#50340

LOGISTIC REGRESSION ANALYSIS OF SAMPLE SURVEY DATA

STATISTICS STATISTIQUE
CANADA CANADA

SEP 16 1993

LIBRARY
BIBLIOTHÈQUE

G. Roberts, J.N.K. Rao

Department of Mathematics and Statistics, Carleton University

and S. Kumar

Household Survey Methods Division, Statistics Canada

Number: CHSM 85-081E

Logistic regression analysis of sample survey data

By G. ROBERTS, J.N.K. RAO

Department of Mathematics and Statistics, Carleton University, Ottawa, Canada

AND S. KUMAR

Household Survey Methods Division, Statistics Canada, Ottawa, Canada

SUMMARY

Standard chi-squared (X^2) or likelihood ratio (G^2) test statistics for logistic regression analysis, involving a binary response variable, are adjusted to take account of the survey design. These adjustments are based on certain generalized design effects (deffs). Logistic regression diagnostics to detect any outlying cell proportions in the table and influential points in the factor space are also developed, taking account of the survey design. Finally, the results are utilized to analyse some data from the October 1980 Canadian Labour Force Survey (CFS).

Some Key Words. Binary response data; chi-squared test statistic, Design effect; Satterthwaite's approximation; Diagnostics.

1. INTRODUCTION

The analysis of variation in the estimated proportions associated with a binary response variable is of considerable interest to researchers in social, behavioural and health sciences. Logistic regression models are extensively used for this purpose (see, for example, the books by Cox (1970), and McCullagh and Nelder (1983)). However, the standard statistical methods

for binomial proportions are often inappropriate for analysing sample survey data due to clustering and stratification used in the survey design. For instance, the standard chi-squared (X^2) and likelihood ratio (G^2) test statistics greatly inflate the type I error rate when a strong, positive clustering is present. As a result, some adjustments to the classical methods that take account of the survey design are necessary in order to make valid inferences from survey data. Section 2 provides adjustments, based on certain generalized design effects (deffs), to standard statistics for testing goodness-of-fit of the model and for testing subhypotheses given a model. A valid estimate of the asymptotic covariance matrix of fitted cell proportions is also obtained.

In addition to formal statistical tests, it is essential to develop diagnostic procedures to detect any outlying cell proportions and influential points in the factor space. Regression diagnostics for the standard linear model have been extensively developed in the literature (see the book by Cook and Weisberg (1982)). Pregibon (1981) developed similar methods for logistic regression with binomial proportions. In Section 3, some of these methods have been modified, by making necessary adjustments to account for the survey design. Finally, the results are utilized in Section 4 to analyse some data from the October 1980 Canadian Labour Force Survey (LFS).

Derivations of asymptotic variances and covariances and of adjustments to test statistics are sketched in the Appendix; details are given in G. Roberts' 1985 Ph.D. thesis at Carleton University.

The methods developed in this article require access to the estimated covariance matrix of cell response proportions. The calculation of standard errors for estimates of regression parameters, fitted cell proportions and

residuals (Section 2.1) requires knowledge of the entire estimated covariance matrix. On the other hand, simple bounds for some adjustments have been developed to facilitate secondary analysis from published tables (Rao and Scott, 1985). These bounds require knowledge only of estimated cell deffs or certain generalized deffs not depending on any hypothesis; reporting of these should be feasible.

Holt and Ewings (1985) have studied the effect of survey design on standard logistic regression analysis under a general cluster effects superpopulation model.

Although a logistic regression (logit) model for binary data can be viewed as an alternative specification of a suitable loglinear model, the objectives behind the two approaches are quite different; hence, the logit model should not be discarded merely as a special case (McCullogh, 1980). In particular, the loglinear models which correspond to logit models are eliminated at an early stage in the usual approaches to loglinear modelling, so that the final loglinear model usually does not correspond to any logit model (Kalbfleisch, 1984). Moreover, the standard errors or parameter estimates and the adjustments to X^2 in the loglinear set-up depend on the covariance matrix of the cells in the extended table appropriate for loglinear model analysis. This covariance matrix may not be available since the computer program is usually set up to provide only the estimated cell response proportions and their estimated covariance matrix.

2. TEST STATISTICS

Suppose that the population of interest is partitioned into I cells (domains) according to the levels of one or more factors. Let $\hat{N}_i$ denote

the survey estimate of the i-th domain size, $\hat{N}_i$ $(i=1,\ldots,I;\ \Sigma N_i = N)$. The corresponding estimate of the i-th domain total, N_{i1}, of a binary (0,1) response variable is denoted by $\hat{N}_{i1}$. The ratio estimate $p_i = \hat{N}_{i1}/\hat{N}_i$ is often used to estimate the population proportion $\pi_i = N_{i1}/N_i$. Standard sampling theory provides an estimate of the covariance matrix of the p_i's.

A logistic regression (logit) model for the proportions π_i is given by $\pi_i = f_i(\underset{\sim}{\beta})$, where

$$\nu_i = \ln\{f_i(\underset{\sim}{\beta})/(1-f_i(\underset{\sim}{\beta}))\} = \underset{\sim}{x}_i'\underset{\sim}{\beta}\ ,\ i=1,\ldots,I. \tag{2.1}$$

In (2.1), $\underset{\sim}{x}_i$ is an s-vector of known constants derived from the factor levels and $\underset{\sim}{\beta}$ is an s-vector of unknown parameters. Under independent binomial sampling in each domain, the maximum likelihood estimates (m.l.e.) $\hat{\underset{\sim}{\beta}}$ and $\hat{\underset{\sim}{f}} = \underset{\sim}{f}(\hat{\underset{\sim}{\beta}}) = (\hat{f}_1,\ldots,\hat{f}_I)'$ are obtained from the following likelihood equations through iterative calculations:

$$X'D(\underset{\sim}{n}/n)\hat{\underset{\sim}{f}} = X'D(\underset{\sim}{n}/n)\underset{\sim}{q}\ , \tag{2.2}$$

where $X' = (\underset{\sim}{x}_1,\ldots,\underset{\sim}{x}_I)$, $D(\underset{\sim}{n}/n) = \text{diag}(n_1/n,\ldots,n_I/n)$, $\underset{\sim}{q}$ is the vector of sample proportions $q_i = n_{i1}/n_i$, n_i is the sample size from the i-th domain $(\Sigma n_i = n)$, and n_{i1} is the i-th sample domain total. For general sample designs, we do not have m.l.e. due to difficulties in obtaining appropriate likelihood functions. Hence, it is a common practice to use a "pseudo m.l.e." of $\underset{\sim}{\beta}$ obtained from (2.2) by replacing n_i/n by the estimated domain relative size $w_i = \hat{N}_i/\hat{N}$, and q_i by the ratio estimate p_i:

$$X'D(\underset{\sim}{w})\hat{\underset{\sim}{f}} = X'D(\underset{\sim}{w})\underset{\sim}{p}\ , \tag{2.3}$$

where $D(\underset{\sim}{w}) = \text{diag}(w_1,\ldots,w_I)$ and $\underset{\sim}{p} = (p_1,\ldots,p_I)'$. The resulting estimates, $\hat{\underset{\sim}{\beta}}$ and $\hat{\underset{\sim}{f}} = f(\hat{\underset{\sim}{\beta}})$, are asymptotically consistent. Equation (2.3) may also be written as

$$X'\hat{\underset{\sim}{N}}_1(m) = X'\hat{\underset{\sim}{N}}_1 , \tag{2.4}$$

where $\hat{\underset{\sim}{N}}_1$ is the vector of estimated counts $\hat{N}_{i1}$ and $\hat{\underset{\sim}{N}}_1(m)$ is the vector of "pseudo m.l.e." $\hat{N}_{i1}(m) = \hat{N}_i\hat{f}_i$ of N_{i1}.

2.1. Estimated asymptotic variances and covariances

Let $n^{-1}\hat{V}$ denote the survey estimate of the covariance matrix of $\underset{\sim}{p}$. Then the estimated asymptotic covariance matrix of $\hat{\underset{\sim}{\beta}}$ is given by

$$\hat{V}_\beta = \frac{1}{n}(X'\hat{\Delta}X)^{-1}(X'D(\underset{\sim}{w})\hat{V}D(\underset{\sim}{w})X)(X'\hat{\Delta}X)^{-1}, \tag{2.5}$$

where $\hat{\Delta} = \text{diag}(w_1\hat{f}_1(1-\hat{f}_1),\ldots,w_I\hat{f}_I(1-\hat{f}_I))$ (see Appendix I). In the binomial case, (2.5) reduces to the standard formula $(X'\hat{\Delta}_b X)^{-1}$, where $\hat{\Delta}_b = \text{diag}(n_1^{-1}\hat{f}_1(1-\hat{f}_1),\ldots,n_I^{-1}\hat{f}_I(1-\hat{f}_I))$.

