^4} \right\}$$

$$+ (n_1 - 1)^2 E \left\{ \frac{\sum_{j=1}^{n_1} x_{ijr} e_{ij}^2 (\sum_{h=1}^{n_1} x_{ih} (e_{ih} - \bar{e}_i)^2)^2}{(\sum (e_{ij} - \bar{e}_i)^2)^4} \right\}.$$

With assumption (3) and the independence of $v_1$ and the $e_{ij}$'s,

$$E \left\{ \frac{\sum_{h=1}^{n_1} x_{ih} (e_{ih} - \bar{e}_i)^2}{(\sum (e_{ij} - \bar{e}_i)^2)^4} \right\} < n_0 x^2 U \sigma_v^2 E \left\{ \frac{(e_{im} - \bar{e}_i)^2}{(\sum (e_{ij} - \bar{e}_i)^2)^4} \right\}$$

$$< c E(U^{-3}).$$
where $U \sim \chi^2_{n_1} - 1$, which is finite for $n_1 > 8$. Also,

\[
\begin{aligned}
&\mathbb{E} \left\{ \left( \sum_{i=1}^{n_1} x_{ij} \epsilon_{ij} \right)^2 \left[ \sum_{h=1}^{n_1} x_{ih} (e_{ih} - \bar{e}_{i,\cdot}) \right]^2 \right\} \\
&\quad \frac{1}{(\Sigma (e_{ij} - \bar{e}_{i,\cdot})^2)^n} \\
&< c \mathbb{E} \left\{ \frac{e_{ij}^2 (e_{im} - \bar{e}_{i,\cdot})^2}{(\Sigma (e_{ij} - \bar{e}_{i,\cdot})^2)^n} \right\} \\
&< c \mathbb{E} \left\{ \frac{e_{ij}^2}{(\Sigma (e_{ij} - \bar{e}_{i,\cdot})^2)^3} \right\} \\
&= c \mathbb{E} \left\{ \frac{(e_{ij} - \bar{e}_{i,\cdot})^2}{(\Sigma (e_{ij} - \bar{e}_{i,\cdot})^2)^3} \right\} + c \mathbb{E}(\bar{e}_{i,\cdot}^2) \mathbb{E}\{(\Sigma (e_{ij} - \bar{e}_{i,\cdot})^2)^{-3}\} \\
&< c \mathbb{E}(U^{-2}) + c \sigma_i^2 \mathbb{E}(U^{-3})
\end{aligned}
\]

which is finite for $n_1 > 8$. Therefore, (5.32) is finite, and

\[
\begin{aligned}
t^{-1(1+\delta)} \sum_{i=1}^{n_1} \mathbb{E} \left\{ (n_i - 1) - 1 \tilde{\sigma}_i^{-\alpha} \sum_{j=1}^{n_i} x_{ij} \epsilon_{ij} \sum_{h=1}^{n_i} x_{ih} (e_{ih} - \bar{e}_{i,\cdot}) \right\}^{1+\delta} \\
&= o(1).
\end{aligned}
\]
By exercise 4.5 on page 146 of Rao (1973),

\[ t^{-1} \sum_{i=1}^{t} (n_i - 1)^{-1} \sim -4 \sigma_i \sum_{j=1}^{n_i} \sum_{k=1}^{n_i} x_{ijr} u_{ij} x_{ih} x_{hs} (e_{ih} - \bar{e}_i). \]

converges in probability to its expectation. But

\[
E\{(n_i - 1)^{-1} \sim -4 \sigma_i \sum_{j=1}^{n_i} x_{ijr} u_{ij} x_{ih} x_{hs} (e_{ih} - \bar{e}_i)\}
\]

\[
= E\{(n_i - 1)^{-1} \sim -4 \sigma_i \sum_{j=1}^{n_i} x_{ijr} e_{ij} x_{ih} x_{hs} (e_{ih} - \bar{e}_i)\},
\]

\[
= (n_i - 1)^{-1} \sum_{j=1}^{n_i} x_{ijr} \sum_{h=1}^{n_i} x_{ih} E\left\{\frac{(e_{ij} - \bar{e}_i)(e_{ih} - \bar{e}_i)}{\sim -4 \sigma_i}\right\},
\]

using the fact that \( v_i \) is independent of the \( e \)'s and the fact that \( \bar{e}_i \) is independent of \( (e_{ih} - \bar{e}_i) \). Therefore,

\[
- \frac{1}{2} t \sum_{i=1}^{t} \sum_{j=1}^{n_i} x_{ijr} u_{ij} (n_i - 1)^{-1} \sim -4 \sigma_i \sum_{k=1}^{n_i} \sum_{h=1}^{n_i} x_{ih} x_{hs} (\beta - \beta_s)(e_{ih} - \bar{e}_i).
\]

\[
= - \frac{1}{2} t \sum_{i=1}^{t} \sum_{j=1}^{n_i} x_{ijr} (n_i - 1) \sum_{h=1}^{n_i} E\left\{\frac{(e_{ij} - \bar{e}_i)(e_{ih} - \bar{e}_i)}{\sim -4 \sigma_i}\right\} \sum_{s=1}^{n_i} x_{ih} x_{hs} (\beta - \beta_s) - \beta_s + o_p(1).
\]
By similar arguments,

\[
2^{t - 1/2} \sum_{i=1}^{n_i} \sum_{j=1}^{n_j} \xi_{ijr} \xi_{ijv} (n_i - 1)^{-1} (\tilde{\sigma}_i^2 + n_i \tilde{\sigma}_v^2)^{-2} 
\]

\[
\cdot \sum_{h=1}^{n_i} x_{ih} (\beta_s - \tilde{\beta}_s) (e_{ih} - \tilde{e}_i).
\]

\[
= 2^{t - 1/2} \sum_{i=1}^{n_i} \sum_{j=1}^{n_j} \xi_{ijr} \tilde{\sigma}_v^2 (n_i - 1)^{-1} \sum_{h=1}^{n_i} E \left\{ \frac{\xi_{ij} (e_{ih} - \tilde{e}_i)}{(\tilde{\sigma}_i^2 + n_i \tilde{\sigma}_v^2)^2} \right\} \sum_{s=1}^\infty x_{ih} (\beta_s - \tilde{\beta}_s) + o_p(1). 
\]

But since \( \tilde{e}_i \) is independent of \( e_{ih} - \tilde{e}_i \) and of \( \tilde{\sigma}_i^2 \),

\[
E \left\{ \frac{n_i \tilde{e}_i (e_{ih} - \tilde{e}_i)}{(\tilde{\sigma}_i^2 + n_i \tilde{\sigma}_v^2)^2} \right\} = 0.
\]

Then, the r-th element of \(-\frac{1}{2} X' \tilde{\gamma}^{-1} y\) is

\[
2^{t - 1/2} \sum_{i=1}^{n_i} \sum_{j=1}^{n_j} \xi_{ijr} \xi_{ijv} \sum_{h=1}^{n_i} E \left\{ \frac{\xi_{ij} (e_{ih} - \tilde{e}_i)}{(\tilde{\sigma}_i^2 + n_i \tilde{\sigma}_v^2)^2} \right\} 
\]

\[
+ 2^{t - 1/2} \sum_{i=1}^{n_i} \sum_{j=1}^{n_j} \xi_{ijr} \xi_{ijv} (n_i - 1)^{-1} E \left\{ \frac{(e_{ij} - \tilde{e}_i)(e_{ih} - \tilde{e}_i)}{\tilde{\sigma}_i^2} \right\}
\]
Thus, \[ t^{-1/2} x' \hat{\Psi}^{-1} y = t^{-1/2} x' \hat{\Psi}^{-1} y \]

\[ + 2t^{-1/2} x' E\{\text{block diag}[\tilde{\sigma}_i^{-4}(e_{ij} - \bar{e}_{ij})^2] \} (\bar{\beta} - \hat{\beta}) + o_p(1) \] .

(5.32a)

The expectation in expression (5.32a) can be evaluated. The expectations of the diagonal elements \( \tilde{\sigma}_i^{-4}(e_{ij} - \bar{e}_{ij})^2 \) can be evaluated using Lemma 4.1 as follows:

\[ E\{\tilde{\sigma}_i^{-4}(e_{ij} - \bar{e}_{ij})^2\} = \tilde{\sigma}_i^{-2} (n_i - 1) n_i E\left\{ \frac{U_1}{(U_1 + U_2)^2} \right\} , \]

where \( U_1 \sim \chi_1^2 \), \( U_2 \sim \chi_{n_i - 2}^2 \), and \( U_1 \) and \( U_2 \) are independent. By Lemma 4.1,
\[
\sigma_i^{-2} (n_i - 1) n_i E \left( \frac{U_1}{(U_1 + U_2)^2} \right) = \sigma_i^{-2} \frac{(n_i - 1) n_i \Gamma \left( \frac{3}{2} \right) \Gamma \left( \frac{n_i - 1}{2} - 1 \right)}{\Gamma \left( \frac{1}{2} \right)^{-1} \Gamma \left( \frac{n_i - 1}{2} + 1 \right)} = n_i (n_i - 3)^{-1} \sigma_i^{-2} ,
\]

and

\[
E(\sigma_i^{-4} (e_{ij} - \bar{e}_{1.})^2) = n_i (n_i - 3)^{-1} \sigma_i^{-2} .
\]

The expectations of the off-diagonal elements of the second term of (5.32a) can be evaluated by means of a transformation. For each \( i \), there exists a nonsingular transformation matrix \( \Psi_i \) such that

\[
\Psi_i (e_{ij} - \bar{e}_{1.}) = \Psi_i ,
\]

where the first \( (n_i - 1) \) elements of \( \Psi_i \) are independent standard normal random variables and \( \Psi_{n_i} = 0 \). Then,

\[
E(\sigma_i^{-4} (e_{ij} - \bar{e}_{1.})(e_{il} - \bar{e}_{1.})') = \Psi_i^{-1} E(\sigma_i^{-4} \Psi_i (e_{ij} - \bar{e}_{1.})(e_{il} - \bar{e}_{1.})' \Psi_i') \Psi_i^{-1}
\]

\[
\Psi_i^{-1} E(\sigma_i^{-4} \Psi \Psi') \Psi_i^{-1} .
\]
For \( j \neq k \neq n_1 \),

\[
E(\widetilde{\sigma}_i^{-4} w_j w_k) = E(\widetilde{\sigma}_i^{-4} w_j w_k | w_j w_k > 0) + E(\widetilde{\sigma}_i^{-4} w_j w_k | w_j w_k < 0) = E(\widetilde{\sigma}_i^{-4} w_j w_k | w_j w_k > 0) - E(\widetilde{\sigma}_i^{-4} w_j w_k | w_j w_k > 0) = 0,
\]

using the symmetry and independence of the distributions of \( w_j \) and \( w_k \). Therefore,

\[
t^{-1/2} x'\hat{\Sigma}^{-1} y = t^{-1/2} x'\Sigma^{-1} y + 2t^{-1/2} x' \Sigma x(\hat{\beta} - \beta) + o_p(1).
\]

With these lemmas, the asymptotic distribution of \((\hat{\beta} - \beta)\) can be determined.

**Theorem 5.2.** Let model (5.1) - (5.2) hold with assumptions (1) - (4). Let \( \hat{\beta} \) be defined as in (5.21). Then, \( n^{1/2}(\hat{\beta} - \beta) \) has a
limiting normal distribution with mean $0$ and a covariance matrix given by

$$\lim_{n \to \infty} n(X'BX)^{-1} D(X'BX)^{-1},$$

where

$$B = \text{block diag}\{n_i \sigma_i^{-2} [(n_i - 3)^{-1} I_{n_i} - \frac{1}{n_i} \omega_i J_{n_i} x n_i]\}$$

$$= K - \text{block diag}\{\sigma_v \omega_i J_{n_i} x n_i\},$$

$$\omega_i = E[(U_i + n_i^2 \sigma_v^{-2})^{-1}], \quad U_i \sim \chi_{n_i}^2 - 1,$$

$$D = X'G_1 X + 4M + 4M(X'M X)^{-1}(X'G_2 X)(X'M X)^{-1} M,$$

$$M = (X'K X),$$

$$N = \text{block diag}\{I_{n_i} - n_i^{-1} J_{n_i} x n_i\},$$

$$G_1 = E[\tilde{\nu}^{-1} y y'\tilde{\nu}^{-1}],$$

$$G_2 = E[(\tilde{e} - \bar{e})(\tilde{e} - \bar{e})'] = \text{block diag}\{\sigma_i^2 (I_{n_i} - n_i^{-1} J_{n_i} x n_i)\}$$

$$= \text{block diag}\{\sigma_i^2 I_{n_i}\},$$
and $\tilde{V}$ and $K$ are defined as in Lemma 5.10.

**Proof.** The first step is to show that

$$\lim_{t \to \infty} n^{-1}(X'\tilde{V}^{-1}X) = \lim_{t \to \infty} n^{-1}(X'BX).$$

Consider the $(r, s)$ element of $n^{-1}(X'\tilde{V}^{-1}X)$ given by

$$n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{1} x_{ijr} x_{ijs} \tilde{\sigma}_{i}^{-2} - n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{1} x_{ijr} \sum_{h=1}^{1} x_{ihs} \sigma_{v}^{2} (\sigma_{i}^{2} + \sigma_{v}^{2})^{-1}.$$

By assumptions (1) and (3),

$$\left| n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{1} x_{ijr} x_{ijs} (\tilde{\sigma}_{i}^{-2} - \sigma_{i}^{-2}) \right| \leq c \sum_{i=1}^{n} \sum_{j=1}^{1} x_{ijr} x_{ijs} \sigma_{i}^{-2} - \sigma_{i}^{-2}$$

for some positive constant $c$. By Lemma 5.5,

$$c \sum_{i=1}^{n} \sum_{j=1}^{1} x_{ijr} x_{ijs} \sigma_{i}^{-2} - \sigma_{i}^{-2} \to 0.$$
\[
= \lim_{t \to \infty} n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ijr} x_{ijs} n_i (n_i - 3)^{-1} \sigma_i^{-2}.
\]

Similarly,

\[
| n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ijr} x_{ihs} (\frac{\hat{\sigma}_v^2}{\sigma_i^2 + n_i \sigma_v^2} - \frac{\sigma_v^2}{\tilde{\sigma}_i^2 + n_i \sigma_v^2}) | \leq c t^{-1} \sum_{i=1}^{n} \frac{\hat{\sigma}_v^2}{\sigma_i^2 + n_i \sigma_v^2} - \frac{\sigma_v^2}{\tilde{\sigma}_i^2 + n_i \sigma_v^2} \rightarrow 0 \quad \text{by Lemma 5.7.}
\]

Using Lemma 5.6 with

\[
b_i = \sum_{j=1}^{n} x_{ijr} x_{ihs} \sigma_v^2
\]

and \( c_i = n_i \sigma_v^2 \), it follows that

\[
\lim_{t \to \infty} n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ijr} x_{ihs} \sigma_v^2 (\sigma_i^2 + n_i \sigma_v^2)^{-1}
\]

\[
= \lim_{t \to \infty} n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ijr} x_{ihs} \sigma_v^2 n_i \sigma_i^{-2} \omega_i.
\]
Therefore,

$$\lim_{n \to \infty} n^{-1}(X'V^{-1} X) = \lim_{n \to \infty} n^{-1}(X'B X).$$

The second step is to show that for any nonzero real vector \( \lambda \),

$$n^{-1/2} \lambda'(X'V^{-1} y)$$

has a limiting normal distribution. By Lemma 5.10,

$$n^{-1/2} \lambda'(X'V^{-1} y) = n^{-1/2} \lambda'(X'\tilde{V}^{-1} y)$$

$$+ 2(X'K X)(X'N X)^{-1} X'N'(e - \tilde{e})] + o_p(1)$$

$$= n^{-1/2} \sum_{i=1}^{t} \xi_i + o_p(1),$$

where

$$\xi_i = \lambda_r \sum_{r=1}^{k} n_i^1 x_{ij} u_{ij} \tilde{\sigma}_i^{-2} - \sum_{j=1}^{n_i} x_{ij} \sum_{j=1}^{n_i} u_{ij} \tilde{\sigma}_i^{2} + n_i \sigma_i^{2} - 1]$$

$$+ 2 \lambda'(X'K X)(X'N X)^{-1}(X - \tilde{X})'(e - \tilde{e}).$$

The \( \xi_i \)'s are independently distributed. Furthermore, by a basic result in \( L^p \) theory, each \( \xi_i \) has a bounded \( 2 + \delta \) moment if each term in the sum comprising \( \xi_i \) has a bounded \( 2 + \delta \) moment for
0 < \delta < 1. Since each \((e_{ij} - \bar{e}_{i.})\) is normally distributed, and since assumptions (1) - (3) hold, \(2 \lambda'(X'KX)(X'NX)(X_i - \bar{X}_i)'(e_i - \bar{e}_i)\)
has a bounded \(2 + \delta\) moment. The other terms in \(\xi_1\) have bounded \(2 + \delta\) moments if
\[
v_i \left( \sum_{j=1}^{n_1} (e_{ij} - \bar{e}_{ij})^2 \right)^{-1}
\]
and
\[
e_{i1} \left( \sum_{j=1}^{n_1} (e_{ij} - \bar{e}_{i.})^2 \right)^{-1}
\]
have bounded \(2 + \delta\) moments.