The estimated asymptotic covariance matrix of the fitted cell proportions $\hat{\underset{\sim}{f}}$ is given by (Appendix I)

$$\hat{V}_f = D(\underset{\sim}{w})^{-1}\hat{\Delta}X\hat{V}_\beta X'\hat{\Delta}D(\underset{\sim}{w})^{-1}. \tag{2.6}$$

The smoothed estimates $\hat{\underset{\sim}{f}}$ can be considerably more efficient than the survey estimates $\hat{\underset{\sim}{p}}$, especially for cells with a small sample, if the model (2.1) provides an adequate fit to $\hat{\underset{\sim}{p}}$ (see Section 3.3). It may be remarked that the estimates $\hat{f}_i$ are similar to the so-called synthetic estimates employed in small area estimation.

The estimated asymptotic covariance matrix of the residual vector

$\underset{\sim}{r} = \underset{\sim}{p} - \hat{\underset{\sim}{f}}$ is given by (Appendix I)

$$\hat{V}_r = n^{-1} A \hat{V} A' , \tag{2.7}$$

where

$$A = I - D(\underset{\sim}{w})^{-1} \hat{\Delta} X (X' \hat{\Delta} X)^{-1} X' D(\underset{\sim}{w}) \tag{2.8}$$

and I is the identity matrix. The diagonal elements $\hat{V}_{ii,r}$, of (2.7) are needed to calculate the standardized residuals $r_i / \hat{V}^{\frac{1}{2}}_{ii,r}$ which are useful in detecting outlying cell proportions (Section 2.4).

2.2. Goodness-of-fit of the model

The standard X^2 and G^2 tests of goodness-of-fit of the model (2.1) are given by

$$X^2 = n \sum_{i=1}^{I} (p_i - \hat{f}_i)^2 w_i / [\hat{f}_i (1-\hat{f}_i)] = \sum_{i=1}^{I} X_i^2 \text{ (say)} \tag{2.9}$$

and

$$G^2 = 2n \sum_{i=1}^{I} w_i [p_i \ell n(p_i/\hat{f}_i) + (1-p_i) \ell n\{(1-p_i)/(1-\hat{f}_i)\}]$$

$$= \sum_{i=1}^{I} G_i^2 \text{ (say)}. \tag{2.10}$$

Note that G_i^2 is defined at $p_i = 0$ and 1, respectively, by the quantities $-2nw_i \ell n(1-\hat{f}_i)$ and $-2nw_i \ell n\, \hat{f}_i$. Under independent binomial sampling, it is well-known that both X^2 and G^2 are asymptotically distributed as a χ^2 variable with $I-s$ degrees of freedom (d.f.) when the model (2.1) holds, but for general sample designs this result is no longer valid. In fact, X^2 (or G^2) is asymptotically distributed as a weighted sum $\Sigma \delta_i Z_i$ of independent χ^2 variables Z_i, each with 1 d.f. (see Appendix III). Here,

the weights δ_i $(i=1,\ldots,I-s)$ are estimated by $\hat{\delta}_i$, the eigenvalues of $\hat{V}_{0\phi}^{-1}\hat{V}_\phi$, where

$$\hat{V}_\phi = n^{-1}H'\hat{\Delta}^{-1}D(\underset{\sim}{w})\hat{V}D(\underset{\sim}{w})\hat{\Delta}^{-1}H\ , \tag{2.11}$$

$$\hat{V}_{0\phi} = n^{-1}H'\hat{\Delta}^{-1}H \tag{2.12}$$

and H is any $I \times (I-s)$ matrix of rank $I-s$ such that $H'X = 0$. The matrix $\hat{V}_{0\phi}^{-1}\hat{V}_\phi$ and $\hat{\delta}_i$ are termed a "generalized deff matrix" and a "generalized deff" respectively since they reduce to I and 1 respectively under binomial sampling.

An adjustment to X^2 or G^2 is obtained by treating $X_c^2 = X^2/\hat{\delta}_.$ or $G_c^2 = G^2/\hat{\delta}_.$ as a χ^2 variable with $I-s$ d.f., where $\hat{\delta}_. = \Sigma\hat{\delta}_i/(I-s)$ may be computed from the following expression:

$$(I-s)\hat{\delta}_. = n\sum_{i=1}^{I}\hat{V}_{ii,r}w_i/[\hat{f}_i(1-\hat{f}_i)]. \tag{2.13}$$

The adjusted statistics X_c^2 or G_c^2 should be satisfactory if the coefficient of variation (CV) of the δ_i's is small. A better adjustment, based on the well-known Satterthwaite approximation, treats $X_S^2 = X_c^2/(1+\hat{a}^2)$ or $G_S^2 = G_c^2/(1+\hat{a}^2)$ as a χ^2 variable with $(I-s)/(1+\hat{a}^2)$ d.f., where

$$\hat{a}^2 = \sum_{i=1}^{I-s}(\hat{\delta}_i - \hat{\delta}_.)^2/[(I-s)\hat{\delta}_.^2] \tag{2.14}$$

is the $(CV)^2$ of the $\hat{\delta}_i$'s, and $\Sigma\hat{\delta}_i^2$ is obtained from

$$\sum_{i=1}^{I-s}\hat{\delta}_i^2 = \sum_{i=1}^{I}\sum_{j=1}^{I}\hat{V}_{ij,r}^2(nw_i)(nw_j)/[\hat{f}_i\hat{f}_j(1-\hat{f}_i)(1-\hat{f}_j)], \tag{2.15}$$

where $\hat{V}_{ij,r}$ is the (i,j)-th element of $\hat{V}_r$. The test statistics X_S^2 and

G_S^2 take account of the variation in the δ_i's unlike X_c^2 and G_c^2.

A Wald statistic, which also takes the survey design into account, is given by

$$X_W^2 = \hat{\underset{\sim}{\nu}}' H \hat{V}_\phi^{-1} H' \hat{\underset{\sim}{\nu}} , \qquad (2.16)$$

where $\hat{\underset{\sim}{\nu}}$ is the vector of logits $\hat{\nu}_i = \ell n\{p_i/(1-p_i)\}$. The statistic X_W^2 is asymptotically distributed as a χ^2 variable with I-s d.f. when the model (2.1) holds. This result follows from the fact that testing the fit of the model (2.1) is equivalent to testing the hypothesis $H\underset{\sim}{\nu} = \underset{\sim}{0}$, where $\underset{\sim}{0}$ is a vector of zeros and $\underset{\sim}{\nu} = (\nu_1,\ldots,\nu_I)'$. The statistic X_W^2, however, is not defined if $p_i = 0$ or 1 for some i, as in the case of LFS data (Section 4). Moreover, it becomes unstable when any p_i is close to 1 (see Section 4) or when the number of degrees of freedom for $\hat{V}$ is not large relative to I-s (Fay, 1985).

2.3. Nested hypotheses

Suppose that the matrix X is partitioned as (X_1,X_2), where X_1 is $I \times r$ and X_2 is $I \times u$ $(r+u = s)$. The logit model (2.1), say M_1, may then be written as

$$\underset{\sim}{\nu} = X\underset{\sim}{\beta} = X_1\underset{\sim}{\beta}_1 + X_2\underset{\sim}{\beta}_2 , \qquad (2.17)$$

where $\underset{\sim}{\beta}_1$ is $r \times 1$ and $\underset{\sim}{\beta}_2$ is $u \times 1$. We are often interested in testing the null hypothesis $H_{2\cdot 1} : \underset{\sim}{\beta}_2 = \underset{\sim}{0}$, given M_1. Denote the reduced model under $H_{2.1}$ as M_2. The pseudo m.l.e. $\hat{\hat{\underset{\sim}{\beta}}}$ of $\underset{\sim}{\beta}$ under M_2 can be obtained from the equations

$$X_1' D(\underset{\sim}{w})\hat{\hat{\underset{\sim}{f}}} = X_1' D(\underset{\sim}{w})\underset{\sim}{p} \qquad (2.18)$$

again by iterative calculations, where $\hat{\hat{f}} = f(\hat{\hat{\beta}})$. The standard X^2 and G^2 tests of $H_{2.1}$ are given by

$$X^2(2|1) = n \sum_{i=1}^{I} (\hat{f}_i - \hat{\hat{f}}_i)^2 w_i / [\hat{\hat{f}}_i(1 - \hat{\hat{f}}_i)] \tag{2.19}$$

and

$$G^2(2|1) = 2n \sum_{i=1}^{I} w_i [\hat{f}_i \ln\{\hat{f}_i/\hat{\hat{f}}_i\} + (1-\hat{f}_i)\ln\{(1-\hat{f}_i)/(1-\hat{\hat{f}}_i)\}] \tag{2.20}$$

respectively. Under $H_{2.1}$, $X^2(2|1)$ or $G^2(2|1)$ is asymptotically distributed as a weighted sum, $\Sigma\delta_i(2|1)Z_i$, of independent χ^2 variables Z_i, each with 1 d.f. Here the weights $\delta_i(2|1)$ $(i=1,\ldots,u)$ are estimated by $\hat{\delta}_i(2|1)$, the eigenvalues of the generalized deff matrix

$$(\tilde{X}_2'\hat{\Delta}\tilde{X}_2)^{-1}(\tilde{X}_2'D(w)\hat{V}D(w)\tilde{X}_2) , \tag{2.21}$$

where $\tilde{X}_2 = [I - X_1(X_1'\hat{\Delta}X_1)^{-1}X_1'\hat{\Delta}]X_2$ (see Appendix II). In the case of binomial sampling, $\delta_i(2|1) = 1$ for all i so that we get the well-known result that $X^2(2|1)$ or $G^2(2|1)$ is asymptotically distributed as a χ^2 variable with u d.f. under $H_{2.1}$.