Using the independence of \(v_i\) and the \(e_{ij}\)'s,
\[
E\{ |v_i| \left( \sum_{j=1}^{n_1} (e_{ij} - \bar{e}_{ij})^2 \right)^{-1} \}^{2+\delta}
\]
\[
= c E\{ |v_i|^{2+\delta} \} E\{ U^{-(2+\delta)} \},
\]
where \(U \sim \chi^2_{n_1 - 1}\). Since \(v_i\) is normally distributed, and since \(n_1 > 8\), this expectation exists.

For the other term,
\[
E\{ |e_{i1}| \left( \sum_{j=1}^{n_1} (e_{ij} - \bar{e}_{i.})^2 \right)^{-1} \}^{2+\delta}
\]
Using the arguments above,

$$E\left( \frac{|e_{i1} - \bar{e}_{i1}|}{\sum(e_{ij} - \bar{e}_{i1})^2} + \frac{|\bar{e}_{i1}|}{\sum(e_{ij} - \bar{e}_{i1})^2} \right)^{2+\delta} < \infty.$$ 

Let \( \mathbf{L} \) be a lower triangular orthogonal matrix whose first two rows are

$$\begin{pmatrix}
1 & 0 & 0 & 0 & \ldots & 0 \\
0 & 2^{-\frac{1}{2}} & 2^{-\frac{1}{2}} & 0 & \ldots & 0
\end{pmatrix}.$$ 

Then,

$$\mathbf{L}(e_{i1} - \bar{e}_{i1})[\sigma_i^2(n_i - 1)n_i^{-1}]^{-\frac{1}{2}} = \mathbf{z},$$

where

$$z_1 = (e_{i1} - \bar{e}_{i1})[\sigma_i^2(n_i - 1)n_i^{-1}]^{-\frac{1}{2}},$$

$$z_2 = (e_{i2} - e_{i3})[2\sigma_i^2(n_i - 1)n_i^{-1}]^{-\frac{1}{2}}.$$
and all the $z$'s are independently distributed. Since $L$ is orthogonal, it follows that

$$
\sum_{j=1}^{n_1} (e_{ij} - \bar{e}_{i \cdot})^2 [\sigma_i^2 (n_i - 1)n_i^{-1}]^{-1} = \sum_{j=1}^{n_1} z_j^2
$$

> $z_1^2 + z_2^2$

$$
= U_1 + n_i (n_i - 1)^{-1} U_2
$$

> $U_1 + U_2$

where $U_1$ and $U_2$ are independent chi-square random variables, each with one degree of freedom. Then,

$$
E[|e_{i1} - \bar{e}_{i \cdot}|[\Sigma(e_{ij} - \bar{e}_{i \cdot})^2]^{-1}] 
$$

$$
= \left[\sigma_i^2 (n_i - 1)n_i^{-1}\right]^{1/2} E\left\{\frac{|z_1|}{\sum_{j=1}^{n_1} z_j^2}\right\}
$$

$$
< \left[\sigma_i^2 (n_i - 1)n_i^{-1}\right]^{1/2} E\left\{\frac{U_1^{1/2}}{U_1 + U_2}\right\}
$$
which is finite by assumptions (1) - (2) and Lemma 4.1. Thus,

$$|\xi_1|^{2+\delta} = o(1)$$

and

$$\left( n^{-1/2} \right)^{2+\delta} \sum_{i=1}^{t} E[|\xi_1|^{2+\delta}] = o(1).$$

By Liapounov's central limit theorem,

$$n^{-1/2} \sum_{i=1}^{t} \xi_i$$

converges in distribution to a normal random variable.

The asymptotic variance of

$$n^{-1/2} \lambda' \tilde{\mathbf{V}}^{-1} \gamma$$

is

$$n^{-1} \lambda' \left[ \mathbf{X}' \mathbf{E}(\tilde{\mathbf{V}}^{-1} \gamma \gamma' \tilde{\mathbf{V}}^{-1}) \mathbf{X} \right]$$

$$+ 2(\mathbf{X}' \mathbf{E}(\tilde{\mathbf{V}}^{-1} \mathbf{g}(\mathbf{g} - \bar{\mathbf{g}})' \mathbf{N} \mathbf{X})(\mathbf{X}' \mathbf{N} \mathbf{X})^{-1}(\mathbf{X}' \mathbf{K}' \mathbf{X})$$

$$+ 2(\mathbf{X}' \mathbf{K} \mathbf{X})(\mathbf{X}' \mathbf{N} \mathbf{X})^{-1}(\mathbf{X}' \mathbf{W}' \mathbf{E}(\mathbf{g} - \bar{\mathbf{g}})(\mathbf{g} - \bar{\mathbf{g}})' \tilde{\mathbf{V}}^{-1} \mathbf{X})$$

$$+ 4(\mathbf{X}' \mathbf{K} \mathbf{X})(\mathbf{X}' \mathbf{N} \mathbf{X})^{-1}(\mathbf{X}' \mathbf{W}' \mathbf{E}(\mathbf{g} - \bar{\mathbf{g}})(\mathbf{g} - \bar{\mathbf{g}})' \mathbf{N} \mathbf{X})(\mathbf{X}' \mathbf{N} \mathbf{X})^{-1}(\mathbf{X}' \mathbf{K}' \mathbf{X})$$

$$\lambda.$$
if these expectations exist.

The matrix \( E((g - \bar{g})(g - \bar{g})') \) is known to be

\[
G_2 = \text{block diag}\{a^2 (I_{n_1} - n_i^{-1} J_{n_1 	imes n_i})\}
\]

= block diag\( (a_i^2 I_{n_1}) \).

The matrix \( G_1 \) exists if

\[
E(|v_i| |e_{ih} - \bar{e}_{i,1}||\Sigma(e_{ij} - \bar{e}_{i,j})^2|^{-2})
\]

and

\[
E(|e_{ij}| |e_{ih} - \bar{e}_{i,1}||\Sigma(e_{ij} - \bar{e}_{i,j})^2|^{-2})
\]

are finite, which is verified in the proof of Lemma 5.10.

Consider

\[
E(\tilde{\mathbf{\Sigma}}^{-1} g(g - \bar{g})') = E(\tilde{\mathbf{\Sigma}}^{-1} (g - \bar{g})(g - \bar{g})')
\]

The \((1,2)\) block of this matrix is

\[
E(\tilde{\mathbf{\Sigma}}^{-1}_1 (g_{1,1} - \bar{e}_{1,1})(g_{1,2} - \bar{e}_{1,2})') = 0
\]
whenever \( i \neq \ell \). The \( i \)-th diagonal block is

\[
E \left\{ \tilde{V}_i^{-1} (\mathbf{g}_i - \bar{\mathbf{g}}_i)(\mathbf{g}_i - \bar{\mathbf{g}}_i)' \right\}
\]

\[
= E \left\{ \tilde{\sigma}_i^{-2} \left( I - \frac{\sigma_v^2}{\tilde{\sigma}_i^2 + n_i \sigma_v^2} \mathbf{J} \right) (\mathbf{g}_i - \bar{\mathbf{g}}_i)(\mathbf{g}_i - \bar{\mathbf{g}}_i)' \right\}
\]

\[
= E \{ \tilde{\sigma}_i^{-2} (\mathbf{g}_i - \bar{\mathbf{g}}_i)(\mathbf{g}_i - \bar{\mathbf{g}}_i)' \}.
\]

Using Lemma 4.1, the \( j \)-th diagonal element of the \( i \)-th block is

\[
E \left\{ (n_i - 1) \frac{(e_{ij} - \bar{e}_{i,j})^2}{\Sigma (e_{ij} - \bar{e}_{i,j})^2} \right\}
\]

\[
= (n_i - 1)E \left\{ \frac{U_1}{U_1 + U_2} \right\},
\]

where \( U_1 \sim \chi^2_{n_i} \), \( U_2 \sim \chi^2_{n_1 - 2} \), and \( U_2 \) is independent of \( U_1 \),

\[
= (n_i - 1) \frac{\Gamma\left(\frac{3}{2}\right)\Gamma\left(\frac{n_i - 1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n_i - 1}{2} + 1\right)}
\]
For the off-diagonal elements of the i-th block, recall from the proof of Lemma 5.10 the nonsingular transformation

\[ T_l(e_i - e_{i*}) = \underline{w}, \]

where the nonzero elements of \( \underline{w} \) are independent standard normal random variables. Then for \( j \neq k \),

\[
E(\tilde{\sigma}_i^{-2} (e_{ij} - e_{i*})(e_{ik} - e_{i*})) = E(\tilde{\sigma}_i^{-2} w_j w_k) \\
= E(\tilde{\sigma}_i^{-2} w_j w_k \mid w_j w_k > 0) + E(\tilde{\sigma}_i^{-2} w_j w_k \mid w_j w_k < 0) \\
= 0
\]

by the independence and symmetry of the distributions of \( w_j \) and \( w_k \). Therefore,

\[
E(\tilde{\nu}^{-1} (\underline{e} - \underline{\bar{e}})(\underline{e} - \underline{\bar{e}})') = I.
\]

The asymptotic variance of \( n^{-1/2} \lambda' X' \underline{\nu}^{-1} \underline{w} \) reduces to
\[ n^{-1} \lambda' [X'C_1 X + 2(X'H X)(X'N X)^{-1}(X'K'X)] \\
+ 2(X'K X)(X'N X)^{-1}(X'K'X) \\
+ 4(X'K X)(X'N X)^{-1}(X'N C_2 N X)(X'N X)^{-1}(X'K'X)] \lambda \\
= n^{-1} \lambda' [X'C_1 X + 4M \\
+ 4M(X'N X)^{-1}(X'N C_2 X)(X'N X)^{-1} M'] \lambda \\
= n^{-1} \lambda' \lambda. \\
\]

Putting the steps of the proof together,

\[ n^{1/2} (\hat{\beta} - \bar{\beta}) = n^{1/2} (X'v^{-1} X)^{-1} X'v^{-1} y \]

\[ = (n^{-1} X'v^{-1} X)^{-1} n^{-1/2} X'v^{-1} y \]

has an asymptotic normal distribution with covariance matrix

\[ \lim_{t \to \infty} \left( n^{-1} X'B X \right)^{-1} n^{-1} D \left( n^{-1} X'B X \right)^{-1} \]

\[ = \lim_{t \to \infty} n(X'B X)^{-1} D(X'B X)^{-1}. \]
The fact that \( n^{1/2} (\hat{\beta} - \beta) \) has mean zero was established in Theorem 5.1.

3. Effects on the predictor

When the variance components are estimated, then \( \gamma_i \) must also be estimated in the predictor \( \hat{\mu}_i^{(\gamma)} \). Let \( \gamma_i \) be estimated by

\[
\hat{\gamma}_i = \hat{\sigma}_v^2 + n_i^{-1} \hat{\sigma}_1^{-1} \cdot (5.33)
\]

With \( \hat{\gamma}_i \) so defined, the predictor of (5.3) is estimated by

\[
\hat{\mu}_i^{(\gamma)} = \hat{\kappa}_i(p) + \gamma_i (\hat{\gamma}_i - \hat{\kappa}_i) , \quad (5.34)
\]

where \( \hat{\kappa} \) is defined in (5.21).

Even though \( \hat{\kappa} \) has reasonable asymptotic properties, the estimator \( \gamma_i \) is neither unbiased nor consistent under the assumptions of Section V.B. The use of \( \hat{\kappa} \) and \( \hat{\gamma}_i \) in the predictor inevitably causes an increase in the mean squared error of the predictor. Theorem 5.3 analyzes the mean squared error of the predictor when the variance components are estimated.

**Lemma 5.11.** Let model (5.1) - (5.2) and assumptions (1) - (3) of Section V.B hold. Then,

\[
|\hat{\sigma}_i^{-2} - \tilde{\sigma}_i^{-2}| = O_p(t^{-1/2}) ,
\]
where \( \tilde{\sigma}^2_1 \) is defined in (5.12) and \( \tilde{\sigma}^2_1 \) is defined in (5.13a).

**Proof.** The proof is similar to that of Lemma 5.5. Let \( \phi \) be any number in the interval \((0, \frac{1}{2})\). Let \( A_{1t} \) be the event that

\[
\max_j |e_{ij} - \bar{e}_{i|} < t^{-\alpha},
\]

where \( \alpha \) is in the interval \((\phi, \frac{1}{2})\). Let \( A_{t} = A_{1t} \cup A_{2t} \), where \( A_{2t} \) is defined as in Lemma 5.5. Then,

\[
P(\tilde{\sigma}^2_1 - \sigma^2_1 > t^{-\phi} < P(A_{t}) + P(\tilde{\sigma}^2_1 - \sigma^2_1 > t^{-\phi} | A_{t})
\]

\[
< P(A_{t}) + P(\tilde{\sigma}^2_1 - \sigma^2_1 > t^{-\phi} | A_{t})
\]

\[
+ P(\sigma^2_1 > t^{2} q_1 > t^{-\phi}), \quad (5.34a)
\]

where \( b_2 \) is some positive constant and \( q_1 \) is defined in (5.25a).