An adjustment to $G^2(2|1)$ or $X^2(2|1)$ is obtained by treating $G^2(2|1)/\hat{\delta}.(2|1)$ or $X^2(2|1)/\hat{\delta}.(2|1)$ as χ^2 with u d.f. under $H_{2.1}$, where $\hat{\delta}.(2|1) = u^{-1}\Sigma\hat{\delta}_i(2|1)$ may be computed from

$$u\hat{\delta}.(2|1) = n \sum_{i=1}^{I} \tilde{V}_{ii,r} w_i / [\hat{\hat{f}}_i(1-\hat{\hat{f}}_i)] \tag{2.22}$$

and $\tilde{V}_{ii,r}$ is the i-th diagonal element of the estimated covariance matrix of residuals, $r_i(2|1) = \hat{f}_i - \hat{\hat{f}}_i$, given by

$$\tilde{V}_r = n^{-1}D(w)^{-1}\hat{\Delta}\tilde{X}_2\hat{A}\tilde{X}_2'\hat{\Delta}D(w)^{-1} \tag{2.23}$$

(2.23)

(see equations (A.11) and (A.13) in Appendix II) and

$$\hat{A} = (\tilde{X}_2'\hat{\Delta}\tilde{X}_2)^{-1}(\tilde{X}_2'D(\underset{\sim}{w})\hat{V}D(\underset{\sim}{w})\tilde{X}_2)(\tilde{X}_2'\hat{\Delta}\tilde{X}_2)^{-1}. \tag{2.24}$$

The standardized residuals $r_i(2|1)/\tilde{V}^{\frac{1}{2}}_{ii,r}$ can also be computed. As in the case of goodness-of-fit, a better adjustment based on the Satterthwaite approximation can be obtained, utilizing the elements of $\tilde{V}_r$.

A Wald statistic for testing $H_{2.1}$ is given by

$$X_W^2(2|1) = \hat{\underset{\sim}{\beta}}_2'\hat{V}_{2\beta}^{-1}\hat{\underset{\sim}{\beta}}_2 , \tag{2.25}$$

where $\hat{V}_{2\beta}$ is the principal submatrix of (2.5) corresponding to $\underset{\sim}{\beta}_2$. Under $H_{2.1}$, the statistic $X_W^2(2|1)$ is asymptotically distributed as a χ^2 with u d.f. In particular, if β_2 is a scalar, then we can treat $\hat{\beta}_2/\text{s.e.}(\hat{\beta}_2)$ as $N(0,1)$ or $\hat{\beta}_2^2/\text{var}(\hat{\beta}_2)$ as χ^2 with 1 d.f., under $H_{2.1}$. Note that $X_W^2(2|1)$ is well-defined even if $p_i = 0$ or 1 for some i, unlike X_W^2. The Wald statistic (2.25) is computationally simpler than the adjusted X^2 or G^2 statistics.

2.4. Diagnostics

It would be desirable to make a critical assessment of the logit fit by identifying any outlying cell proportions and influential points in the factor space. For this purpose, the vector of residuals, $\underset{\sim}{r}$, and a projection matrix in the factor space provide useful tools. However, the residuals can be defined on different scales, unlike in the case of the standard linear model. A natural choice that takes account of the survey design is the vector of standardized residuals $e_i = r_i/\hat{V}^{\frac{1}{2}}_{ii,r}$.

Since the e_i are approximately $N(0,1)$ under the model, the expected numbers of $|e_i|$ exceeding 1.96, 2.33 and 2.58 are roughly equal to 0.05I, 0.02I and 0.01I respectively, where I is the number of residuals (cells). These expected numbers provide a rough guide for identifying any outlying cells. Ignoring the design and hence using standardized residuals under binomial sampling could lead to erroneous diagnostics.

The standardized residuals e_i, however, become unreliable for these cells with $p_i = 1$ or close to 1. To circumvent this difficulty, we suggest the use of components of X_c^2 or G_c^2, viz. $\tilde{X}_i = X_i/\hat{\delta}^{\frac{1}{2}}$ or $\tilde{G}_i = G_i/\hat{\delta}^{\frac{1}{2}}$, $i = 1,\ldots,I$, for residual analysis; Pregibon (1981) used X_i or G_i in the binomial case. In either case, large individual components should roughly indicate cells poorly accounted for by the model. Index plots $\tilde{X}_i$ vs. i and $\tilde{G}_i$ vs. i are useful for displaying these components. A normal probability plot of $\tilde{X}_i$ or $\tilde{G}_i$ (i.e., the ordered values plotted against standard normal quantiles) is also useful for detecting deviations from the model, i.e., deviations from a straight line configuration.

Following Pregibon (1981), we suggest the use of diagonal elements, m_{ii}, of the projection matrix

$$M = I - \hat{\Delta}^{\frac{1}{2}}X(X'\hat{\Delta}X)^{-1}X'\hat{\Delta}^{\frac{1}{2}} = I - T, \quad \text{say} \tag{2.26}$$

to detect influential points. The matrix M arises naturally in solving the likelihood equations (2.4) by the method of iteratively reweighted least squares (Pregibon, 1981), and small values of m_{ii} call attention to extreme points in the factor space. Again an index plot m_{ii} vs. i

provides a useful display. It may be noted that the design effect does not come into the picture with m_{ii} since we are using pseudo m.l.e. based on binomial sampling.

Another useful plot which effectively summarises the information in the index plots X_i vs. i and m_{ii} vs. i is given by the scatter plot of $\tilde{X}_i^2/X_c^2 = X_i^2/X^2$ vs. t_{ii}, where t_{ii} is the i-th diagonal element of T given by (2.26). Again, the deff does not come into the picture.

The diagnostic measures e_i, X_i (or G_i) and m_{ii} are useful for detecting extreme points, but not for assessing their impact on various aspects of the fit, including parameter estimates, $\hat{\underset{\sim}{\beta}}$, fitted values, $\hat{\underset{\sim}{f}}$, and goodness-of-fit measures $X^2/\hat{\delta}_.$ and $G^2/\hat{\delta}_.$ or others. Following Pregibon (1981), we suggest three measures which quantify the effect of extreme cells (points) on the fit. These measures take account of the design effect.

(1) Coefficient sensitivity. Let $\hat{\beta}_j(-\ell)$ denote the pseudo m.l.e. of β_j obtained after deleting the ℓ-th cell from the data. Then the quantity $\Delta_j(\ell) = \{\hat{\beta}_j - \hat{\beta}_j(-\ell)\}/\text{s.e.}(\hat{\beta}_j)$ provides a measure of the j-th coefficient sensitivity to the ℓ-th cell (point). The index plots $\Delta_j(\ell)$ vs. ℓ for each j provide useful displays, but the task of "looking" at the index plots could become unmanageable unless the number of coefficients in the model is small.

(2) Sensitivity of fitted values. Significant changes in coefficient estimates when the ℓ-th point is deleted from the data set does not

necessarily imply that the fitted values $\underset{\sim}{f}$ also vary significantly from $\underset{\sim}{f}(-\ell) = \underset{\sim}{f}(\hat{\beta}(-\ell))$, where $\hat{\underset{\sim}{\beta}}(-\ell)$ is the estimate of $\underset{\sim}{\beta}$ obtained by deleting the ℓ-th cell; i.e. $\|\hat{\underset{\sim}{f}}-\hat{\underset{\sim}{f}}(-\ell)\|$ could be small. We therefore use $\{G^2-\tilde{G}^2(-\ell)\}/\hat{\delta}.$ or $\{X^2-\tilde{X}^2(-\ell)\}/\hat{\delta}.$ to assess the impact of the ℓ-th point on the fitted values $\hat{\underset{\sim}{f}}$, where $\tilde{G}^2(-\ell)$ and $\tilde{X}^2(-\ell)$ are given by (2.10) and (2.9) respectively when $\hat{\underset{\sim}{f}} = \underset{\sim}{f}(\hat{\underset{\sim}{\beta}})$ is replaced by $\hat{\underset{\sim}{f}}(-\ell)$.