Considering each term of (5.34a) in turn,

\[
P(A_{1t}) = P(\max_j |e_{ij} - \bar{e}_{i|} < t^{-\alpha})
\]

\[
< P(|e_{i|} - \bar{e}_{i|} < t^{-\alpha})
\]

\[
= P(|w| < t^{-\alpha} \sigma^{-1} \sigma^{-1} (n_{i|} - 1)),
\]
where \( \omega \sim N(0, 1) \),

\[
< t^{-\alpha} \sigma_1^{-1} n_1^{-1} (n_1 - 1)
\]

\[
= o(t^{-\alpha}).
\]

As before, \( P(A_{2t}) \longrightarrow 0 \). So,

\[
\lim_{t \to \infty} P(A_t) = 0.
\]

When \( A_t \) is true, then \( |g_{ij}| < (2t)^{-\alpha} \) for all \( i \) and \( j \), and \( \max_j |e_{ij} - \bar{e}_{ij}| > t^{-\alpha} \) for all \( i \). Then,

\[
|\hat{\sigma}_1^2 - \sigma_1^2| = (n_1 - 1) \frac{\sum_{i=1}^{n_1} [2(e_{ij} - \bar{e}_{ij}) - g_{ij}] g_{ij}}{\sum_{i=1}^{n_1} (e_{ij} - \bar{e}_{ij})^2} \frac{\sum_{i=1}^{n_1} (\hat{e}_{ij} - \bar{e}_{ij})^2}{\sum_{i=1}^{n_1} (\hat{e}_{ij} - \bar{e}_{ij})^2}
\]

\[
= (n_1 - 1) b_1 \frac{\sum_{i=1}^{n_1} |e_{ij} - \bar{e}_{ij}| |g_{ij}|}{(\sum_{i=1}^{n_1} (e_{ij} - \bar{e}_{ij})^2)^2}
\]

\[
< (2t)^{-\alpha}(n_1 - 1)b_1 q_1,
\]

using the results of (5.25) and (5.28). So,
\[ P\left( \left| \tilde{\sigma}_1^2 - \tilde{\sigma}_i^2 \right| > b_2 t^{-\alpha} q_1 \big| \tilde{A}_t \right) \]

\[ < P\left( (2t)^{-\alpha}(n_i - 1)b_1 q_1 > b_2 t^{-\alpha} q_1 \right) \]

\[ = 0 \]

for some positive constant \( b_2 \). Finally, since \( \alpha > \phi \),

\[ P\{b_2 t^{-\alpha} q_1 > t^{-\phi}\} \rightarrow 0 \]

Therefore,

\[ P\left( \left| \tilde{\sigma}_1^2 - \tilde{\sigma}_i^2 \right| > t^{-\phi} \right) \rightarrow 0 \]

But

\[ \left| \tilde{\sigma}_1^2 - \tilde{\sigma}_i^2 \right| = O_p\left( t^{-\phi} \right) \]

is not the strongest statement that can be made. Let \( \phi_1 \) be any number in the interval \((\phi, \frac{1}{2})\). Then, \( \phi_1 = \phi + \delta \) for some positive number \( \delta \). If \( \alpha \) is restricted to the interval \((\phi_1, \frac{1}{2})\), the arguments above still hold, and

\[ \left| \tilde{\sigma}_1^2 - \tilde{\sigma}_i^2 \right| = O_p\left( t^{-\phi_1} \right) = O_p\left( t^{-\phi+\delta} \right) . \]
Let \( \varepsilon > 0 \) be given. By definition there is a number \( M_\varepsilon \) such that

\[
P\left\{ \left| \hat{\sigma}_1^{-2} - \tilde{\sigma}_1^{-2} \right| > t^{-(\phi + \delta)M_\varepsilon} \right\} < \frac{\varepsilon}{2}
\]

for all \( t \). Suppose \( t > (M_\varepsilon/\varepsilon)^{-\delta} \). Then,

\[
P\left\{ \frac{\left| \hat{\sigma}_1^{-2} - \tilde{\sigma}_1^{-2} \right|}{t^{-\phi}} > \varepsilon \right\} = P\left\{ \frac{\left| \hat{\sigma}_1^{-2} - \tilde{\sigma}_1^{-2} \right|}{t^{-(\phi + \delta)}} > t^\delta \varepsilon \right\}
\]

\[
< P\left\{ \frac{\left| \hat{\sigma}_1^{-2} - \tilde{\sigma}_1^{-2} \right|}{t^{-(\phi + \delta)}} > M_\varepsilon \right\}
\]

\[
< \frac{\varepsilon}{2} < \varepsilon.
\]

By the definition of \( o_p \),

\[
\left| \hat{\sigma}_1^{-2} - \tilde{\sigma}_1^{-2} \right| = o_p\left(t^{-\phi}\right).
\]

Now let \( \delta \) be any number in the interval \((0, 1/2)\), and let \( \varepsilon > 0 \) be given. Since

\[
\left| \hat{\sigma}_1^{-2} - \tilde{\sigma}_1^{-2} \right| = o_p\left(t^{-1/2 + \delta}\right),
\]

there is a \( t_0 \) such that
whenever $t > t_o$. Let $M_e = t_o^\delta$, and suppose $t > t_o$. Then,

$$
P\left\{ \left| \hat{\sigma}_1^{-2} - \bar{\sigma}_1^{-2} \right| > t^{-1/2} M_e \right\}$$

$$= P\left\{ \left| \hat{\sigma}_1^{-2} - \bar{\sigma}_1^{-2} \right| > t^{-1/2} + \delta (t^{-1} t_o^{-1}) \delta \right\}

< \varepsilon. $$

Therefore,

$$\left| \hat{\sigma}_1^{-2} - \bar{\sigma}_1^{-2} \right| = o_p(t^{-1/2}).$$

**Theorem 5.3.** Let model (5.1) - (5.2) and assumptions (1) - (4) hold. Then,

$$ \hat{(Y)}_{ij} - \hat{u}_i \hat{(Y)}_{ij} = (\hat{\mu}_i - \mu_i)^2 + (\hat{\gamma}_i - \gamma_i)^2 \hat{u}_i^2 - 2(1 - \gamma_i) v_i$$
where

\[ \hat{\gamma}_1 = \sigma_v^2 (\sigma_v^2 + n_1^{-1} \check{\sigma}_1^2)^{-1}, \]

and

\[ \check{\sigma}_1^2 = (n_1 - 1)^{-1} \sum_{j=1}^{n_1} (e_{1j} - \bar{e}_1)^2. \]

Furthermore,

\[
E\{(\hat{\mu}_1 - \mu_1)^2 + (\hat{\gamma}_1 - \gamma_1)^2 \check{\mu}_1^2 - 2[(1 - \gamma_1) \check{\mu}_1 - \gamma_1 e_1]\},
\]

\[
E\{(\hat{\mu}_1 - \mu_1)^2\}
\]

\[
+ 2(n_1 - 1)^{-1} n_1^{-1} \sigma_1^2 (1 - \gamma_1). \]
Proof. Write

\[ \dot{u}_1 - u_1 = \bar{x}_1(p) \dot{e} + \gamma_1 (-\bar{x}_1 \bar{e} + v_1 + \bar{e}_1. - \bar{x}_1 \bar{e}) \]

\[ - \bar{x}_1(p) \dot{e} - \gamma_1 v_1 \]

\[ = (\bar{x}_1(p) - \gamma_1 \bar{x}_1.)(\dot{\bar{e}} - \bar{e}) - (1 - \gamma_1) v_1 + \gamma_1 \bar{e}_1. \]

\[ = (\bar{x}_1(p) - \gamma_1 \bar{x}_1.)(\dot{\bar{e}} - \bar{e}) - (\gamma_1 - \gamma_1) \bar{x}_1. (\dot{\bar{e}} - \bar{e}) \]

\[ - (1 - \gamma_1) v_1 + \gamma_1 \bar{e}_1. + (\gamma_1 - \gamma_1) \bar{u}_1. \]

\[ = (\bar{x}_1(p) - \gamma_1 \bar{x}_1.)(\dot{\bar{e}} - \bar{e}) + (\bar{x}_1(p) - \gamma_1 \bar{x}_1.)(\dot{\bar{e}} - \bar{e}) \]

\[ - (\gamma_1 - \gamma_1) \bar{x}_1. (\dot{\bar{e}} - \bar{e}) - [(1 - \gamma_1) v_1 \]

\[ - \gamma_1 \bar{e}_1. + (\gamma_1 - \gamma_1) \bar{u}_1. \]

\[ = (\dot{u}_1 - u_1) + (\bar{x}_1(p) - \gamma_1 \bar{x}_1.)(\dot{\bar{e}} - \bar{e}) \]

\[ - (\gamma_1 - \gamma_1) \bar{x}_1. (\dot{\bar{e}} - \bar{e}) + (\gamma_1 - \gamma_1) \bar{u}_1. \, . \]

(5.35)
By Theorem 5.2,

\[ \hat{\beta} - \beta = o_p(n^{-1/2}) = o_p(t^{-1/2}) , \]

and

\[ \hat{\beta} - \tilde{\beta} = (\hat{\beta} - \beta) - (\tilde{\beta} - \beta) \]

\[ = o_p(t^{-1/2}) . \]

So (5.35) becomes

\[ \mu_1 - \mu_1 = (\mu_1 - \mu_1) + (\gamma_1 - \gamma_1)\bar{u}_1, + o_p(t^{-1/2}) . \]

(5.36)

Treating \( \hat{\gamma}_1 \) as a function of one variable, \( \hat{\sigma}_v^2 \), the Taylor series expansion of \( \hat{\gamma}_1 \) is

\[
\hat{\gamma}_1 = \frac{\sigma^2_v}{\sigma^2_v + n_1^{-1} \hat{\sigma}_i^2} + \frac{n_1^{-1} \hat{\gamma}_1^2}{(\sigma^2_v + n_1^{-1} \hat{\sigma}_i^2)^2} (\hat{\sigma}_v^2 - \sigma_v^2) + \text{remainder}
\]

\[ = \frac{\sigma^2_v}{\sigma^2_v + n_1^{-1} \hat{\sigma}_i^2} + o_p(t^{-1/2}) . \]
Furthermore, using the result of Lemma 5.11,

\[
\left| \frac{\sigma_v^2}{\sigma_v^2 + n_1^{-1} \hat{\sigma}_1^2} - \frac{\sigma_v^2}{\sigma_v^2 + n_1^{-1} \tilde{\sigma}_1^2} \right| = n_1^{-1} \sigma_v^2 \left| \frac{\tilde{\sigma}_1^2 - \hat{\sigma}_1^2}{(\sigma_v^2 + n_1^{-1} \hat{\sigma}_1^2)(\sigma_v^2 + n_1^{-1} \tilde{\sigma}_1^2)} \right|
\]

\[
\leq n_1 \sigma_v^2 \left| \frac{\tilde{\sigma}_1^2 - \hat{\sigma}_1^2}{\sigma_v^2 \tilde{\sigma}_1^2 \hat{\sigma}_1^2} \right|
\]

\[
= n_1 \sigma_v^2 \left| \tilde{\sigma}_1^{-2} - \hat{\sigma}_1^{-2} \right|
\]

\[
= O_p(t^{-1/2}) .
\]

Thus, if

\[
\hat{\gamma}_1 = \sigma_v^2 (\sigma_v^2 + n_1^{-1} \tilde{\sigma}_1^2)^{-1},
\]

then

\[
\hat{\gamma}_1 - \tilde{\gamma}_1 = O_p(t^{-1/2}) .
\]

Therefore, (5.36) becomes
and the square is

\[ \hat{\sigma}^2(\gamma) = \mu_1^2 - \mu_1^2 = (\tilde{\mu}_1 - \mu_1)^2 + (\tilde{\gamma}_1 - \gamma_1)^2 \tilde{u}_1^2, \]

\[ - 2[(1 - \gamma_1)v_1 - \gamma_1 e_1,] (\tilde{\gamma}_1 - \gamma_1) \tilde{u}_1, \]

\[ + o_p(t^{-1/2}). \]

The expectation

\[ E[(\tilde{\mu}_1 - \mu_1)^2] + E[(\tilde{\gamma}_1 - \gamma_1)^2 \tilde{u}_1^2], \]

\[ - 2E[(1 - \gamma_1)v_1 - \gamma_1 e_1,] (\tilde{\gamma}_1 - \gamma_1) \tilde{u}_1, \]

(5.39)

will now be evaluated. Now \( e_{ij} \) and \( \tilde{u}_i \) are independent, so \( \tilde{\gamma}_i - \gamma_i \) and \( \tilde{u}_i \) are independent. Then,

\[ E[(1 - \gamma_1)v_1 - \gamma_1 e_1,] (\tilde{\gamma}_1 - \gamma_1) \tilde{u}_1, \]
\[
E\{[(1 - \gamma_i)\eta_i - \gamma_i \bar{e}_i,](\tilde{\gamma}_i - \gamma_i)\bar{u}_i,]\}
\]
\[
= E\{[(1 - \gamma_i)\eta_i \bar{u}_i, - \gamma_i (1 - \gamma_i)\bar{u}_i,](\tilde{\gamma}_i - \gamma_i)\bar{u}_i,\}
\]
\[
= 0 . \tag{5.39a}
\]

The middle term of (5.39) is

\[
E((\tilde{\gamma}_i - \gamma_i)^2 \bar{u}_i,^2) = E(\bar{u}_i,^2)E((\tilde{\gamma}_i - \gamma_i)^2)
\]
\[
= (\sigma_v^2 + n_i^{-1} \sigma_e^2)E((\tilde{\gamma}_i - \gamma_i)^2) .
\]

An overestimate of \(E((\tilde{\gamma}_i - \gamma_i)^2)\) is evaluated here. Write

\[
(\tilde{\gamma}_i - \gamma_i)^2 = \left[\frac{\sigma_v^2}{\sigma_v^2 + n_i^{-1} \sigma_1^2} - \frac{\sigma_1^2}{\sigma_v^2 + n_i^{-1} \sigma_1^2}\right]^2
\]
\[
= \left[\frac{\sigma_v^2 n_i^{-1}(\sigma_1^2 - \sigma_i^2)}{(\sigma_v^2 + n_i^{-1} \sigma_1^2)(\sigma_v^2 + n_i^{-1} \sigma_1^2)}\right]^2
\]
Then,

\[
E((\tilde{y}_1 - y_1)^2) < \frac{n_i^{-2} v(\tilde{\sigma}_1^2)}{(\sigma_v^2 + n_i^{-1} \sigma_i^2)^2}
\]

\[
= \frac{2 \sigma_i^4}{n_i^2(n_i - 1)(\sigma_v^2 + n_i^{-1} \sigma_i^2)^2}.
\]

Therefore,

\[
E((\tilde{y}_1 - y_1)^2 \tilde{u}_i^2) < \frac{2 \sigma_i^4}{n_i^2(n_i - 1)(\sigma_v^2 + n_i^{-1} \sigma_i^2)^2}
\]

\[
= 2(n_i - 1)^{-1} n_i^{-1} \sigma_i^2(1 - \gamma_i). \quad (5.39b)
\]

Substituting the results of (5.39a) and (5.39b) into (5.39) leads to the final result.
4. Estimation with prior estimates

As the number of clusters in the sample increases, the number of variance components to be estimated also increases. For small $t$, the use of prior estimates may be particularly desirable if the number of sample clusters is too small to give an adequate estimate of $\sigma_v^2$.

The procedure for incorporating prior estimates is the same as for the homogeneous nested-error model (Section IV.C). The prior and sample estimates are expressed as the true values plus errors. New estimates are obtained from the resulting linear model by generalized least squares. In this section, formulas for the new estimates are given without derivation.

Two cases of prior information are considered. For the first case, suppose prior estimates are given for $\sigma_v^2$ and $\sigma_i^2$, $i = 1, \ldots, t$. Denote the priors by $\bar{\sigma}_v^2$ and $\bar{\sigma}_i^2$, and let $\bar{W}$ be their covariance.

$$E[(\hat{\mu}_i - \mu_1)^2 + (\hat{\gamma}_i - \gamma_i)^2 \bar{\mu}_i^2 - 2(1 - \gamma_i)\bar{\gamma}_i]$$

$$< E[(\hat{\mu}_i - \mu_1)^2]$$

$$+ 2(n_i - 1)^{-1} n_i^{-1} \gamma_i^2 (1 - \gamma_i)$$
matrix. Let \( \mathbf{W} \) be the estimated covariance matrix of the sample estimates. The new estimates of the variance components are given by

\[
\begin{pmatrix}
\hat{\sigma}_v^2 \\
\hat{\sigma}_e^2
\end{pmatrix} = \begin{pmatrix}
\hat{\sigma}_v^2 \\
\hat{\sigma}_e^2
\end{pmatrix} + \begin{pmatrix}
(\mathbf{W}^{-1} + \mathbf{W}^{-1})^{-1} \\
-1
\end{pmatrix} \begin{pmatrix}
\hat{\sigma}_v^2 - \hat{\sigma}_v^2 \\
\hat{\sigma}_e^2 - \hat{\sigma}_e^2
\end{pmatrix}.
\]  

(5.40)

The covariance matrix of the new estimates is given by

\[
\begin{pmatrix}
\hat{\sigma}_v^2 \\
\hat{\sigma}_e^2
\end{pmatrix} = \begin{pmatrix}
(\mathbf{W}^{-1} + \mathbf{W}^{-1})^{-1}
\end{pmatrix}.
\]  

(5.41)

A test of the consistency of the sample and prior estimates is

\[
\mathcal{F}_{w+1} = (t + 1)^{-1} \begin{pmatrix}
\hat{\sigma}_v^2 - \hat{\sigma}_v^2 \\
\hat{\sigma}_e^2 - \hat{\sigma}_e^2
\end{pmatrix} (\mathbf{W} + \mathbf{W})^{-1} \begin{pmatrix}
\hat{\sigma}_v^2 - \hat{\sigma}_v^2 \\
\hat{\sigma}_e^2 - \hat{\sigma}_e^2
\end{pmatrix}.
\]  

(5.42)

For the second case, a prior estimate \( \hat{\sigma}_v^2 \) and its variance \( \hat{\sigma}_v^2 \) are given. The new estimates of the variance components are given by
\[
\begin{pmatrix}
\hat{\sigma}_v^2 \\
\hat{\sigma}_e^2
\end{pmatrix}
= \begin{pmatrix}
\sigma_v^2 \\
\sigma_e^2
\end{pmatrix}
+ \begin{pmatrix}
\mathbb{W}^{-1} & \mathbb{Q}' \\
\mathbb{Q} & \mathbb{0}
\end{pmatrix}
+ \begin{pmatrix}
\mathbb{W}^{-1} & \mathbb{0} \\
\mathbb{0} & \mathbb{Q}
\end{pmatrix}
\begin{pmatrix}
\sigma_v^2 \\
\sigma_e^2
\end{pmatrix}
\]  
(5.43)

with covariance matrix

\[
\begin{pmatrix}
\hat{\sigma}_v^2 \\
\hat{\sigma}_e^2
\end{pmatrix}
\]  
\[
\begin{pmatrix}
\mathbb{W}^{-1} & \mathbb{Q}' \\
\mathbb{Q} & \mathbb{0}
\end{pmatrix}
\]  
.  
(5.44)

A test for consistency of \(\hat{\sigma}_v^2\) and \(\hat{\sigma}_e^2\) is

\[
t_m = (\mathbb{W} + \sqrt{\mathbb{V}(\hat{\sigma}_v^2)})^{-\frac{1}{2}} |\hat{\sigma}_v^2 - \hat{\sigma}_e^2| .
\]  
(5.45)

### C. Tests of the Model

When the variance components are not known, the question arises as to whether the homogeneous or the heterogeneous nested-error model is correct. If the \(n_i\)’s are small, estimation of the heterogeneous variance components is not feasible. But assuming the \(n_i\)’s are large enough to make the estimation of either model an option, some tests are needed to help determine which model is more appropriate.