(3) <u>Goodness-of-fit sensitivity</u>. A measure of goodness-of-fit sensitivity is given by $\{G^2-G^2(-\ell)\}/\hat{\delta}.$ or $\{X^2-X^2(-\ell)\}/\hat{\delta}.$, where $X^2(-\ell) = n\sum_{i\neq\ell}(p_i-\hat{f}_i(-\ell))^2 w_i/\{\hat{f}_i(-\ell)(1-\hat{f}_i(-\ell))\}$ and $G^2(-\ell)$ similarly defined using (2.10). Note that $X^2(-\ell) \neq \tilde{X}^2(-\ell)$ and $G^2(-\ell) \neq \tilde{G}^2(-\ell)$.

3. APPLICATION TO LFS DATA

We have applied the methods in Section 2 to some data from the October 1980 Canadian Labour Force Survey (LFS). The sample consisted of males aged 15-64 who were in the labour force and not full-time students. We have chosen two factors, age and education, to explain the variation in nonemployment rates via logit models. Age-group levels were formed by dividing the interval [15,64] into ten groups with the j-th age group being the interval $[10+5j,\ 14+5j]$, $j=1,\ldots,10$ and then using the midpoint of each interval, $A_j = 12+5j$ as the value of age for all persons in that age group. Similarly, the levels of education, E_k, were formed by assigning to each person a value based on the median years of schooling resulting in the following six levels: 7, 10, 12, 13, 14 and 16. The resultant age by education cross-classificatin provided a two-way table of $I = 60$ cell proportions (employment rates), π_{jk}.

The LFS design employed stratified multi-stage cluster sampling with two stages in the self-representing (SR) urban areas and three or four stages in the non-self-representing (NSR) areas in each province. The survey estimates, p_{jk}, of π_{jk} were adjusted for post-stratification using the projected census age-sex distribution at the provincial level. The estimated covariance matrix, $\hat{V}/n$, of the estimates p_{jk} was based on more than 450 first-stage units so that the degrees of freedom for V was large compared to $I = 60$. A detailed description of the sampling plan and associated estimation procedures for the LFS is given in Statistics Canada (1977).

3.1. Formal tests of hypotheses

Scatter plots of the logits $\hat{\nu}_{jk} = \ln\{p_{jk}/(1-p_{jk})\}$ against age levels A_j, at each education level E_k, indicate that $\hat{\nu}_{jk}$ increases with age to a maximum and then decreases. Hence, the following model might be suitable to explain the variation in the π_{jk}:

$$\nu_{jk} = \ln\{\pi_{jk}/(1-\pi_{jk})\} = \beta_0 + \beta_1 A_j + \beta_2 A_j^2 + \beta_3 E_k + \beta_4 E_k^2 ,$$

$$j = 1,\ldots,10;\ k = 1,\ldots,6 . \qquad (3.1)$$

Some previous work in the sociological literature also supports such a model (Block and Smith, 1977). Applying the results of Section 2, we obtain the following values for testing the goodness-of-fit of the model (3.1):

$$X^2 = 98.9, \quad G^2 = 101.2$$

$$X^2/\hat{\delta}. = 52.5, \quad G^2/\hat{\delta}. = 53.7 \text{ and } \hat{\delta}. = 1.88.$$

Since the value of X^2 or G^2 is larger than $\chi^2_{0.05}(55) = 77.3$, the upper 5% point of χ^2 with $I-s = 55$ d.f., we would reject the model (3.1) if the sample design is ignored. On the other hand, the value of $X^2/\hat{\delta}.$ or $G^2/\hat{\delta}.$ indicates that the model is adequate, the significance level (or P-value) being approximately equal to 0.52. The value of Satterthwaite's statistic X^2_S when adjusted to refer to $\chi^2_{0.05}(55)$ is equal to 47.7 which is also not significant at the 5% level. Moreover, in the present context with $s(=5)$ relatively small compared to $I(=60)$, the simple correction $\hat{d}. = \sum\sum \hat{d}_{jk}/60$, the average cell deff, is very close to $\hat{\delta}.$: $\hat{d}. = 1.905$ compared to $\hat{\delta}. = 1.88$, where $\hat{d}_{jk} = \text{var}(p_{jk})/[(nw_{jk})^{-1}p_{jk}(1-p_{jk})]$ is the estimated cell deff and w_{jk} is the estimated relative size for the (j,k)-th cell. Rao and Scott (1985) have shown that $\hat{\delta}. \doteq \hat{d}.$ when $I/(I-s) \doteq 1$.

The Wald statistic X^2_W is not defined here since two of the cells have $p_{jk} = 1$, i.e. all employed. We made minor perturbations to the estimated counts to ensure that $p_{jk} < 1$ for all cells and then computed X^2_W. The resulting values of X^2_W are all large compared to $X^2/\hat{\delta}.$, at least 30 times larger than $X^2/\hat{\delta}.$ and vary considerably (1715 to 3061). We thus concluded that the Wald statistic is very unstable for testing goodness-of-fit in the present context. If the two cells having $p_{jk} = 1$ are deleted, then $X^2_W = 68.4 < \chi^2_{0.05}(53) = 71.0$, indicating that the model (3.1) is adequate. However, it is not a good practice to delete cells just to accommodate a chosen statistic. The other problem with X^2_W, noted by Fay (1985), does not arise here since the d.f. for $\hat{V}$ is large as compared to the number of cells in the table.

The pseudo m.l.e. of the β_i, their standard errors and the corresponding standard errors under binomial sampling, all obtained under the model (3.1), are given in Table 1. The Wald statistic $X_W^2(2|1)$ and the G^2 statistic $G^2(2|1)/\hat{\delta}.(2|1)$ for the hypotheses $H_{2.1}: \beta_2 = 0$ and $H_{2.1}: \beta_4 = 0$ conditional on model (3.1), are also given in Table 1. As expected, the true standard errors are larger than the corresponding binomial standard errors. The hypothesis $\beta_4 = 0$ (i.e., no quadratic education effect) is not rejected at the 5% level either by the Wald statistic or the G^2-statistic ($\chi^2_{0.05}(1) = 3.84$). On the other hand, the coefficient β_2 of A_j^2 is highly significant, indicating a quadratic age effect.

We have also tested two more nested hypotheses given the model (3.1): $H_{2.1}: \beta_3 = \beta_4 = 0$ (i.e., no education effect); $H_{2.1}: \beta_2 = \beta_4 = 0$ (i.e., no quadratic effects). Both hypotheses are rejected at the 1% level:

$$G^2(2|1)/\hat{\delta}.(2|1) = 282.2/1.64 = 172.1,\ X_W^2(2|1) = 165.6 \text{ for } H_{2.1}: \beta_3 = \beta_4 = 0$$

$$G^2(2|1)/\hat{\delta}.(2|1) = 242.2/2.28 = 106.3,\ X_W^2(2|1) = 162.1 \text{ for } H_{2.1}: \beta_2 = \beta_4 = 0,$$

as compared to $\chi^2_{0.01}(2) = 9.21$. Note that the Wald statistic is stable for testing nested hypotheses, unlike in the case of goodness-of-fit, and leads to values close to the corresponding values of $G^2(2|1)/\hat{\delta}.(2|1)$.

By the above tests of goodness-of-fit and nested hypotheses, we arrived at the following simple model involving only four parameters:

$$\ln\left\{\frac{\hat{p}_{dk}}{1-\hat{p}_{dk}}\right\} = \underset{(0.247)}{-3.10} + \underset{(0.013)}{0.211A_j} - \underset{(0.000172)}{0.00218A_j^2} + \underset{(0.0115)}{0.1509E_k} \qquad (3.2)$$

The standard errors of parameter estimates are given in brackets in (3.2). The diagnostics in Section 3.2 will be based on the fitted model (3.2).

3.2. Diagnostics

We now apply to the LFS data the diagnostics developed in Section 2.4.

(i) Residual analysis

The sixty cells in the two-way table were numbered lexicographically and the standardized residuals e_i were computed under the model (3.2). The cells numbered 6 and 54 with $p_i = 1$ lead to very large e_i-values: (66.2 and 6.2 respectively, which are unreliable as noted earlier). Among the remaining e_i, the residuals numbered 7, 27 and 59 have values 3.84, 2.73 and 2.52 respectively, whereas the expected number of $|e_i|$ exceeding 2.33 is roughly $60 \times 0.02 = 1.2$. Hence, there is some indication that cells 7 and 27 might correspond to outlying cell proportions.

The normal probability plot of $\tilde{G}_i = G_i/\hat{\delta}^{\frac{1}{2}}$ displayed in Figure 1 indicates no significant deviations from a straight line configuration. The index plot of G_i, Figure 2, is consistent with Figure 1. The plots of $\tilde{X}_i$ are not given to save space but they are similar to those of $\tilde{G}_i$. We thus conclude that there is no evidence of outlying cell proportions when the components $\tilde{G}_i$ or $\tilde{X}_i$ are used for residual analysis.