Snedecor and Cochran (1980) suggested two tests of equality of variances for independent samples of normally distributed random
variables. The tests may be applied to the independent estimated variance components \( \hat{\sigma}_1^2 \) defined by (5.12).

The first test is Bartlett's test of homogeneity when the degrees of freedom differ. Let

\[
C = 1 + \left[3(t-1)\right]^{-1} \left[ \sum_{i=1}^{t} (n_i - 1)^{-1} - (n - t)^{-1} \right]
\]

and

\[
M = 2.3026\{(n - t)\log[(n - t)^{-1} \sum_{i=1}^{t} (n_i - 1)\hat{\sigma}_i^2] - \sum_{i=1}^{t} (n_i - 1)\log(\hat{\sigma}_i^2) \},
\]

where

\[
2.3026 = \log 10 .
\]

Under the hypothesis that the \( \hat{\sigma}_1^2 \)'s are equal, \( M/C \) has a chi-square distribution with \( (t-1) \) degrees of freedom. The homogeneous nested-error model is rejected if \( M/C \) is too large.

An alternative test is Levene's test of homogeneity [Snedecor and Cochran (1980)]. Levene's test is performed exactly like Bartlett's test except that
is used instead of $\sigma_i^2$. Because it uses absolute differences from the cluster mean instead of squared differences, Levene's test is less likely to reject the hypothesis of homogeneity for large-tailed distributions. However, the distribution of the test statistic is only approximately chi-square.
VI. NUMERICAL EXAMPLES

The predictors of cluster means under the nested-error models are illustrated using data provided by the U.S. Department of Agriculture. Since the predictors were developed for predicting crop areas, the homogeneous nested-error model example illustrates the method for predicting county crop areas. Portions of this example were presented by Battese and Fuller (1981, 1982). The heterogeneous nested-error model example illustrates an application of the theory to the estimation of urban areas.

A. The Homogeneous Nested-Error Model

To illustrate the Battese-Fuller predictor for the homogeneous nested-error model, data for 37 area segments from 12 Iowa counties are considered. These data, which were obtained in 1978, are summarized in Table 1. The numbers of segments sampled are given, together with the reported crop hectares for the sample segments, the number of pixels classified as having corn and soybeans in the sample segments, and the population means of the corn and soybean pixels for each county.

The reported hectares of corn and soybeans were initially regressed on an intercept, linear and quadratic terms of both corn and soybean pixels, and an interaction term. In each case, the quadratic terms, the interaction term, and the terms corresponding to the other crop were not significantly different from zero when tested at the 5 percent level.
Table 1. 1978 Segment Data for Corn and Soybeans in Iowa Counties

<table>
<thead>
<tr>
<th>County</th>
<th>Pop. Sam.</th>
<th>No. of segments for sample segments</th>
<th>Reported hectares</th>
<th>Pixels for sample segments</th>
<th>Mean pixels per segment for counties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cerro Gordo</td>
<td>545 1</td>
<td>165.76 8.09 374 55</td>
<td>295.29 189.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Franklin</td>
<td>564 3</td>
<td>162.08 43.50 361 137 318.21 188.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>152.04 71.43 288 206</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>161.75 42.49 369 165</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hamilton</td>
<td>566 1</td>
<td>96.32 106.03 209 218 300.40 196.65</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hancock</td>
<td>569 5</td>
<td>114.12 99.15 313 190 314.28 198.66</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>100.60 124.56 246 270</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>127.88 110.88 353 172</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>116.90 109.14 271 228</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>87.41 143.66 237 297</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hardin</td>
<td>556 6</td>
<td>88.59 102.59 220 262 325.99 177.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>88.59 29.46 340 87</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>165.35 69.28 355 160</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>104.00 99.15 261 221</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>88.63 143.66 187 345</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>153.70 94.49 350 190</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Humboldt</td>
<td>424 2</td>
<td>185.35 6.47 432 96 290.74 220.22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>116.43 63.82 367 178</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kossuth</td>
<td>965 5</td>
<td>93.48 91.05 221 167 298.65 204.61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>121.00 132.33 369 191</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>109.91 143.14 343 249</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>122.66 104.13 342 182</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>104.21 118.57 294 179</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pocahontas</td>
<td>570 3</td>
<td>92.88 105.26 206 218 257.17 247.13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>149.94 76.49 316 221</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>64.75 174.34 145 338</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Webster</td>
<td>687 4</td>
<td>99.96 144.15 252 303 262.17 247.09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>140.43 103.60 293 221</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>98.95 88.59 206 222</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>131.04 115.58 302 274</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 1. (continued)

<table>
<thead>
<tr>
<th>County</th>
<th>Pop. Sam.</th>
<th>No. of segments</th>
<th>Reported hectares for sample segments</th>
<th>Pixels for sample segments</th>
<th>Mean pixels per segment for counties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Winnebago</td>
<td>402</td>
<td>3</td>
<td>127.07</td>
<td>95.67</td>
<td>355</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>133.55</td>
<td>76.57</td>
<td>295</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>77.70</td>
<td>93.48</td>
<td>223</td>
</tr>
<tr>
<td>Worth</td>
<td>394</td>
<td>1</td>
<td>76.08</td>
<td>103.60</td>
<td>253</td>
</tr>
<tr>
<td>Wright</td>
<td>567</td>
<td>3</td>
<td>206.39</td>
<td>37.84</td>
<td>459</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>108.33</td>
<td>131.12</td>
<td>290</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>118.17</td>
<td>124.44</td>
<td>307</td>
</tr>
</tbody>
</table>

Thus, the model that was used by Battese and Fuller in the estimation procedure is

\[ y_{ij} = \beta_0 + x_{ij} \beta_1 + u_{ij}, \]

\[ u_{ij} = v_i + e_i, \quad i = 1, 2, \ldots, 12, \quad j = 1, \ldots, n_i, \]

where \( y_{ij} \) is the reported hectares of corn (soybeans) in the \( j \)-th sampled segment of the \( i \)-th county; \( x_{ij} \) is the pixels of corn (soybeans) for the same segment; and \( u_{ij} \) is the random error structure assumed to be that described for the homogeneous nested-error model (3.2).

The estimates were actually computed by using SUPER CARP [Hidiroglou, Fuller, and Hickman (1980)], a computer package for analyzing survey data that contains an option for estimation of the
nested-error model by the methods suggested by Fuller and Battese (1973). The following parameter and variance estimates were obtained:

for corn: \( \hat{\beta}_0 = 5.5 \quad \hat{\beta}_1 = 0.388 \)
\[ \text{(13.5)} \quad \text{(0.044)} \]
\( \hat{\sigma}_v^2 = 60 \quad \hat{\sigma}_e^2 = 292 \)
\[ \text{(75)} \quad \text{(84)} \]

for soybeans: \( \hat{\beta}_0 = -3.8 \quad \hat{\beta}_1 = 0.475 \)
\[ \text{(9.3)} \quad \text{(0.040)} \]
\( \hat{\sigma}_v^2 = 250 \quad \hat{\sigma}_e^2 = 184 \)
\[ \text{(142)} \quad \text{(53)} \]

where the numbers in parentheses are the estimated standard errors.

For both corn and soybeans, \( \hat{\beta}_0 \) is not significantly different from zero, which is to be expected when there is no bias in the pixel classification process. Therefore, the second model omits the intercept:

\[ y_{ij} = x_{ij} \beta + u_{ij}, \]
\[ u_{ij} = v_i + e_i, \quad i = 1, 2, \ldots, 12, \quad j = 1, \ldots, n_i, \]

with the same error structure as before. The parameter estimates for this model are:
for corn: \[ \hat{\beta} = 0.405 \quad (0.012) \]
\[ \hat{\sigma}_v^2 = 52 \quad (72) \quad \hat{\sigma}_e^2 = 292 \quad (84) \]

for soybeans: \[ \hat{\beta} = 0.462 \quad (0.021) \]
\[ \hat{\sigma}_v^2 = 238 \quad (137) \quad \hat{\sigma}_e^2 = 184 \quad (53) \]

For both soybean models and for the corn model with an intercept, the estimated slope is not significantly different (at the .05 level) from .45, the ratio of one pixel to one hectare. Thus, if the pixels and reported survey hectares were expressed in the same units, the slope would not be significantly different from 1.

It is interesting to note that \( \hat{\sigma}_v^2 \) is more significant for soybeans than for corn. Another way to look at \( \hat{\sigma}_v^2 \) is through the intra-class correlation coefficient defined by

\[ \rho = \frac{\hat{\sigma}_v^2}{\hat{\sigma}_v^2 + \hat{\sigma}_e^2} . \]

This statistic gives the proportion of the total variance associated with counties. For the models with an intercept, the following estimates of the intra-class correlation coefficient were obtained:

for corn: \[ \hat{\rho} = 0.17 \quad (0.19) \]
for soybeans: $\hat{\rho} = 0.58$ . 

Since $\hat{\rho}$ is moderately large for soybeans, the nested-error model seems appropriate, assuming that the other assumptions are valid. Since $\hat{\rho}$ is nearer to zero for corn, nested-error estimation and prediction will provide less improvement over ordinary least squares for corn than for soybeans.

The assumption of normality was also tested. Full normal plots and Shapiro-Wilk tests of the residuals, both $\hat{\mu}$ and $\hat{\epsilon}$, indicate that the normality assumption is easily satisfied for soybeans. Excluding one observation which appears to be an outlier, the normality assumption is also satisfied for corn.

The residual in the corn data which appears to be an outlier is probably due to an error in the dataset. Two segments in Hardin County have exactly the same number of reported hectares of corn. Most likely the duplication is a copy error. The true number of reported hectares of corn for the segment in question was not available, and the questionable observation was allowed to remain in the dataset.

The estimated within-county variances of individual counties with more than one observation vary from 83 (59) for Hancock County to 930 (1320) for Humboldt County. The standard errors of the individual estimates are quite large due to the small number of observations within counties. Since no county has more than six sample segments, estimation of a heterogeneous nested-error model is not recommended. Considering
all of the above tests, the assumption of a homogeneous nested-error model seems reasonable, especially for the soybean data.

Both sets of estimates were used to obtain predictions for the county mean hectares of corn and soybeans per segment, assuming that the estimated variance components were true. The predictors used were the regression predictor $\mu_{(0)}$, the "best" predictor $\mu_{(y)}$, the adjusted sample mean predictor $\mu_{(1)}$, and the sample county mean

$$ y_i = n_i^{-1} \sum_{j=1}^{n_i} y_{ij}. $$

The predictors are defined in Section III.B.3. The predicted values and their standard errors are given in Tables 2 - 5.

For every county, the sample mean from the June Enumerative Survey has a much larger mean squared error than any of the predictors using satellite data, and the predictor $\mu_{(y)}$ does, in fact, have a smaller estimated mean squared error than either $\mu_{(0)}$ or $\mu_{(1)}$. The mean squared errors of $\mu_{(0)}$ are approximately equal to $\hat{\sigma}_v^2$ in each case, whereas those for $\mu_{(1)}$ are approximately $n_i^{-1} \hat{\sigma}_e^2$. Thus, the estimation of the $\beta$-parameters contributes relatively little to the mean squared errors of the predictors. Furthermore, the predictions obtained for the two models (with and without intercept parameters) are substantially the same.

As stated previously, the Statistical Reporting Service found that the regression predictor, $\mu_{(0)}$, is notably better than the direct expansion predictor. If the nested-error model is appropriate, the best
Table 2. Predicted Mean Hectares of Corn per Segment for the Model with Intercept

<table>
<thead>
<tr>
<th>County</th>
<th>$\gamma_1$</th>
<th>$\tilde{\mu}(0)$</th>
<th>$\tilde{\mu}(Y)$</th>
<th>$\tilde{\mu}(1)$</th>
<th>$\tilde{\gamma}_1$.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cerro Gordo</td>
<td>0.17</td>
<td>120.0</td>
<td>122.6</td>
<td>135.2</td>
<td>165.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(8.3)</td>
<td>(7.8)</td>
<td>(16.9)</td>
<td>(18.8)</td>
</tr>
<tr>
<td>Franklin</td>
<td>0.38</td>
<td>128.9</td>
<td>137.1</td>
<td>150.4</td>
<td>158.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(8.0)</td>
<td>(6.5)</td>
<td>(9.8)</td>
<td>(12.6)</td>
</tr>
<tr>
<td>Hamilton</td>
<td>0.17</td>
<td>122.0</td>
<td>123.6</td>
<td>131.8</td>
<td>96.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(8.3)</td>
<td>(7.8)</td>
<td>(16.8)</td>
<td>(18.8)</td>
</tr>
<tr>
<td>Hancock</td>
<td>0.51</td>
<td>127.3</td>
<td>124.2</td>
<td>121.1</td>
<td>109.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(7.8)</td>
<td>(5.8)</td>
<td>(7.7)</td>
<td>(10.9)</td>
</tr>
<tr>
<td>Hardin</td>
<td>0.55</td>
<td>131.9</td>
<td>131.1</td>
<td>130.5</td>
<td>114.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(7.8)</td>
<td>(5.7)</td>
<td>(7.1)</td>
<td>(10.4)</td>
</tr>
<tr>
<td>Humboldt</td>
<td>0.29</td>
<td>118.2</td>
<td>115.4</td>
<td>108.7</td>
<td>150.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(8.2)</td>
<td>(7.2)</td>
<td>(11.8)</td>
<td>(14.4)</td>
</tr>
<tr>
<td>Kossuth</td>
<td>0.51</td>
<td>121.3</td>
<td>112.7</td>
<td>104.4</td>
<td>110.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(7.8)</td>
<td>(5.8)</td>
<td>(7.6)</td>
<td>(10.9)</td>
</tr>
<tr>
<td>Pocahontas</td>
<td>0.38</td>
<td>105.2</td>
<td>109.3</td>
<td>116.0</td>
<td>102.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(7.9)</td>
<td>(6.5)</td>
<td>(9.7)</td>
<td>(12.6)</td>
</tr>
<tr>
<td>Webster</td>
<td>0.45</td>
<td>107.1</td>
<td>111.7</td>
<td>117.2</td>
<td>117.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(7.9)</td>
<td>(6.2)</td>
<td>(8.6)</td>
<td>(11.6)</td>
</tr>
<tr>
<td>Winnebago</td>
<td>0.38</td>
<td>118.6</td>
<td>116.5</td>
<td>113.1</td>
<td>112.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(8.0)</td>
<td>(6.5)</td>
<td>(9.9)</td>
<td>(12.6)</td>
</tr>
<tr>
<td>Worth</td>
<td>0.17</td>
<td>117.8</td>
<td>113.1</td>
<td>90.3</td>
<td>76.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(8.3)</td>
<td>(7.7)</td>
<td>(17.0)</td>
<td>(18.8)</td>
</tr>
<tr>
<td>Wright</td>
<td>0.38</td>
<td>122.3</td>
<td>123.2</td>
<td>124.6</td>
<td>144.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(8.0)</td>
<td>(6.6)</td>
<td>(9.8)</td>
<td>(12.6)</td>
</tr>
</tbody>
</table>
Table 3. Predicted Mean Hectares of Soybeans per Segment for the Model with Intercept

<table>
<thead>
<tr>
<th>Soybeans</th>
<th>$\gamma_i$</th>
<th>$\sim(0)$</th>
<th>$\sim(\gamma)$</th>
<th>$\sim(1)$</th>
<th>$- y_i$.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cerro Gordo</td>
<td>0.58</td>
<td>86.4</td>
<td>78.2</td>
<td>72.1</td>
<td>8.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(15.6)</td>
<td>(11.0)</td>
<td>(13.7)</td>
<td>(20.8)</td>
</tr>
<tr>
<td>Franklin</td>
<td>0.80</td>
<td>85.6</td>
<td>66.1</td>
<td>61.4</td>
<td>52.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(15.3)</td>
<td>(7.1)</td>
<td>(7.8)</td>
<td>(17.7)</td>
</tr>
<tr>
<td>Hamilton</td>
<td>0.58</td>
<td>89.7</td>
<td>93.3</td>
<td>95.9</td>
<td>106.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(15.7)</td>
<td>(10.5)</td>
<td>(13.6)</td>
<td>(20.8)</td>
</tr>
<tr>
<td>Hancock</td>
<td>0.87</td>
<td>90.7</td>
<td>100.5</td>
<td>101.9</td>
<td>117.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(15.2)</td>
<td>(5.8)</td>
<td>(6.2)</td>
<td>(16.9)</td>
</tr>
<tr>
<td>Hardin</td>
<td>0.89</td>
<td>80.4</td>
<td>74.4</td>
<td>73.7</td>
<td>89.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(15.2)</td>
<td>(5.4)</td>
<td>(5.7)</td>
<td>(16.8)</td>
</tr>
<tr>
<td>Humboldt</td>
<td>0.73</td>
<td>100.9</td>
<td>81.8</td>
<td>74.7</td>
<td>35.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(15.6)</td>
<td>(8.7)</td>
<td>(9.9)</td>
<td>(18.5)</td>
</tr>
<tr>
<td>Kossuth</td>
<td>0.87</td>
<td>93.5</td>
<td>119.3</td>
<td>123.1</td>
<td>117.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(15.2)</td>
<td>(5.7)</td>
<td>(6.1)</td>
<td>(16.9)</td>
</tr>
<tr>
<td>Pocahontas</td>
<td>0.80</td>
<td>13.7</td>
<td>113.2</td>
<td>113.1</td>
<td>118.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(15.2)</td>
<td>(7.1)</td>
<td>(7.8)</td>
<td>(17.7)</td>
</tr>
<tr>
<td>Webster</td>
<td>0.84</td>
<td>13.7</td>
<td>109.9</td>
<td>109.2</td>
<td>113.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(15.1)</td>
<td>(6.3)</td>
<td>(6.8)</td>
<td>(17.2)</td>
</tr>
<tr>
<td>Winnebago</td>
<td>0.80</td>
<td>84.3</td>
<td>97.6</td>
<td>100.8</td>
<td>88.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(15.3)</td>
<td>(7.1)</td>
<td>(7.9)</td>
<td>(17.7)</td>
</tr>
<tr>
<td>Worth</td>
<td>0.58</td>
<td>93.8</td>
<td>87.2</td>
<td>82.3</td>
<td>103.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(15.7)</td>
<td>(10.6)</td>
<td>(13.6)</td>
<td>(20.8)</td>
</tr>
<tr>
<td>Wright</td>
<td>0.80</td>
<td>101.5</td>
<td>112.8</td>
<td>115.6</td>
<td>97.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(15.3)</td>
<td>(7.2)</td>
<td>(8.0)</td>
<td>(17.7)</td>
</tr>
</tbody>
</table>
Table 4. Predicted Mean Hectares of Corn per Segment for the Model Without Intercept

<table>
<thead>
<tr>
<th>County</th>
<th>$\gamma_1$</th>
<th>$\tilde{\mu}_1$</th>
<th>$\tilde{\gamma}_1$</th>
<th>$\tilde{\mu}_1$</th>
<th>$\bar{y}_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cerro Gordo</td>
<td>0.15</td>
<td>119.5</td>
<td>121.7</td>
<td>133.9</td>
<td>165.8</td>
</tr>
<tr>
<td></td>
<td>(7.7)</td>
<td>(7.2)</td>
<td>(16.9)</td>
<td>(18.5)</td>
<td></td>
</tr>
<tr>
<td>Franklin</td>
<td>0.35</td>
<td>128.8</td>
<td>136.2</td>
<td>150.1</td>
<td>158.6</td>
</tr>
<tr>
<td></td>
<td>(7.5)</td>
<td>(6.3)</td>
<td>(9.8)</td>
<td>(12.1)</td>
<td></td>
</tr>
<tr>
<td>Hamilton</td>
<td>0.15</td>
<td>121.6</td>
<td>123.3</td>
<td>133.3</td>
<td>96.3</td>
</tr>
<tr>
<td></td>
<td>(7.9)</td>
<td>(7.3)</td>
<td>(17.3)</td>
<td>(18.5)</td>
<td></td>
</tr>
<tr>
<td>Hancock</td>
<td>0.47</td>
<td>127.2</td>
<td>124.6</td>
<td>121.6</td>
<td>109.4</td>
</tr>
<tr>
<td></td>
<td>(7.3)</td>
<td>(5.7)</td>
<td>(7.7)</td>
<td>(10.4)</td>
<td></td>
</tr>
<tr>
<td>Hardin</td>
<td>0.52</td>
<td>131.9</td>
<td>131.5</td>
<td>131.2</td>
<td>114.8</td>
</tr>
<tr>
<td></td>
<td>(7.3)</td>
<td>(5.4)</td>
<td>(7.1)</td>
<td>(9.9)</td>
<td></td>
</tr>
<tr>
<td>Humboldt</td>
<td>0.26</td>
<td>117.6</td>
<td>114.8</td>
<td>106.9</td>
<td>150.9</td>
</tr>
<tr>
<td></td>
<td>(7.4)</td>
<td>(6.6)</td>
<td>(11.8)</td>
<td>(14.0)</td>
<td></td>
</tr>
<tr>
<td>Kossuth</td>
<td>0.47</td>
<td>120.8</td>
<td>113.0</td>
<td>104.1</td>
<td>110.3</td>
</tr>
<tr>
<td></td>
<td>(7.2)</td>
<td>(5.5)</td>
<td>(7.6)</td>
<td>(10.4)</td>
<td></td>
</tr>
<tr>
<td>Pocahontas</td>
<td>0.35</td>
<td>104.1</td>
<td>108.4</td>
<td>116.6</td>
<td>102.5</td>
</tr>
<tr>
<td></td>
<td>(7.5)</td>
<td>(6.2)</td>
<td>(9.9)</td>
<td>(12.1)</td>
<td></td>
</tr>
<tr>
<td>Webster</td>
<td>0.42</td>
<td>106.1</td>
<td>110.7</td>
<td>117.2</td>
<td>117.6</td>
</tr>
<tr>
<td></td>
<td>(7.3)</td>
<td>(5.8)</td>
<td>(8.5)</td>
<td>(11.1)</td>
<td></td>
</tr>
<tr>
<td>Winnebago</td>
<td>0.35</td>
<td>118.1</td>
<td>116.3</td>
<td>113.1</td>
<td>112.8</td>
</tr>
<tr>
<td></td>
<td>(7.5)</td>
<td>(6.2)</td>
<td>(9.9)</td>
<td>(12.1)</td>
<td></td>
</tr>
<tr>
<td>Worth</td>
<td>0.15</td>
<td>117.2</td>
<td>113.2</td>
<td>90.9</td>
<td>76.1</td>
</tr>
<tr>
<td></td>
<td>(7.8)</td>
<td>(7.3)</td>
<td>(17.2)</td>
<td>(18.5)</td>
<td></td>
</tr>
<tr>
<td>Wright</td>
<td>0.35</td>
<td>121.9</td>
<td>122.5</td>
<td>123.8</td>
<td>144.3</td>
</tr>
<tr>
<td></td>
<td>(7.4)</td>
<td>(6.2)</td>
<td>(9.7)</td>
<td>(12.1)</td>
<td></td>
</tr>
</tbody>
</table>
Table 5. Predicted Mean Hectares of Soybeans per Segment for the Model Without Intercept

<table>
<thead>
<tr>
<th>County</th>
<th>$\gamma_i$</th>
<th>$\mu_i^{(0)}$</th>
<th>$\mu_i^{(\gamma)}$</th>
<th>$\mu_i^{(1)}$</th>
<th>$\hat{\gamma}_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cerro Gordo</td>
<td>0.56</td>
<td>87.7</td>
<td>77.9</td>
<td>70.3</td>
<td>8.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(15.8)</td>
<td>(10.8)</td>
<td>(14.0)</td>
<td>(20.5)</td>
</tr>
<tr>
<td>Franklin</td>
<td>0.80</td>
<td>86.9</td>
<td>66.4</td>
<td>61.1</td>
<td>52.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(15.2)</td>
<td>(7.1)</td>
<td>(7.9)</td>
<td>(17.3)</td>
</tr>
<tr>
<td>Hamilton</td>
<td>0.56</td>
<td>90.9</td>
<td>93.9</td>
<td>96.2</td>
<td>106.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(15.3)</td>
<td>(10.3)</td>
<td>(13.5)</td>
<td>(20.5)</td>
</tr>
<tr>
<td>Hancock</td>
<td>0.87</td>
<td>91.8</td>
<td>100.9</td>
<td>102.3</td>
<td>117.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(14.5)</td>
<td>(5.5)</td>
<td>(5.9)</td>
<td>(16.5)</td>
</tr>
<tr>
<td>Hardin</td>
<td>0.89</td>
<td>81.8</td>
<td>75.0</td>
<td>74.2</td>
<td>89.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(14.4)</td>
<td>(5.1)</td>
<td>(5.4)</td>
<td>(16.3)</td>
</tr>
<tr>
<td>Humboldt</td>
<td>0.72</td>
<td>101.8</td>
<td>81.4</td>
<td>73.6</td>
<td>35.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(15.5)</td>
<td>(8.6)</td>
<td>(9.9)</td>
<td>(18.1)</td>
</tr>
<tr>
<td>Kossuth</td>
<td>0.87</td>
<td>94.6</td>
<td>119.1</td>
<td>122.9</td>
<td>117.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(14.3)</td>
<td>(5.6)</td>
<td>(5.9)</td>
<td>(16.5)</td>
</tr>
<tr>
<td>Pocahontas</td>
<td>0.80</td>
<td>114.2</td>
<td>113.4</td>
<td>113.2</td>
<td>118.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(14.8)</td>
<td>(7.0)</td>
<td>(7.8)</td>
<td>(17.3)</td>
</tr>
<tr>
<td>Webster</td>
<td>0.84</td>
<td>114.2</td>
<td>110.1</td>
<td>109.3</td>
<td>113.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(14.7)</td>
<td>(6.2)</td>
<td>(6.8)</td>
<td>(16.8)</td>
</tr>
<tr>
<td>Winnebago</td>
<td>0.80</td>
<td>85.7</td>
<td>97.4</td>
<td>100.4</td>
<td>88.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(15.2)</td>
<td>(7.1)</td>
<td>(7.9)</td>
<td>(17.3)</td>
</tr>
<tr>
<td>Worth</td>
<td>0.56</td>
<td>94.9</td>
<td>88.1</td>
<td>82.9</td>
<td>103.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(15.2)</td>
<td>(10.3)</td>
<td>(13.4)</td>
<td>(20.5)</td>
</tr>
<tr>
<td>Wright</td>
<td>0.80</td>
<td>102.3</td>
<td>112.4</td>
<td>115.1</td>
<td>97.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(15.2)</td>
<td>(7.2)</td>
<td>(8.0)</td>
<td>(17.3)</td>
</tr>
</tbody>
</table>
predictor, \( \tilde{\mu}_1^{(\gamma)} \), is even better than the regression predictor. By using the best predictor instead of the regression predictor, the mean squared errors decrease by an amount depending on the number of sampled segments in the counties. For corn, the mean squared error decreases by about 12 percent for counties with 1 sampled segment and almost 45 percent for the county with 6 sampled segments. For soybeans, the mean squared error decreases by 53 percent for counties with 1 sampled segment and 87 percent for the county with 6 sampled segments. The relative efficiencies of the best predictor compared to the regression predictor, range from about 1.1 to 1.9 for corn and from 2.0 to 8.0 for soybeans. Thus, the best predictor using the nested-error model is substantially better than the regression estimator, especially when the among-counties variance is relatively large and there are several sample segments observed for the given county.

Table 6 shows the estimated mean squared conditional bias (MSCB) of the predictions for the model with an intercept. The MSCB is defined by (3.13). As expected, the MSCBs of the predictor \( \tilde{\mu}_1^{(1)} \) are near zero, and the MSCBs for the predictor \( \tilde{\mu}_1^{(0)} \) are the largest, some in the neighborhood of \( \hat{\sigma}_v^2 \).

The standard errors of the predictors in Tables 2 - 5 and the MSCBs of Table 6 were computed by assuming that the estimated variance components were true. Table 7 shows the effect of estimating the variance components for the model with an intercept. The mean squared errors in the "Known" columns correspond to the standard errors in
Table 6. Mean Squared Conditional Biases of the Predictors for the Model with Intercept

<table>
<thead>
<tr>
<th>County</th>
<th>Corn</th>
<th></th>
<th>Soybeans</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>$\gamma_1$</td>
<td>1</td>
</tr>
<tr>
<td>Cerro Gordo</td>
<td>61.5</td>
<td>42.9</td>
<td>1.22</td>
</tr>
<tr>
<td>Franklin</td>
<td>54.4</td>
<td>21.0</td>
<td>0.09</td>
</tr>
<tr>
<td>Hamilton</td>
<td>61.4</td>
<td>43.0</td>
<td>1.64</td>
</tr>
<tr>
<td>Hancock</td>
<td>52.5</td>
<td>13.0</td>
<td>0.18</td>
</tr>
<tr>
<td>Hardin</td>
<td>51.6</td>
<td>10.7</td>
<td>0.32</td>
</tr>
<tr>
<td>Humboldt</td>
<td>58.9</td>
<td>32.3</td>
<td>2.32</td>
</tr>
<tr>
<td>Kossuth</td>
<td>52.0</td>
<td>12.7</td>
<td>0.05</td>
</tr>
<tr>
<td>Pocahontas</td>
<td>51.3</td>
<td>20.3</td>
<td>0.24</td>
</tr>
<tr>
<td>Webster</td>
<td>51.7</td>
<td>15.5</td>
<td>0.00</td>
</tr>
<tr>
<td>Winnebago</td>
<td>55.4</td>
<td>21.1</td>
<td>0.00</td>
</tr>
<tr>
<td>Worth</td>
<td>61.1</td>
<td>42.1</td>
<td>0.26</td>
</tr>
<tr>
<td>Wright</td>
<td>55.3</td>
<td>22.1</td>
<td>0.51</td>
</tr>
</tbody>
</table>
Table 7. Mean Squared Errors of the Best Predictors with Known and Estimated Variance Components

<table>
<thead>
<tr>
<th>County</th>
<th>Corn Known</th>
<th>Corn Estimated</th>
<th>Soybeans Known</th>
<th>Soybeans Estimated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cerro Gordo</td>
<td>60.1</td>
<td>72.5</td>
<td>119.9</td>
<td>130.3</td>
</tr>
<tr>
<td>Franklin</td>
<td>42.6</td>
<td>57.1</td>
<td>50.6</td>
<td>53.4</td>
</tr>
<tr>
<td>Hamilton</td>
<td>60.1</td>
<td>72.5</td>
<td>111.1</td>
<td>121.4</td>
</tr>
<tr>
<td>Hancock</td>
<td>34.1</td>
<td>45.2</td>
<td>33.8</td>
<td>35.0</td>
</tr>
<tr>
<td>Hardin</td>
<td>32.0</td>
<td>41.5</td>
<td>29.3</td>
<td>30.1</td>
</tr>
<tr>
<td>Humboldt</td>
<td>52.3</td>
<td>67.4</td>
<td>76.5</td>
<td>81.6</td>
</tr>
<tr>
<td>Kossuth</td>
<td>33.2</td>
<td>44.3</td>
<td>32.7</td>
<td>33.9</td>
</tr>
<tr>
<td>Pocahontas</td>
<td>42.8</td>
<td>57.3</td>
<td>50.3</td>
<td>53.1</td>
</tr>
<tr>
<td>Webster</td>
<td>38.0</td>
<td>50.9</td>
<td>39.5</td>
<td>41.3</td>
</tr>
<tr>
<td>Winnebago</td>
<td>42.6</td>
<td>57.1</td>
<td>50.8</td>
<td>53.7</td>
</tr>
<tr>
<td>Worth</td>
<td>59.6</td>
<td>72.0</td>
<td>111.7</td>
<td>122.0</td>
</tr>
<tr>
<td>Wright</td>
<td>43.2</td>
<td>57.6</td>
<td>52.3</td>
<td>55.1</td>
</tr>
</tbody>
</table>

Tables 2 and 3 and were computed according to the formulas in Section III.B.3. The mean squared errors in the other columns were computed according to Theorem 4.1 for predictors using estimated variance components.