(ii) Influential cells

The index plot of m_{ii} displayed in Figure 3 clearly points to cells 2, 3 and 55. Figure 4 gives the plot of $\tilde{X}_i^2/X_c^2 = X_i^2/X^2$ vs. t_{ii}, where the line with slope -1 is given by $X_i^2/X^2 + t_{ii} = 3\text{ave}(t_{ii}^*)$. Here $t_{ii}^* = t_{ii} + X_i^2/X^2$, and the values of t_{ii}^* near to unity correspond to cells which are outlying or influential or both (Pregibon, 1981) and appear above the line in Figure 3. It is clear that cells 2, 3 and 55

warrant further examination.

(iii) Coefficient sensitivity

The index plots for measuring coefficiency sensitivity ($\Delta_j(\ell)$ vs. ℓ) are displayed in Figures 5, 6, 7 and 8 for β_0, β_1, β_2 and β_3 respectively. It is clear from these plots that cells 2 and 3 cause instability in $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$, whereas $\hat{\beta}_3$ is affected by cell 7.

(iv) Sensitivity of fitted values

Figure 9 displays the plot of $[G^2 - G^2(-\ell)]/\hat{\delta}. = c_\ell$ vs. ℓ for assessing the impact of individual cells on fitted values. Significant peaks in this figure correspond to cells 2 and 3 and to a lesser extent to cell 7. Following Cook (1977) and Pregibon (1981), it may be noted that the comparison of c_ℓ to the percentage point of χ^2 with s d.f. ($s = 4$ in the model (3.1)) gives a rough guide as to which contour of the confidence region the pseudo m.l.e. is displaced due to deletion of the ℓ-th cell. The value $c_\ell = 2.1$ for cell 2 roughly corresponds to the 78% contour of the confidence region.

(v) Goodness-of-fit sensitivity

Figure 10 displays the plot of $[G^2-G^2(-\ell)]/\hat{\delta}.$ vs. ℓ ; the plot of $[X^2-X^2(-\ell)]/\hat{\delta}.$ is similar here but the former plot is preferred (Pregibon, 1981). Significant peaks in this figure correspond to cells 2, 3, 7, 27 39 and 54 (values ≥ 3), the most significant being cell 7 with the value 5.4. By deleting cell 7 and recomputing the adjusted statistic $G_c^2(-7) = G^2(-7)/\hat{\delta}.(-7)$ where $\hat{\delta}.(-7)$ is the corresponding estimate of $\delta.$, we get $G_c^2(-7) = 48.43$ with 55 d.f. compared to $G^2/\hat{\delta}. = 55.3$ with 56 d.f.

Our investigation indicates on the whole that cells 7, 2 and 3 are possible candidates for deletion, but we feel that their impact is not significant enough to warrant this action.

3.3. Smoothed estimates

The coefficient of variation of survey estimates, $1-p_{jk}$, of unemployment rates is quite large for cells with small samples, ranging from 6.8% (for cell 3) to 98.5% (for cell 59). Because of this, we computed the coefficient of variation of smoothed estimates, $1-\hat{f}_{jk}$, under the model (3.1), using formula (2.6). The smoothed estimates lead to a dramatic reduction in coefficient of variation: the coefficient of variation of $1-\hat{f}_{jk}$ ranges from 3.3% (cell 8) to 12.4% (cell 60); the coefficient of variation for cell 59 is reduced from 98.5% to 11.0%. The average coefficient of variation of $1-p_{jk}$ (over the 58 cells with $1-p_{jk} > 0$) is 32.1% compared to 6.2%, the average coefficient of variation of $1-\hat{f}_{jk}$ (over all the 60 cells). Moreover, the bias of smoothed estimates should be relatively small since model (3.1) provides an adequate fit to the data.

APPENDIX

Outline of derivation of main results

I. Asymptotic variances and covariances

The pseudo m.l.e. are obtained from the binomial likelihood, $L(\underset{\sim}{\beta})$, say, by replacing n_i by nw_i and n_{i1} by $(nw_i)p_i$ and then minimizing with respect to $\underset{\sim}{\beta}$. It is easily seen that

$$-2 \ln L(\underset{\sim}{\beta}) = 2nG^2(\underset{\sim}{a}^*, \underset{\sim}{b}^*(\underset{\sim}{\beta})) + \text{terms not involving } \underset{\sim}{\beta} ,$$

where $\underset{\sim}{a}^* = \{w_1p_1,\ldots,w_Ip_I; w_1(1-p_1),\ldots,w_I(1-p_I)\}$ and $\underset{\sim}{b}^* = b^*(\beta) = \{w_1f_1,\ldots,w_If_I; w_1(1-f_1),\ldots,w_I(1-f_I))\}$ and $G^2(\underset{\sim}{a}^*,\underset{\sim}{b}^*) = \Sigma a_i^* \ln(a_i^*/b_i^*)$, $\Sigma a_i^* = \Sigma b_i^* = 1$. Hence, noting that maximizing $L(\beta)$ is equivalent to minimizing $G^2(\underset{\sim}{a}^*,\underset{\sim}{b}^*(\beta))$, we can use the results of Birch (1964) to get

$$\sqrt{n}(\hat{\underset{\sim}{\beta}}-\underset{\sim}{\beta}) \sim \sqrt{n}[(B'B)^{-1}B'D(\underset{\sim}{b})^{-\frac{1}{2}}(\underset{\sim}{a}-\underset{\sim}{b}(\beta))] \tag{A.1}$$

where ~ denotes "asymptotic equivalence". Here $\underset{\sim}{a}$ and $\underset{\sim}{b}$ are derived from $\underset{\sim}{a}^*$ and $\underset{\sim}{b}^*$ respectively by replacing w_i with $W_i = N_{i1}/N_i$, $w_i - W_i = o_p(1)$, $D(b) = \text{diag}(b_1,\ldots,b_I)$, and $B = D(\underset{\sim}{b})^{-\frac{1}{2}}(\partial\underset{\sim}{b}/\partial\beta)$. In the case of logit model (2.1), Birch's (1964) regularity conditions are satisfied and (A.1) reduces to

$$\sqrt{n}(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta}) \sim (X'\Delta X)^{-1}X'D(\underset{\sim}{W})\{\sqrt{n}(\underset{\sim}{p} - \underset{\sim}{f})\} , \tag{A.2}$$

where $\Delta = \text{diag}\{W_1f_1(1-f_1),\ldots,W_If_I(1-f_I)\}$ and $D(\underset{\sim}{W}) = \text{diag}(W_1,\ldots,W_I)$. Now assuming that $\sqrt{n}(\underset{\sim}{p} - \underset{\sim}{f})$ converges in distribution to $N_I(\underset{\sim}{0},V)$, we get, from (A.2), the asymptotic covariance matrix of $\hat{\underset{\sim}{\beta}}$:

$$V_{\underset{\sim}{\beta}} = \frac{1}{n}(X'\Delta X)^{-1}(X'D(\underset{\sim}{W})VD(\underset{\sim}{W})X)(X'\Delta X)^{-1}. \tag{A.3}$$

Replacing the parameters in (A.3) by their estimates, we get (2.5). Similarly, noting that

$$\sqrt{n}(\hat{\underset{\sim}{f}} - \underset{\sim}{f}) \sim (\frac{\partial \underset{\sim}{f}}{\partial \underset{\sim}{\beta}}) \{\sqrt{n}(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})\} \tag{A.4}$$

$$= D(\underset{\sim}{W})^{-1}\Delta X\{\sqrt{n}(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})\}$$

and

$$\sqrt{n}(\underset{\sim}{p} - \underset{\sim}{\hat{f}}) = \sqrt{n}\,\underset{\sim}{r} \sim \{I - D(W)^{-1}\Delta X(X'\Delta X)^{-1}X'D(W)\}\{\sqrt{n}(\underset{\sim}{p}-\underset{\sim}{f})\} \tag{A.5}$$

we get (2.6) and (2.7).