The increases in the mean squared errors of corn predictors range from 21 percent to 34 percent, but the increases are all less than 10 percent for soybean predictors. This is due to the fact that \( \hat{\sigma}_v^2 \) is much larger for soybeans than for corn, and the increase term of Theorem
4.1 is a decreasing function of $\sigma^2_v$. Counties with a single observation have the smallest percentage increases for corn and the largest percentage increases for soybeans. In fact, the percentage increase appears to be a monotone decreasing function of sample size for soybeans, but not for corn.

Walker and Sigman (1982) illustrated the stratification procedures of Section III.C by predicting crop areas in six counties of eastern South Dakota. From their results, it may be inferred that the ratio $R = \sigma^2_v / \sigma^2_e$ is approximately .47 with an approximate standard error of .36 for corn in the stratum of PSUs which are at least 75 percent cultivated. Since the 12 Iowa counties of this example are from the same stratum and are not far from eastern South Dakota, the approximate ratio from the South Dakota data may be used as prior information. The test of consistency of the sample and prior estimates is

$$t_\infty = .53$$

which is not significant.

When the prior ratio is combined with the sample ratio according to (4.48), the new estimates are:

$$\tilde{R} = .33 \quad \tilde{\sigma}^2_e = 271$$
$$\quad (.25) \quad \quad \quad \quad (74)$$

$$\tilde{\sigma}^2_v = 89 \quad \quad \quad \quad (56)$$
With these variance components, the other parameters of the model are estimated by:

\[
\hat{\beta}_0 = 5.1 \quad \hat{\beta}_1 = 0.389 .
\]

(13.5) (0.043)

These estimates are all similar to the estimates obtained without prior information.

The best predictors of the mean hectares of corn per segment were computed according to (4.12) using the new estimated parameters. In Table 8, these predictors are compared with the best predictor obtained according to (4.13) with \( \hat{\gamma}_1 \) defined by (4.14), using only the sample data. The standard errors were computed assuming that the new estimated variance components are true. Notice that the standard errors in Table 8 are larger than the standard errors in Table 2. This is due to the fact that the new estimate of \( \sigma^2 \), which is presumably closer to the true \( \sigma^2 \), is larger than the sample estimate of \( \sigma^2 \). Although the apparent standard errors have increased, the true standard errors have actually decreased. The amount of the decrease depends on the consistency of the sample and prior estimates, and on the size of their variances.

B. The Heterogeneous Nested-Error Model

To illustrate small area prediction using the heterogeneous nested-error model, data from the 1982 National Resources Inventory (NRI) are considered. The NRI is a nationwide survey of natural resources sponsored by the Soil Conservation Service of the U.S. Department of
Table 8. Predicted Mean Hectares of Corn per Segment With and Without Prior Information

<table>
<thead>
<tr>
<th>County</th>
<th>With priors</th>
<th>Sample only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cerro Gordo</td>
<td>123.8 (8.8)</td>
<td>123.8 (9.5)</td>
</tr>
<tr>
<td>Franklin</td>
<td>139.6 (7.0)</td>
<td>139.6 (7.7)</td>
</tr>
<tr>
<td>Hamilton</td>
<td>124.5 (8.8)</td>
<td>124.4 (9.5)</td>
</tr>
<tr>
<td>Hancock</td>
<td>123.6 (6.1)</td>
<td>123.5 (6.6)</td>
</tr>
<tr>
<td>Hardin</td>
<td>131.1 (5.8)</td>
<td>131.0 (6.3)</td>
</tr>
<tr>
<td>Humboldt</td>
<td>114.4 (8.0)</td>
<td>114.5 (8.8)</td>
</tr>
<tr>
<td>Kossuth</td>
<td>110.8 (6.0)</td>
<td>110.8 (6.6)</td>
</tr>
<tr>
<td>Pocahontas</td>
<td>110.6 (7.0)</td>
<td>110.6 (7.7)</td>
</tr>
<tr>
<td>Webster</td>
<td>112.9 (6.5)</td>
<td>112.8 (7.1)</td>
</tr>
<tr>
<td>Winnebago</td>
<td>115.9 (7.0)</td>
<td>115.9 (7.7)</td>
</tr>
<tr>
<td>Worth</td>
<td>111.0 (8.7)</td>
<td>111.0 (9.4)</td>
</tr>
<tr>
<td>Wright</td>
<td>123.5 (7.1)</td>
<td>123.5 (7.8)</td>
</tr>
</tbody>
</table>

Agriculture. One of the many items of interest in the survey is the amount of urban and built-up land. In this example, 506 primary sampling units (PSUs) from 5 Alabama counties are used. For each PSU,
the number of urban and built-up acres (in a unit of at least 40 acres) and the total acreage are known. Also available are census-type county base data estimates of percentages of urban acres for the five Alabama counties. These county base percentages serve as the auxiliary variable, both for individual PSUs and for entire counties. The county means are given in Table 9. It is desired to predict the mean percentage of urban and built-up acres per PSU in each of the five counties.

The model that was used in the prediction process is

\[ y_{ij} = \beta x_{ij} + u_{ij} + e_{ij}, \quad i = 1, \ldots, 5, \quad j = 1, \ldots, n_i, \]

where

\[ y_{ij} = \frac{\text{urban acres}}{\text{total acres}} \times 100 \]

for the j-th PSU in the i-th county, \( x_{ij} \) is the county base data percentage of urban acres, and \( u_{ij} \) is the random error with the heterogeneous nested-error structure described by (5.2). This model was used even though the sample distribution of \( y \) is heavily skewed toward zero, indicating nonnormality. Using the SAS computer package, [SAS Institute Inc. (1982)], the following variance component estimates were obtained:
The terms are clearly not homogeneous. Since the number of degrees of freedom associated with each $\hat{\sigma}_1^2$ is large, the differences cannot be attributed solely to sampling error. The fact that the estimated variance components vary so much among counties is not surprising. Jefferson County contains the city of Birmingham and is by far the most populous county in the state. At the other extreme, Blount and St. Clair Counties have no towns with a population of 10,000 or more. Consequently, the vast majority of sample segments in the less populous counties have no urban acres. Blount County has only three sample segments with reported urban acres. St. Clair County has six sample segments with reported urban acres, but of these nonzero observations, only two have more than 35 percent urban acres. Therefore, the sample variance within Blount and St. Clair Counties should be relatively small. Jefferson County has 21 nonzero
observations, many of which have at least 80 percent urban acres. Both
the number and magnitude of the nonzero observations in Jefferson County
will cause the Jefferson County variance to be relatively large.
Bartlett's test of homogeneity gives

$$\chi^2_4 = 454.9,$$

which is highly significant. Thus, the heterogeneous nested-error model
is more appropriate for these data than is the homogeneous nested-error
model.

Unfortunately, the number of counties is small, and the variation
among counties is small relative to the variation within counties. Not
only does the estimate of $\sigma^2_v$ have a large standard error, but the
estimate is negative. Ordinarily, one would set the negative estimate
of $\sigma^2_v$ equal to zero and set $\gamma_i = 0$ for all $i$, which is not very
interesting. For the sake of illustration, suppose a prior estimate of
$\sigma^2_v$ is available, namely, $\tilde{\sigma}^2_v = 3.0$ with a standard error of 0.7.
After combining $\tilde{\sigma}^2_v$ with the unbiased $\hat{\sigma}^2_v$ according to (5.43), the new
estimated parameters are:

$$\tilde{\sigma}^2_v = 2.55 \quad \sigma^2_1 = 165$$
$$\hat{\beta} = 0.420 \quad \sigma^2_2 = 207$$
$$\tilde{\sigma}^2_2 = 431$$
The estimates of the $\sigma_i^2$ components are very similar to the previous estimates, and the new estimate of $\sigma_v^2$ is close to the prior estimate.

Assuming these estimated variance components were true, the mean percentages of urban acres per PSU for each county were predicted according to (5.3). The results are included in Table 9.

The predictions and sample means, $\bar{y}_{1.}$, are very similar, but the standard errors of the predictors are consistently smaller than those of

<table>
<thead>
<tr>
<th>County</th>
<th>Sample size</th>
<th>County base data</th>
<th>Sample %</th>
<th>Predicted %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n_i)</td>
<td>(\bar{x}_{1.})</td>
<td>(\bar{y}_{1.})</td>
<td></td>
</tr>
<tr>
<td>Blount</td>
<td>111</td>
<td>4.8</td>
<td>2.1</td>
<td>2.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(2.0)</td>
<td>(0.98)</td>
</tr>
<tr>
<td>Calhoun</td>
<td>100</td>
<td>11.0</td>
<td>4.1</td>
<td>4.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(2.1)</td>
<td>(1.1)</td>
</tr>
<tr>
<td>Etowah</td>
<td>89</td>
<td>15.2</td>
<td>6.4</td>
<td>6.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(2.7)</td>
<td>(1.5)</td>
</tr>
<tr>
<td>Jefferson</td>
<td>92</td>
<td>39.3</td>
<td>17.2</td>
<td>16.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(4.0)</td>
<td>(2.9)</td>
</tr>
<tr>
<td>St. Clair</td>
<td>114</td>
<td>6.3</td>
<td>2.2</td>
<td>2.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1.9)</td>
<td>(0.82)</td>
</tr>
</tbody>
</table>
The predictions are less than half of the county base data estimates, $\bar{y}_i$. One assumption of the prediction procedure is that the observed $y$-values are true. The $x$-variable is considered to be an auxiliary variable, a supplement to the $y$-variable in making predictions. For this reason, $\bar{y}_i$ has more influence on the predictor than $\bar{x}_i$.

The mean squared errors of the predictors using known and estimated variance components are compared in Table 10. The known mean squared errors, computed according to (5.8), assume that the estimated variance components using the prior information are true, and they correspond to the standard errors of the predictors in Table 9. The estimated mean squared errors are the approximate mean squared errors of Theorem 5.3, assuming that the variance components are estimated. The increases in mean squared errors due to the estimation of the variance components are all less than 4 percent. Recall that the percentage increases in the

<table>
<thead>
<tr>
<th>County</th>
<th>Known</th>
<th>Estimated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blount</td>
<td>0.957</td>
<td>0.967</td>
</tr>
<tr>
<td>Calhoun</td>
<td>1.28</td>
<td>1.30</td>
</tr>
<tr>
<td>Etowah</td>
<td>2.24</td>
<td>2.32</td>
</tr>
<tr>
<td>Jefferson</td>
<td>8.41</td>
<td>8.66</td>
</tr>
<tr>
<td>St. Clair</td>
<td>0.670</td>
<td>0.674</td>
</tr>
</tbody>
</table>
mean squared errors of Table 7 are much larger. In the case of the homogeneous nested-error model, the squared error in Theorem 4.1 is percentage increases in the mean squared errors of Table 7 are much larger. In the case of the homogeneous nested-error model, the squared error in Theorem 4.1 is approximated to $O_p(\max[t^{-1}, (n-t)^{-1}])$, whereas the squared error is approximated to $O_p(t^{-1/2})$ in Theorem 5.3 for the heterogeneous nested-error model. Since $t$ is small in this example, the $O_p(t^{-1/2})$ term that is omitted in the expectation of the squared error in Theorem 5.3 may be substantial. The penalty term in Theorem 5.3 which causes the increase is an increasing function of $\hat{\sigma}_1^2$ and a decreasing function of sample size. In Table 10, the largest increase in mean squared error occurs in Jefferson County, which also has the largest estimated variance $\hat{\sigma}_1^2$. Etowah County has the fewest sample segments and the largest percentage increase in mean squared error.
VII. SUMMARY

Many small area estimation techniques have been proposed for a variety of small area problems. The techniques differ considerably in their assumptions, use of available data, and applications. All small area techniques have the same general goal - to provide reasonable estimates for small subgroups of a population.

The particular small area estimation problem considered here is that of estimating hectares of crops at the county level using data from a national agricultural survey and auxiliary data from LANDSAT satellites. To estimate the county crop mean hectares, Battese and Fuller (1981, 1982) proposed a prediction procedure which assumes a homogeneous nested-error model:

\[ y_{ij} = x_{ij} \beta + u_{ij}, \]

\[ u_{ij} = v_{i} + e_{ij}, \quad i = 1, \ldots, t, \quad j = 1, \ldots, n_{i}, \]

where \( y_{ij} \) is the survey reported hectares of the crop for the \( j \)-th sample segment of the \( i \)-th county, \( x_{ij} \) is a \((1 \times k)\) vector of auxiliary variables including the satellite estimate of the crop acreage for the \( j \)-th segment of the \( i \)-th county, and \( \beta \) is a \((k \times 1)\) vector of fixed parameters. The error, \( u_{ij} \), has a random county component, \( v_{i} \), and a random error, \( e_{ij} \). The random components are normally and
independently distributed with zero means and covariance structure given by

$$E(u_{ij}^T u_{jk}) = \begin{cases} 
\sigma_v^2 + \sigma_e^2, & i = \ell, j = k \\
\sigma_v^2, & i = \ell, j \neq k \\
0, & i \neq \ell 
\end{cases}$$

Assuming that the variance components are known, the best predictor of the mean hectares per segment of a particular crop in county \( i \) is

$$\tilde{\mu}_i = \bar{x}_i(p) \bar{\beta} + \gamma_i (\bar{y}_i - \bar{x}_i \bar{\beta}),$$

where \( \bar{x}_i(p) \) is the mean of the auxiliary variables for the entire \( i \)-th county, \( \bar{\beta} \) is the generalized least squares estimator of \( \beta \), and

$$\gamma_i = \sigma_v^2 (\sigma_v^2 + n_i^{-1} \sigma_e^2)^{-1}.$$  

The difference

$$\bar{y}_i - \bar{x}_i \bar{\beta}$$

is an estimate of the county component, \( \nu_i \), and \( \gamma_i \) is a shrinker. Thus, the estimator is a shrinkage estimator, and it has the form of a James-Stein estimator. The predictor is also the best linear unbiased
predictor of the county mean with the random county component. When $\beta$ is known, the predictor is the expected value of the county mean, conditioned on the observable error means, $\bar{u}_i$. The predictor is biased, conditioned on the random county components. The predictor and its properties are discussed in Section III.B.

If the variance components are not known, estimates of the variance components are used in the prediction procedure. The fitting-of-constants estimators of the variance components are used in this study. The asymptotic properties of the estimated variance components and the estimated generalized least squares vector $\hat{\beta}$ are known. The use of estimated variance components in the Battese-Fuller predictor causes an increase in the mean squared error of the predictor. The increase in the mean squared error is the subject of Theorem 4.1. With estimated variance components, the predictor is an empirical Bayes predictor, and Theorem 4.1 is based on the empirical Bayes work of Efron and Morris (1973).

Prior estimates of the variance components are helpful in making current estimates of the variance components. Prior estimates and estimates based on the sample data may be combined by a generalized least squares procedure to obtain improved estimates of the variance components.

A prediction procedure analogous to the Battese-Fuller procedure is developed in Chapter V for the heterogeneous nested-error model. The heterogeneous model differs from the homogeneous model in that the
within-county variance component is different for each county. The heterogeneous nested-error model may be more appropriate for states whose counties are agriculturally diverse. When the variance components are known, the expressions for the i-th county predictor, its mean squared error, and its mean squared conditional bias are identical to the corresponding expressions for the homogeneous nested-error model, except that the individual within-county variance, \( \sigma_i^2 \), is used in place of the pooled within-county variance, \( \sigma_e^2 \). For example, the best predictor of the county mean is

\[
\hat{\mu}_i(y) = \bar{x}_{i(p)} \tilde{\beta} + \gamma_i (\bar{y}_i - \bar{x}_{i(p)} \tilde{\beta}),
\]

where

\[
\gamma_i = \frac{\sigma_i^2 \sigma_y^2 + n_i^{-1} \sigma_i^2 \sigma_{vi}^2}{\sigma_y^2}.
\]

\( \tilde{\beta} \) is the generalized least squares estimator of \( \beta \) for the heterogeneous model, and \( \bar{x}_{i(p)} \) is defined as before. Again, the predictor is the best linear unbiased predictor of the county mean.

If the variance components must be estimated, estimators similar to the fitting-of-constants estimators may be used. The asymptotic properties of the estimators are derived with the added assumption of bounded sample sizes within counties. Theorems 5.1 and 5.2 and the supporting lemmas are extensions of the work on the estimation of \( \tilde{\beta} \) by
Fuller and Battese (1973) for the homogeneous nested-error model, and by Fuller and Rao (1978) for a linear regression model with a diagonal covariance matrix with unequal variances. Theorem 5.3 discusses the mean squared error of the predictor when the variance components are estimated. Theorem 5.3 is similar to Theorem 4.1 for the homogeneous nested-error model, but Theorem 5.3 estimates the increase in the mean squared error of the estimated predictor with less precision than Theorem 4.1.

The homogeneous nested-error model example in Chapter VI shows the Battese-Fuller predictor to be a reasonable solution to the problem of estimating crop hectares at the county level. The Statistical Reporting Service of the U.S. Department of Agriculture is implementing the computer software for the Battese-Fuller prediction procedure, including the estimation of the variance components with options for the inclusion of prior estimates of the variance components. Walker and Sigman (1982) of the Statistical Reporting Service extended the procedure for counties whose sample segments fall within different land use strata. The heterogeneous nested-error model prediction procedure is suggested for regions in which the amount of variation within counties differs among neighboring counties.
VIII. REFERENCES


IX. ACKNOWLEDGEMENTS

I wish to express my appreciation to Professor Wayne A. Fuller for his expert and patient direction of my Ph.D. program. His experience and insight have been invaluable help, both in the preparation of the dissertation and in my preparation for a career.

A special thanks is given to Professors J. Jeffery Goebel and George E. Battese for their assistance and supervision of my work during earlier stages of this research project.

I am grateful to the Survey Section for the use of its word processor. I was extremely fortunate to have Mrs. Jane Stowe as typist, and I thank her for efficiently and cheerfully typing the many drafts of the dissertation.

With gratitude I acknowledge the bequest from my grandmother, Besse Harter. Her gift for my education provided for the final production of the dissertation.

My sincere thanks go to my family and friends, who have been a constant source of encouragement throughout my graduate career. Their friendship and moral support have made graduate school much more enjoyable.

This research was supported by Research Agreement 58-319T-1-0054X with the Statistical Reporting Service of the U.S. Department of Agriculture.
X. APPENDIX: AN ALTERNATIVE ESTIMATOR OF $\gamma_1$ FOR THE HOMOGENEOUS NESTED-ERROR MODEL

In Chapter IV, the fitting-of-constants estimators were used to estimate the variance components of the homogeneous nested-error model. The ratio $\gamma_1 = (\sigma^2 + n_1^{-1} \sigma^2)^{-1} \sigma^2$ was estimated by a function of the estimated variance components. In Theorem 4.1, the mean squared error of the predictor with estimated variance components was approximated by the mean squared error of the predictor with known variance components plus a penalty term for estimating the variance components. For the special case in which $x_{ij} = 1$ and $n_i = r$ for all $i$, it was shown that the penalty term matches the exact penalty term (4.39) given by Peixoto (1982). Unfortunately, the proof of Theorem 4.1 required the rather strong assumption (4).

In this section, an alternative estimator of $\gamma_1$ is considered. A theorem analogous to Theorem 4.1 is presented, without the strong assumption about the sequence $\{n_i\}$. It will be shown that for the special case in which $x_{ij} \equiv 1$ and $n_i = r$ for all $i$, the penalty term matches Peixoto's result.

Let model (3.1) - (3.2) hold. Let

$$u_{ij} = n_i^{-1} \sum_{j=1}^{n_i} u_{ij}.$$  

Since
$$E\{\bar{u}_o^2\} = \sigma_v^2 + n\sigma_e^2,$$

it follows that

$$z_i = \bar{u}_o^2 - n\sigma_e^2$$

(10.1)

is an unbiased estimator of $\sigma_v^2$, where $\sigma_e^2$ is defined by (4.1). Let

$$\hat{\sigma}_v^2 = \frac{\sum_{i=1}^{t} \hat{w}_i^{-1} z_i}{\sum_{i=1}^{t} \hat{w}_i^{-1}},$$

(10.2)

where

$$\hat{w}_i = \hat{\sigma}_v^2 + n\sigma_e^2,$$

and $\hat{\sigma}_v^2$ is defined by (4.4).

The estimator $\hat{\sigma}_v^2$ can be thought of as an estimated generalized least squares estimator. Write

$$z = \lambda \sigma_v^2 + \xi,$$

where $\lambda$ is a column of ones, and $E(\xi) = 0$. Suppose, artificially, that
$$V(e) = \Omega = \text{diag}(w_1, \ldots, w_t),$$

where

$$w_i = \sigma_v^2 + n_i^{-1} \sigma_e^2.$$

Then, the generalized least squares estimator of $\sigma_v^2$ would be

$$\hat{\sigma}_v^2 = \frac{\sum_{i=1}^{t} \frac{1}{w_i} \sum_{i=1}^{t} z_i^2}{\sum_{i=1}^{t} \frac{1}{w_i}}.$$

which is an unbiased estimator of $\sigma_v^2$. With $\hat{w}_i$ replacing $w_i$ in $\Omega$, the estimated generalized least squares estimator of $\sigma_v^2$ would be $\hat{\sigma}_v^2$.

The true error vector $\theta$ is unobservable, so $\sigma_v^2$ cannot be computed exactly. In practice, one would use the estimator

$$\sigma_v^2 = \frac{\sum_{i=1}^{t} \frac{1}{w_i} \hat{u}_i^2}{\sum_{i=1}^{t} \frac{1}{w_i}}, \quad (10.3)$$

where
\[ \hat{u}_i = y_i - \hat{E}_i \hat{\beta}, \]

and \( \hat{\beta} \) is the estimated generalized least squares estimator of \( \beta \) given by (4.10).

Define \( G \) (generalized mean square among clusters) by

\[
G = n_G \left[ \sigma^2 + \left( \sum_{i=1}^{t} w_i^{-1} \right)^{-1} \sum_{i=1}^{t} n_i^{-1} \sigma^2_e \right]^{-1} \sum_{i=1}^{t} w_i^{-1} u_i^2.
\]

where

\[
n_G = \left( \sum_{i=1}^{t} w_i^{-1} n_i^{-1} \right)^{-1} \sum_{i=1}^{t} w_i^{-1}.
\]

Like MSA in Theorem 4.1, \( G \) will be approximated by a function of chi-square random variables. Assumptions (2) and (3) of Section IV.B.2 are retained, but assumption (4) is dropped. Recall that assumption (1) holds automatically for the nested-error model.

Recall from (4.5) and (4.8) that

\[
\hat{\sigma}_e^2 - \sigma_e^2 = o_p((n-t)^{-1/2})
\]

and
\[
\hat{\sigma}_v^2 - \sigma_v^2 = O_p(\max\{t^{-1/2}, (n-t)^{-1/2}\}).
\]

It follows that

\[
t^{-1} \sum_{i=1}^t w_i^{-1} = t^{-1} \sum_{i=1}^t w_i^{-1} - t^{-1} \sum_{i=1}^t w_i^{-2} \hat{w}_i (w_i - \hat{w}_i) + \text{remainder}
\]

\[
= t^{-1} \sum_{i=1}^t w_i^{-1} + (\sigma_v^2 - \sigma_v^2) t^{-1} \sum_{i=1}^t (\sigma_v^2 + \sigma_e^{-1} \sigma_e^{-2})
\]

\[
+ (\sigma_v^2 - \sigma_v^2) t^{-1} \sum_{i=1}^t n_i^{-1} (\sigma_v^2 + \sigma_e^{-1} \sigma_e^{-2})^2 + \text{remainder}
\]

\[
= t^{-1} \sum_{i=1}^t w_i^{-1} + O_p(\max\{t^{-1/2}, (n-t)^{-1/2}\}).
\]

Similarly,

\[
t^{-1} \sum_{i=1}^t w_i^{-1} (z_i - \sigma_v^2) = t^{-1} \sum_{i=1}^t w_i^{-1} (z_i - \sigma_v^2)
\]

\[
- t^{-1} \sum_{i=1}^t w_i^{-2} (z_i - \sigma_v^2) (\hat{w}_i - w_i) + \text{remainder}
\]

\[
= t^{-1} \sum_{i=1}^t w_i^{-1} (z_i - \sigma_v^2)
\]

\[
- (\sigma_v^2 - \sigma_v^2) t^{-1} \sum_{i=1}^t w_i^{-2} (z_i - \sigma_v^2)
\]

\[
- (\sigma_v^2 - \sigma_v^2) t^{-1} \sum_{i=1}^t n_i^{-1} (z_i - \sigma_v^2)
\]
+ remainder.

Since

$$E\{t^{-1} \sum_{i=1}^{t} w_i^{-2}(z_i - \sigma_v^2)\} = 0,$$

and

$$V\{t^{-1} \sum_{i=1}^{t} w_i^{-2}(z_i - \sigma_v^2)\} = 2t^{-2} \sum_{i=1}^{t} w_i^{-2}$$

$$+ 2(n-k-t-\lambda)^{-1} \sigma_e^4(t^{-1} \sum_{i=1}^{t} w_i^{-2} n_i^{-1})^2$$

$$= O(\max[t^{-1}, (n-t)^{-1}]),$$

it follows that

$$t^{-1} \sum_{i=1}^{t} w_i^{-2}(z_i - \sigma_v^2) = O_p(\max[t^{-1/2}, (n-t)^{-1/2}]).$$

Therefore,

$$t^{-1} \sum_{i=1}^{t} w_i^{-1}(z_i - \sigma_v^2) = t^{-1} \sum_{i=1}^{t} w_i^{-1}(z_i - \sigma_v^2)$$

$$+ O_p(\max[t^{-1}, (n-t)^{-1}]).$$
Then,

\[
\frac{\sigma^2_v - \sigma^2_{v'}}{\sigma^2_v} = \frac{t^{-1} \sum_{i=1}^{t} w_i^{-1}(z_i - \sigma_v^2)}{t^{-1} \sum_{i=1}^{t} w_i^{-1}}
\]

\[
= \frac{t^{-1} \sum_{i=1}^{t} w_i^{-1}(z_i - \sigma_v^2)}{t^{-1} \sum_{i=1}^{t} w_i^{-1}} + O_p(\max[t^{-1}, (n-t)^{-1}])
\]

\[
= \frac{t^{-1} \sum_{i=1}^{t} w_i^{-1}(z_i - \sigma_v^2)}{t^{-1} \sum_{i=1}^{t} w_i^{-1}} - \frac{t^{-1} \sum_{i=1}^{t} w_i^{-1}(z_i - \sigma_v^2)(t^{-1} \sum_{i=1}^{t} w_i^{-1} - t^{-1} \sum_{i=1}^{t} w_i^{-1})}{(t^{-1} \sum_{i=1}^{t} w_i^{-1})^2}
\]

+ remainder + O_p(\max[t^{-1}, (n-t)^{-1}])

\[
= \frac{t^{-1} \sum_{i=1}^{t} w_i^{-1}(z_i - \sigma_v^2)}{t^{-1} \sum_{i=1}^{t} w_i^{-1}} + O_p(\max[t^{-1}, (n-t)^{-1}])
\]

\[
= \sigma^2_v - \sigma^2_{v'} + O_p(\max[t^{-1}, (n-t)^{-1}]) .
\]

Therefore,
\[ \sigma_v^2 - \sigma_v^* = O_p \left( \max \left[ t^{-1}, (n - t)^{-1} \right] \right). \]

By the same arguments, it can be shown that

\[ \frac{t^{-1} \sum_{i=1}^{t} \hat{w}_i \hat{n}_i (\sigma^2_e - \sigma^2_v)}{t^{-1} \sum_{i=1}^{t} \hat{w}_i \hat{n}_i} = \frac{t^{-1} \sum_{i=1}^{t} \hat{w}_i \hat{n}_i (\sigma^2_e - \sigma^2_v)}{t^{-1} \sum_{i=1}^{t} \hat{w}_i \hat{n}_i} + O_p \left( \max \left[ t^{-1}, (n - t)^{-1} \right] \right), \]

so

\[ \frac{t \sum_{i=1}^{t} \hat{w}_i \hat{n}_i \sigma^2_e}{t \sum_{i=1}^{t} \hat{w}_i \hat{n}_i} - \frac{t \sum_{i=1}^{t} \hat{w}_i \hat{n}_i \sigma^2_v}{t \sum_{i=1}^{t} \hat{w}_i \hat{n}_i} = O_p \left( \max \left[ t^{-1}, (n - t)^{-1} \right] \right). \]

Thus,

\[ G = n_G \left[ \sigma_v^2 + \left( \sum_{i=1}^{t} \hat{w}_i \hat{n}_i \right) \right] \]

\[ = n_G \left[ \sigma_v^* + \left( \sum_{i=1}^{t} \hat{w}_i \hat{n}_i \right) \right] + O_p \left( n_G \max \left[ t^{-1}, (n - t)^{-1} \right] \right). \]
Let

\[ G = \sum_{i=1}^{t-\lambda} w_i^{-1} u_i^2. \]  
(10.6)

So,

\[ G - \tilde{G} = O_p(n_G \max[t^{-1}, (n-t)^{-1}]). \]

Note that both \( G \) and \( \tilde{G} \) are \( O_p(n_G) \). Also, since \( \tilde{G} \) is a function of \( \bar{u}_i \), \( t = 1, \ldots, t \), \( \tilde{G} \) and \( \hat{\sigma}_2^2 \) are independent.

Let

\[ \gamma_1 = \frac{1 - \hat{\phi}}{1 + (b_1 - 1)\hat{\phi}}, \]  
(10.7)
where

\[ \phi = \frac{t - \lambda - 2}{t - \lambda} \frac{\hat{\sigma}_e^2}{G} , \]  
(10.8)

and

\[ b_i = n_i^{-1} n_G . \]  
(10.9)

The estimator \( \hat{\phi} \) is approximately unbiased for

\[ \phi = \frac{\sigma^2}{n_G \sigma_v^2 + \sigma_e^2} . \]  
(10.10)

When \( n_i = r \) for all \( i \), then \( b_i = 1 \) and \( \gamma_1 \) is an unbiased estimator of

\[ \gamma_1 = \frac{1 - \phi}{1 + (b_i - 1)\phi} . \]  
(10.11)

The estimator \( \gamma_1 \) is the alternative estimator of \( \gamma_1 \) referred to at the beginning of this section.

The cluster mean, given by (3.4), is predicted by

\[ \mu_1^{(y)} = \bar{x}_i(p) \hat{\theta} + \gamma_1 (\bar{y}_i - \bar{x}_i \hat{\theta}) , \]  
(10.12)
where $\overline{y}_1.$ and $\overline{x}_1.$ are the cluster sample means of the variables, $\overline{x}_1(p)$ is the cluster population mean vector of the auxiliary variables, and $\hat{\beta}$ is the generalized least squares estimator of $\beta$ defined by (4.10).