II. <u>Asymptotic null distribution of $X^2(2|1)$</u>

The statistic $X^2(2|1)$ (given by (2.19)) for testing the nested hypothesis $H_{2.1} : \underset{\sim}{\beta}_2 = \underset{\sim}{0}$ is asymptotically equivalent to

$$n(\underset{\sim}{\hat{f}} - \underset{\sim}{\hat{\hat{f}}})'D(\underset{\sim}{W})\Delta^{-1}D(W)(\underset{\sim}{\hat{f}} - \underset{\sim}{\hat{\hat{f}}}) \tag{A.6}$$

under $H_{2.1}$. Now, similar to (A.4) we have

$$\sqrt{n}(\underset{\sim}{\hat{\hat{f}}} - \underset{\sim}{f}) \sim D(\underset{\sim}{W})^{-1}\Delta X_1\{\sqrt{n}(\underset{\sim}{\hat{\hat{\beta}}}_1 - \underset{\sim}{\beta}_1)\}, \tag{A.7}$$

where

$$\sqrt{n}(\underset{\sim}{\hat{\hat{\beta}}}_1 - \underset{\sim}{\beta}_1) \sim (X_1'\Delta X_1)^{-1}X_1'D(\underset{\sim}{W})\{\sqrt{n}(\underset{\sim}{p} - \underset{\sim}{f})\}. \tag{A.8}$$

Hence, from (A.4) and (A.6),

$$\sqrt{n}(\underset{\sim}{\hat{f}}-\underset{\sim}{\hat{\hat{f}}}) \sim D(\underset{\sim}{W})^{-1}\Delta\{X_1\sqrt{n}(\underset{\sim}{\hat{\beta}}_1-\underset{\sim}{\beta}_1)+X_2\sqrt{n}\,\hat{\beta}_2 - X_1\sqrt{n}(\underset{\sim}{\hat{\hat{\beta}}}_1-\underset{\sim}{\beta}_1)\} \tag{A.9}$$

under $H_{2.1}$. Now, following Rao and Scott (1984), we express $X'\Delta X$ as a partitioned matrix

$$X'\Delta X = \begin{pmatrix} X_1'\Delta X_1 & X_1'\Delta X_2 \\ X_2'\Delta X_1 & X_2'\Delta X_2 \end{pmatrix}$$

and then use the standard formula for the inverse of a partioned matrix to get, after simplification,

$$\sqrt{n}(\underset{\sim}{\hat{\hat{\beta}}}_1 - \underset{\sim}{\beta}_1) \sim \sqrt{n}(\underset{\sim}{\hat{\beta}}_1 - \underset{\sim}{\beta}_1) + (X_1'\Delta X_1)^{-1}(X_1'\Delta X_2)\sqrt{n}\,\underset{\sim}{\beta}_2 . \tag{A.10}$$

Substituting (A.10) into (A.9) we get

$$\sqrt{n}(\hat{f} - f) \approx D(W)^{-1}\Delta\tilde{X}_2\sqrt{n}\,\hat{\beta}_2 , \qquad \text{(A.11)}$$

where

$$\tilde{X}_2 = X_2 - X_1(X_1'\Delta X_1)^{-1}(X_1'\Delta X_2) .$$

As a result, we get the following asymptotic representation from (A.5) and (A.11):

$$X^2(2|1) \approx n\,\hat{\beta}_2'(\tilde{X}_2'\Delta\tilde{X}_2)\hat{\beta}_2 . \qquad \text{(A.12)}$$

Also it follows from (A.3) and the formula for the inverse of a partitioned matrix that the asymptotic covariance matrix of $\hat{\beta}_2$ may be written as

$$V_{\hat{\beta}_2} = \frac{1}{n}(\tilde{X}_2'\Delta\tilde{X}_2)^{-1}(\tilde{X}_2'D(W)VD(W)\tilde{X}_2)(\tilde{X}_2'\Delta\tilde{X}_2)^{-1} \qquad \text{(A.13)}$$

so that $\hat{\beta}_2$ is approximately $N_u(0,V_{\hat{\beta}_2})$ under $H_{2.1}$. Hence, $X^2(2|1)$ is asymptotically distributed as $\sum_1^u \delta_i(2|1)Z_i$, using a standard result on the distribution of a quadratic form in normal variables, where the $\delta_i(2|1)$ are eigenvalues of $(\tilde{X}_2'\Delta\tilde{X}_2)^{-1}(\tilde{X}_2'D(W)VD(W)\tilde{X}_2)$. Replacing Δ, W and V by their estimates $\hat{\Delta}$, $\hat{W}$ and $\hat{V}$ respectively, we get (2.21). It can be shown that $G^2(2|1)$ is asymptotically equivalent to $X^2(2|1)$ under $H_{2.1}$ so that the above result also holds in the case of $G^2(2|1)$.

A Wald statistic under binomial sampling is sometimes used, instead of $X^2(2|1)$ or $G^2(2|1)$, to test $H_{2.1}$. Noting that $H_{2.1}$ is equivalent to $H_{2.1}'$: $\zeta = H'\nu = 0$, where $\nu = (\nu_1,\ldots,\nu_I)'$ and H is any $I \times u$ matrix of rank u with $H'X_1 = 0$ and $H'X_2$ nonsingular, the Wald statistic is given by

$$\tilde{X}^2_W(2|1) = \hat{\phi}' \hat{V}^{-1}_{0\phi} \hat{\phi} . \tag{A.14}$$

Here $\hat{\phi} = H'\hat{\nu}$, $\hat{\nu}_i = \ln\{\hat{f}_i/(1-\hat{f}_i)\}$ and $\hat{V}_{0\phi}$ is the estimate of $V_{0\phi}$ given below. As in the case of $X^2(2|1)$ the true asymptotic null distribution of $\tilde{X}^2_W(2|1)$ is a weighted sum, $\Sigma\gamma_i(2|1)Z_i$, of independent $\chi^2(1)$ variables with weights $\gamma_1(2|1),\ldots,\gamma_u(2|1)$ given by the eigenvalues of $V^{-1}_{0\phi}V_\phi$, where $V_\phi = H'XV_\beta X'H$ and $V_{0\phi} = n^{-1}H'X(X'\Delta X)^{-1}X'H$ is the corresponding expression under binomial sampling. The formula for V_ϕ follows from the approximation

$$\sqrt{n}(\hat{\phi}-\phi) \sim H'\Delta^{-1}D(W)\{\sqrt{n}(\hat{f}-f)\} \sim H'X\{\sqrt{n}(\hat{\beta}-\beta)\}, \tag{A.15}$$

using (A.4). It follows from (A.14), (A.15) and (A.12) that $\tilde{X}^2_W(2|1)$ is asymptotically equivalent to $X^2(2|1)$ under $H_{2.1}$, noting that $H'X(\hat{\beta}-\beta) = H'X_2\hat{\beta}_2$ and $H'X(X'\Delta X)^{-1}X'H = H'X_2(\tilde{X}_2'\Delta\tilde{X}_2)^{-1}\tilde{X}_2'H_2$. This result implies that

$$\{\gamma_1(2|1),\ldots,\gamma_u(2|1)\} \text{ is identical to } \{\delta_1(2|1),\ldots,\delta_u(2|1)\}. \tag{A.16}$$

III. Asymptotic null distribution of X^2

The asymptotic null distribution of X^2 (or G^2) can be obtained as a special case of the result for nested hypothesis $H_{2.1}$, by treating the model M_1 as a saturated model. We have $\hat{f} = p$ in the saturated case so that from (A.15) $V_\phi = n^{-1}H'\Delta^{-1}D(W)VD(W)\Delta^{-1}H$ and $V_{0\phi} = n^{-1}H'\Delta^{-1}X$. Hence, from (A.16), X^2 is asymptotically distributed as $\Sigma\delta_iZ_i$, where $\delta_1,\ldots,\delta_{I-s}$ are the eigenvalues of $V^{-1}_{0\phi}V_\phi$.

REFERENCES

BIRCH, M.W. (1964). A new proof of the Pearson-Fisher theorem. *Ann. Math. Statist.* *35*, 818-24.

BLOCH, F.E. & SMITH, S.P. (1977). Human capital and labour market employment. *J. Human Resources 12*, 550-9.

COOK, R.D. (1977). Detection of influential observations in linear regression. *Technometrics 19*, 15-18

COOK, R.D. & WEISBERG, S. (1982). *Residuals and Influence in Regression.* London: Chapman and Hall.

COX, D.R. (1970). *Analysis of Binary Data.* London: Chapman and Hall.

FAY, R.E. (1985). Replication approaches to the log-linear analysis of data from complex samples. In *Recent Developments in the Analysis of Large-scale Data Sets*, pp. 95-118, Luxembourg: Office for Official Publications of the European Communities.

HOLT, D. & EWINGS, P.O. (1985). Personal communication.

KALBFLEISCH, J.G. (1984). Aspects of categorical data analysis. In *Topics in Applied Statistics* (Y.P. Chaubey and T.D. Dwivedi, eds.), Department of Mathematics, Concordia University, Montreal, 139-150.

McCULLAGH, P. (1980). Regression models for ordinal data (with discussion). *J.R. Statist. Soc.*, *B*, *42*, 104-42.

McCULLAGH, P. & NELDER, J.A. (1983). *Generalized Linear Models.* London: Chapman and Hall.

PREGIBON, D. (1981). Logistic regression diagnostics. *Ann. Statist. 9*, 705-24.

RAO, J.N.K. & SCOTT, A.J. (1984). On chi-squared tests for multiway contingency tables with cell proportions estimated from survey data. *Ann. Statist.* *12*, 46-60.