Theorem 10.1 below is analogous to Theorem 4.1 with $\gamma_1$ estimated by $\hat{\gamma}_1$ instead of $\gamma_1$ of (4.14). In the proof of the theorem, $\phi$ will be approximated by $\tilde{\phi}$, a function of $\tilde{G}$.

**Theorem 10.1.** Assume model (3.1) - (3.2), and let assumptions (2) and (3) hold. Let $\hat{\gamma}(\gamma)$ be defined as in (10.12). Then,

$$
(\hat{\gamma}(\gamma) - \gamma_1)^2 = (\gamma_1 - \gamma_1)^2 + \frac{(t-1)b_1^2(\tilde{\phi} - \phi)^2}{t[1 + (b_1 - 1)\phi]^4} \overline{u}_1^2 \\
+ 2[(1 - \gamma_1)v_1 - \gamma_1 \overline{e}_1.] \frac{b_1(\tilde{\phi} - \phi)}{[1 + (b_1 - 1)\phi]^2} \overline{u}_1. \\
+ O_p(\max[t^{-1}, (n - t)^{-1}]),
$$

where

$$
\tilde{\phi} = \frac{\gamma - 1 - 2}{t - \lambda} \frac{\hat{\sigma}_e^2}{\hat{G}}, \tag{10.13}
$$
\( \tilde{G} \) is defined by (10.6), and \( b_1, \phi, \) and \( \tilde{\mu}_1^{(\gamma)} \) are defined by (10.9), (10.8), and (4.12), respectively. Furthermore,

\[
E \left\{ (\tilde{\mu}_1^{(\gamma)} - \mu_1)^2 + \frac{(t-1)b_1^2(\tilde{\phi} - \phi)^2}{t[1 + (b_1 - 1)\phi]^4} \tilde{u}_1^2 \right\} \\
+ 2[(1 - \gamma_1)\nu_1 - \gamma_1\tilde{e}_1, \gamma_1] \frac{b_1(\tilde{\phi} - \phi)}{[1 + (b_1 - 1)\phi]^2} \tilde{u}_1, \nu_1
\]

\[
= E\{ (\tilde{\mu}_1^{(\gamma)} - \mu_1)^2 \} + \frac{n_1^{-1} \sigma_v^2(1 - \gamma_1)(\sigma_v^2 + n_1^{-1} \sigma_e^2)(t-1)}{(\sigma_v^2 + n_1^{-1} \sigma_e^2)^2 t}
\]

\[
\times \left[ \frac{(\nu_1 + 2)(\nu_2 - 2)}{\nu_1^2 \nu_2^3} \left( \sum_{i=1}^{t-\lambda} \frac{\sigma_v^2 + n_G^{-1} \sigma_e^2}{\sigma_v^2 + n_1^{-1} \sigma_e^2} \right)^2 \right.
\]

\[
- \frac{2(\nu_2 - 2)}{\nu_2^2} \sum_{i=1}^{t} \frac{\sigma_v^2 + n_G^{-1} \sigma_e^2}{\sigma_v^2 + n_1^{-1} \sigma_e^2} + 1 \right],
\]

where \( \nu_1 = n - k - t + \lambda \), and \( \nu_2 = t - \lambda \).
Proof. Following the steps of the proof of Theorem 4.1,

\[
\hat{\mu}_1 - \mu_1 = \hat{\xi}_1(p) \hat{\xi} + \hat{\gamma}_1(\bar{\gamma}_1 - \bar{\xi}_1.) \hat{\xi} - \bar{\xi}_1(p) \hat{\xi} - \nu_1
\]

\[
= (\hat{\xi}_1(p) - \gamma_1 \bar{\xi}_1.)(\hat{\xi} - \xi) + (\bar{\xi}_1(p) - \gamma_1 \bar{\xi}_1.)(\hat{\xi} - \bar{\xi}) - (\gamma_1 - \gamma_1) \bar{\xi}_1. (\hat{\xi} - \bar{\xi})
\]

\[
- [(1 - \gamma_1) \nu_1 - \gamma_1 \tilde{e}_1.]
\]

\[
+ (\gamma_1 - \gamma_1) \bar{\mu}_1.
\]

\[
= (\hat{\mu}_1 - \mu_1) + (\bar{\xi}_1(p) - \gamma_1 \bar{\xi}_1.)(\hat{\xi} - \bar{\xi})
\]

\[
- (\gamma_1 - \gamma_1) \bar{\xi}_1. (\hat{\xi} - \xi) + (\gamma_1 - \gamma_1) \bar{\mu}_1.
\]

(10.14)

As before,

\[
\hat{\xi} - \xi = O_p(n^{-1/2})
\]

and
\[ \hat{g} - \tilde{g} = O_p(\max[t^{-1}, (n - t)^{-1}]). \]

The Taylor expansion of \( \hat{g} \) is

\[
\hat{g} = \gamma_1 \left\{ \frac{1 - \phi}{1 + (b_1 - 1)\phi} - \frac{b_1}{[1 + (b_1 - 1)\phi]^2} (\phi - \hat{\phi}) \right\} + \frac{2b_1(b_1 - 1)}{[1 + (b_1 - 1)\phi]^4} (\phi - \hat{\phi})^2 + \text{remainder.}
\]

Now

\[
\hat{\phi} - \tilde{\phi} = \frac{t - \lambda - 2}{t - \lambda} \sigma_e^2 (G^{-1} - \tilde{G}^{-1})
\]

\[= \frac{t - \lambda - 2}{t - \lambda} \sigma_e^2 \left( \frac{\tilde{G} - G}{G \tilde{G}} \right) \]

\[= O_p(n^{-1} \max[t^{-1}, (n - t)^{-1}]), \]

so
\[
\gamma_1 = \frac{1 - \phi}{1 + (b_1 - 1)\phi} - \frac{b_1}{[1 + (b_1 - 1)\phi]^2} (\tilde{\phi} - \phi) \\
+ \frac{2b_1(b_1 - 1)}{[1 + (b_1 - 1)\phi]^4} + \text{remainder} \\
+ o_p \left( n^{-1} \max\{t^{-1}, (n - t)^{-1}\} \right).
\]

Expanding \( \tilde{\phi} \) as a function of \( \hat{\sigma}_e^2 \) and \( \tilde{G} \) gives

\[
\tilde{\phi} = \frac{t - \lambda - 2}{t - \lambda} \phi + \frac{t - \lambda - 2}{t - \lambda} \frac{\hat{\sigma}_e^2 - \sigma_e^2}{n_G \sigma_v^2 + \sigma_e^2} \\
- \frac{t - \lambda - 2}{t - \lambda} \frac{\hat{\sigma}_e^2 (\tilde{G} - n_G \sigma_v^2 - \sigma_e^2)}{(n_G \sigma_v^2 + \sigma_e^2)^2} + \text{remainder}.
\]

Recall that

\[
\hat{\sigma}_e^2 - \sigma_e^2 = o_p \left( (n - t)^{-1/2} \right).
\]

Now
\[ v(\tilde{G} - n_G \sigma_v^2 - \sigma_e^2) = v \left\{ \frac{\Sigma_{i=1}^{t-\lambda} \left( \sigma_v^2 + n_i^{-1} \sigma_e^2 \right)^{-1} u_i^2}{n_G \Sigma_{i=1}^{t-\lambda} \left( \sigma_v^2 + n_i^{-1} \sigma_e^2 \right)^{-1}} \right\} \]

\[ = \frac{2n_G^2 (t - \lambda)}{t-\lambda} \left[ \Sigma_{i=1}^{t-\lambda} \left( \sigma_v^2 + n_i^{-1} \sigma_e^2 \right)^{-1} \right]^2 \]

\[ \leq \frac{2n_G^2 \sigma_v^4}{(t - \lambda)} . \]

Also,

\[ E(\tilde{G} - n_G \sigma_v^2 - \sigma_e^2) = 0 , \]

so,

\[ \tilde{G} - n_G \sigma_v^2 - \sigma_e^2 = O_p(n_G^{-1} t^{-1/2}) . \]

Then,

\[ \frac{(\tilde{G} - n_G \sigma_v^2 - \sigma_e^2)}{(n_G \sigma_v^2 + \sigma_e^2)^2} = O_p(n_G^{-1} t^{-1/2}) , \]
and

\[ \gamma_i - \gamma_t + o_p(\max[t^{-1/2}, (n-t)^{-1/2}]) \]

Returning to (10.14), it follows that

\[ \hat{\gamma}(\gamma) - \mu_i = (\hat{\gamma}(\gamma) - \mu_i) + (\gamma_i - \gamma_t) \vec{u}_i. \]

\[ + o_p(\max[t^{-1}, (n-t)^{-1}]) \]

\[ = (\hat{\gamma}(\gamma) - \mu_i) + \frac{b_i(\phi - \phi)}{[1 + (b_i - 1)\phi]^2} \vec{u}_i. \]

\[ + o_p(\max[t^{-1}, (n-t)^{-1}]) \]

Then,

\[ (\hat{\gamma}(\gamma) - \mu_i)^2 = (\hat{\gamma}(\gamma) - \mu_i)^2 + \frac{(t-1)b_i^2(\phi - \phi)^2}{t[1 + (b_i - 1)\phi]^4} \vec{u}_i^2 \]

\[ + 2[(1 - \gamma_t)\nu_i - \gamma_t\vec{e}_i] \frac{b_i(\phi - \phi)}{[1 + (b_i - 1)\phi]^2} \vec{u}_i. \]
\[ + O_p(\max[t^{-1}, (n - t)^{-1}]). \] (10.15)

Notice that multiplying the second term of (10.15) by \( t^{-1}(t-1) \) does not change the truth of the statement.

Now consider the expectation of the first three terms of (10.15). As in (4.30),

\[
E \left\{ \left[ (1 - \gamma_1)v_1 - \gamma_1 e_{1.} \right] \frac{b_1(\tilde{\phi} - \phi)}{[1 + (b_1 - 1)\phi]^2} \tilde{u}_{1.} \right\} = 0.
\]

Also,

\[
E \left\{ \frac{(t-1)b_1^2(\tilde{\phi} - \phi)^2}{t[1 + (b_1 - 1)\phi]^4} \tilde{u}_{1.}^2 \right\}
\]

\[
= \frac{(t-1)b_1^2 \phi^2(\sigma^2 + n_1^{-1} \sigma^2)}{t[1 + (b_1 - 1)\phi]^4} E \left\{ \frac{\tilde{\phi} - \phi}{\tilde{\phi} \phi^2} \tilde{u}_{1.}^2 \right\}.
\] (10.16)

Now,
The expectation in the second term of (10.17) is

$$E\left\{ \left( \frac{\tilde{v} - \phi}{\phi} \right)^2 \frac{\tilde{u}_{1,}^2}{V(\tilde{u}_{1,})} \right\} = E\{ \phi^{-2} \tilde{v}^2 \tilde{u}_{1,}^2 (V(\tilde{u}_{1,}))^{-1} \}$$

$$- 2E\{ \phi^{-1} \tilde{v}_1^2 (V(\tilde{u}_{1,}))^{-1} \} + 1.$$  

(10.17)

where $U_1$ and $U_2$ are independent chi-square random variables with 1 and $(t - \lambda - 1)$ degrees of freedom, respectively. By Lemma 4.1,

$$E\{ \phi^{-1} \tilde{v}_{1,}^2 (V(\tilde{u}_{1,}))^{-1} \} = \frac{t-\lambda-2}{t-\lambda} \sum_{i=1}^{t-\lambda-1} \frac{\sigma^2 + n_{i}^{-1} \sigma_{e}^2}{\sigma^2 + \sum_{j=1}^{n_{i}} \sigma_{e}^2} \cdot \left( \frac{U_1}{U_1 + U_2} \right)$$

$$= \frac{t-\lambda-2}{t-\lambda} \left( \frac{\sigma^2}{\sigma^2 + \sum_{i=1}^{n_{i}} \sigma_{e}^2} \right) E\left\{ \left( \frac{U_1}{U_1 + U_2} \right) \right\}.$$
Similarly,

\[
E\{\phi^{-2} \tilde{\phi}^2 \tilde{u}_{1,\ast}^2 (V(\tilde{u}_{1,\ast}))^{-1}\} = \left(\frac{t^{-\lambda} - 2}{t^{-\lambda}}\right)^2 E\left\{\sum_{i=1}^{\lambda} \frac{\tilde{e}^i}{\sigma^i_e}\right\} \left(\frac{t^{-\lambda}}{\Sigma} \frac{\sigma^2 + n^{-1}_G \sigma^2_e}{\sigma^2_v + n^{-1}_1 \sigma^2_e}\right)^2
\]

\[
	imes \left(\frac{U_1}{(U_1 + U_2)^2}\right)
\]

\[
= \frac{(t^{-\lambda} - 2)}{(t^{-\lambda})^2} \frac{(v_1 + 2)}{v_1} \left(\frac{t^{-\lambda}}{\Sigma} \frac{\sigma^2 + n^{-1}_G \sigma^2_e}{\sigma^2_v + n^{-1}_1 \sigma^2_e}\right)^2.
\]

Therefore, (10.17) becomes

\[
\frac{(v_2 - 2)}{v_2^2} \frac{(v_1 + 2)}{v_1} \frac{t^{-\lambda}}{\Sigma} \sum_{i=1}^{\lambda} \left(\frac{\sigma^2_v + n^{-1}_G \sigma^2_e}{\sigma^2_v + n^{-1}_1 \sigma^2_e}\right)^2
\]

\[
- 2 \frac{(v_2 - 2)}{v_2^2} \frac{t^{-\lambda}}{\Sigma} \sum_{i=1}^{\lambda} \left(\frac{\sigma^2_v + n^{-1}_G \sigma^2_e}{\sigma^2_v + n^{-1}_1 \sigma^2_e}\right)^2 + 1.
\]
The constant multiplier of (10.17) in (10.16) is

\[
\frac{(t-1)b_i^2 \phi^2(\sigma_v^2 + n_i^{-1} \sigma_e^2)}{t[1 + (b_i - 1)\phi]^4} = \frac{n_i^{-1} \sigma_e^2(1 - \gamma_i)(\sigma_v^2 + n_G^{-1} \sigma_e^2)^2(t-1)}{(\sigma_v^2 + n_i^{-1} \sigma_e^2)^2 t},
\]

following the steps in Theorem 4.1. Therefore, expression (10.16) is

\[
\frac{n_i^{-1} \sigma_v^2(1 - \gamma_i)(\sigma_v^2 + n_G^{-1} \sigma_e^2)^2(t-1)}{(\sigma_v^2 + n_i^{-1} \sigma_e^2)^2 t} \left[ \sum_{i=1}^{t} \frac{(v_1 + 2)(v_2 - 2)}{v_1 v_2} \right]^{t-\lambda} \left[ \sum_{i=1}^{t} \frac{\sigma_v^2 + n_G^{-1} \sigma_e^2}{\sigma_v^2 + n_i^{-1} \sigma_e^2} \right] + 1,
\]

and Theorem 10.1 is proved.

For the special case in which \( \xi_{ij} = 1 \) for all \( i \) and \( j \) and \( n_i = r \), \( i = 1,2,\ldots,t \), then \( n_i = n_G = r \). The penalty term (10.18) above simplifies to

\[
r_i^{-1} \sigma_e^2(1 - \gamma) \frac{(t-1)}{t} \left[ \frac{(rt - t + 2)(t - 3)}{(rt - t)(t - 1)^3} \frac{(t - 2)}{t} \right],
\]
\[
\left( -\frac{2(t - 3)}{(t - 1)^2} (t - 1) + 1 \right)
\]

\[
\frac{r^{-1} \sigma^2(1 - \gamma)}{t^2(r - 1)} \left[ (rt - t + 2)(t - 3) - 2(t - 3)t(r - 1) \right.
\]

\[
+ t(t - 1)(r - 1) \right]
\]

\[
\frac{2r^{-1} \sigma^2(1 - \gamma)(rt - 3)}{t^2(r - 1)} ,
\]

which matches Peixoto's result.