RAO, J.N.K. & SCOTT, A.J. (1985). On simple adjustments to chi-squared tests with survey data. Unpublished manuscript.

STATISTICS CANADA (1977). *Methodology of the Canadian Labour Force Survey, 1976*. Catalogue 71-526 occasional, Ottawa: Statistics Canada.

Table 1. Pseudo maximum likelihood estimates $\hat{\beta}_i$ and corresponding standard errors for the LFS data under model (3.1). Also, $X_W^2(2|1) = \hat{\beta}_i^2/\mathrm{var}(\hat{\beta}_i)$ and $G^2(2|1)/\hat{\delta}.(2|1)$ for the nested hypotheses $H_{2.1}:\beta_2 = 0$ and $H_{2.1}:\beta_4 = 0$.

i	$\hat{\beta}_i$	s.e. $(\hat{\beta}_i)$		$X_W^2(2\|1)$	$G^2(2\|1)/\hat{\delta}.(2\|1)$
		True	Binomial		
0	-2.76				
1	0.209	0.013	0.012		
2	-0.00217	0.000173	0.000136	157.3	102.1
3	0.0913	0.089	0.068		
4	0.00276	0.0041	0.0030	0.45	0.46

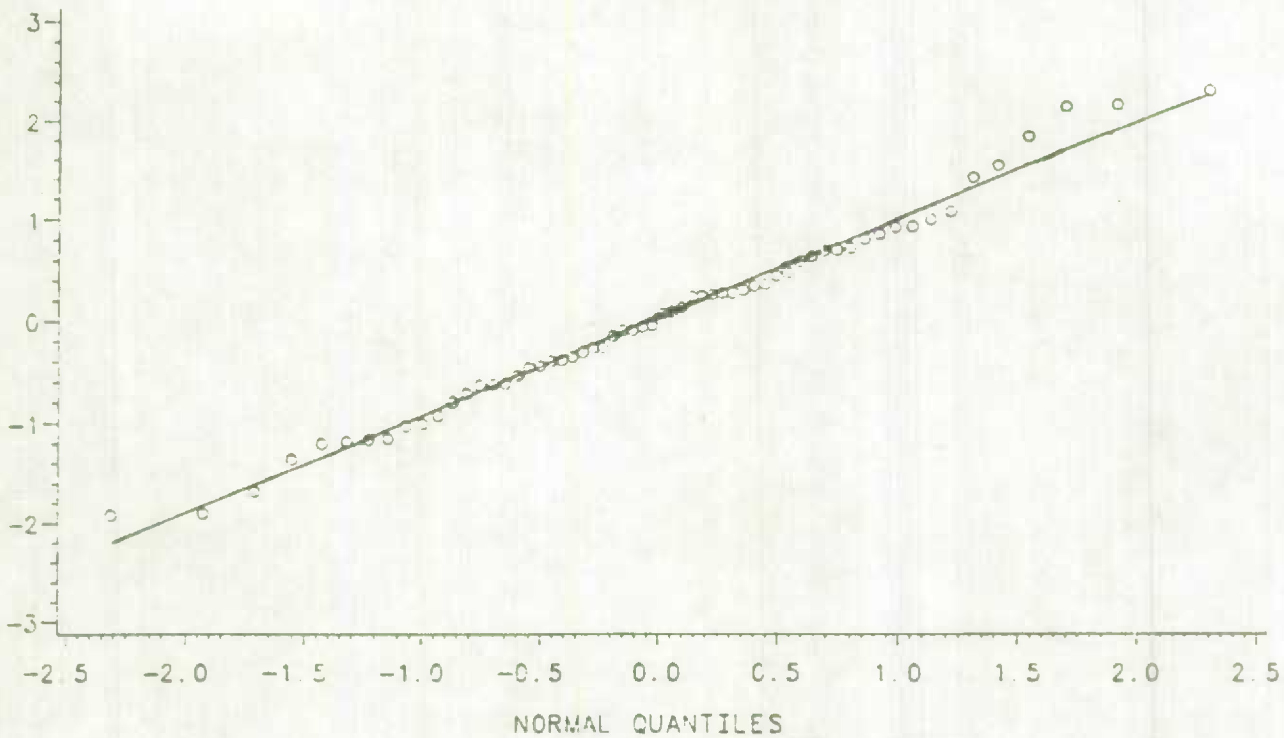

Figure 1: Normal Probability Plot of $\widetilde{G}_i$

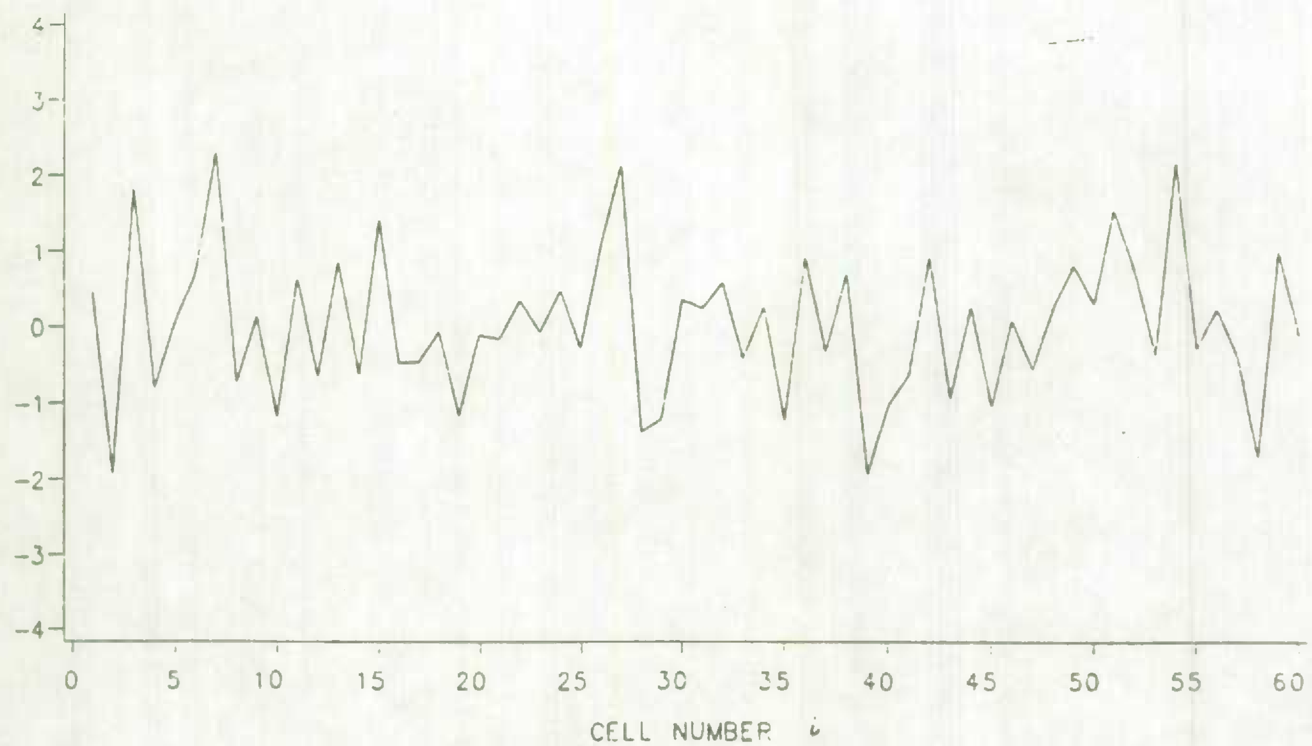

Figure 2: Index Plot of $\widetilde{G}_i$

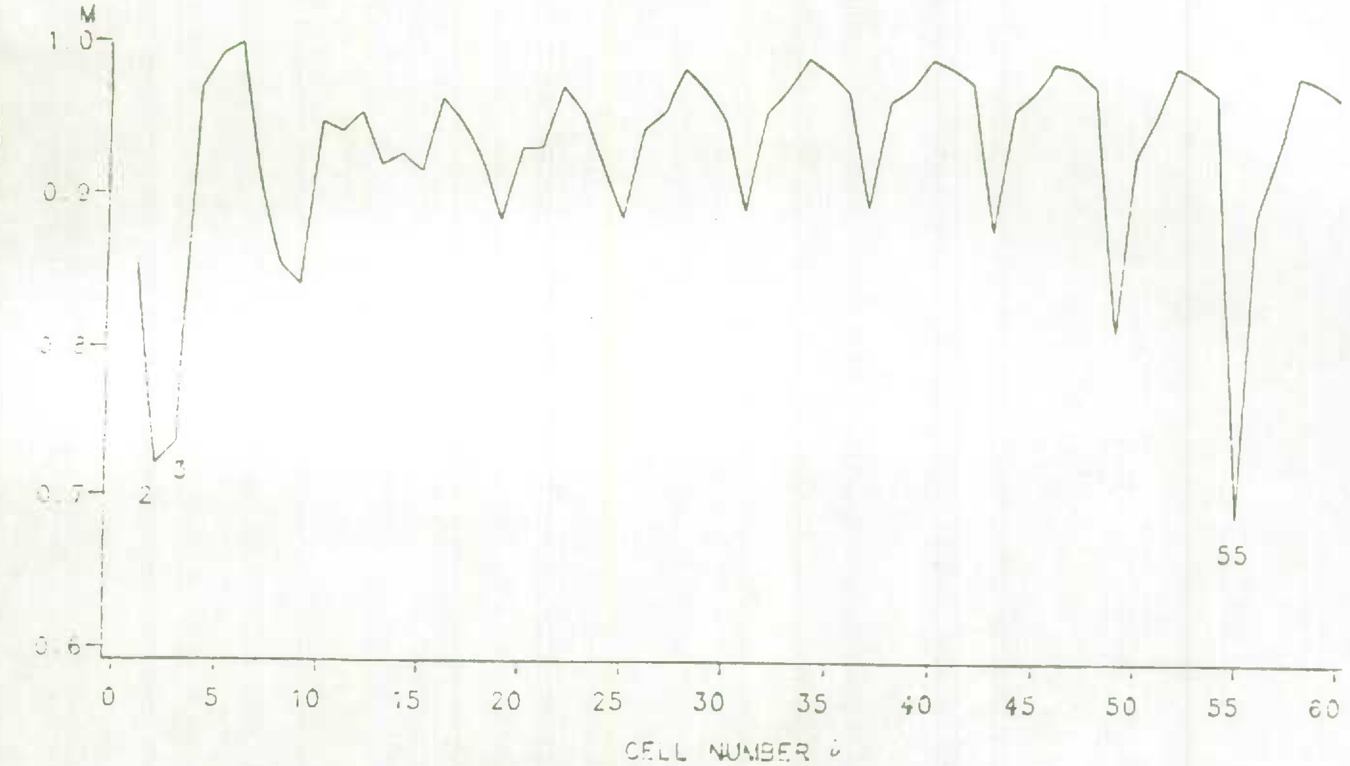

Figure 3: Index Plot of m_{ii}

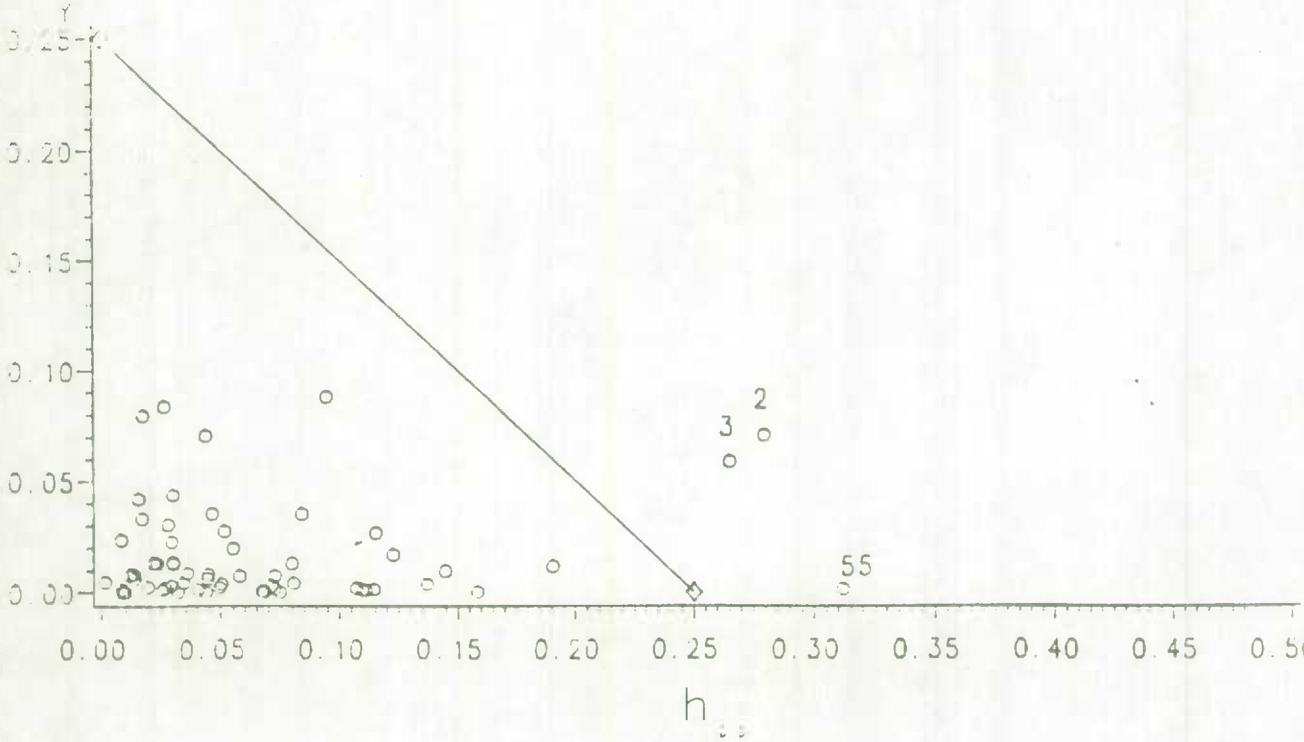

Figure 4: Scatter Plot of χ_i^2/χ^2 vs h_{ii}

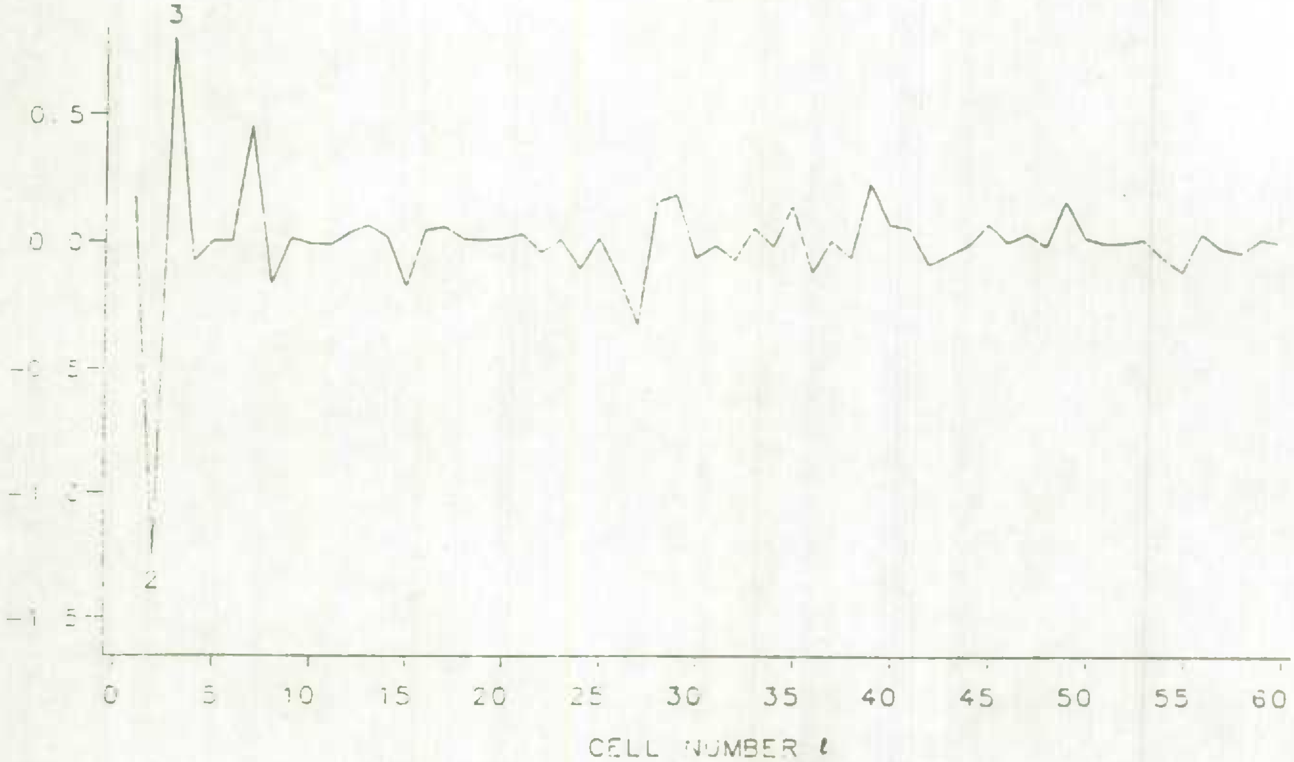

Figure 5: Index Plot of $\{\hat{\beta}_0 - \hat{\beta}_0(-\ell)\}/\text{s.e.}(\hat{\beta}_0)$

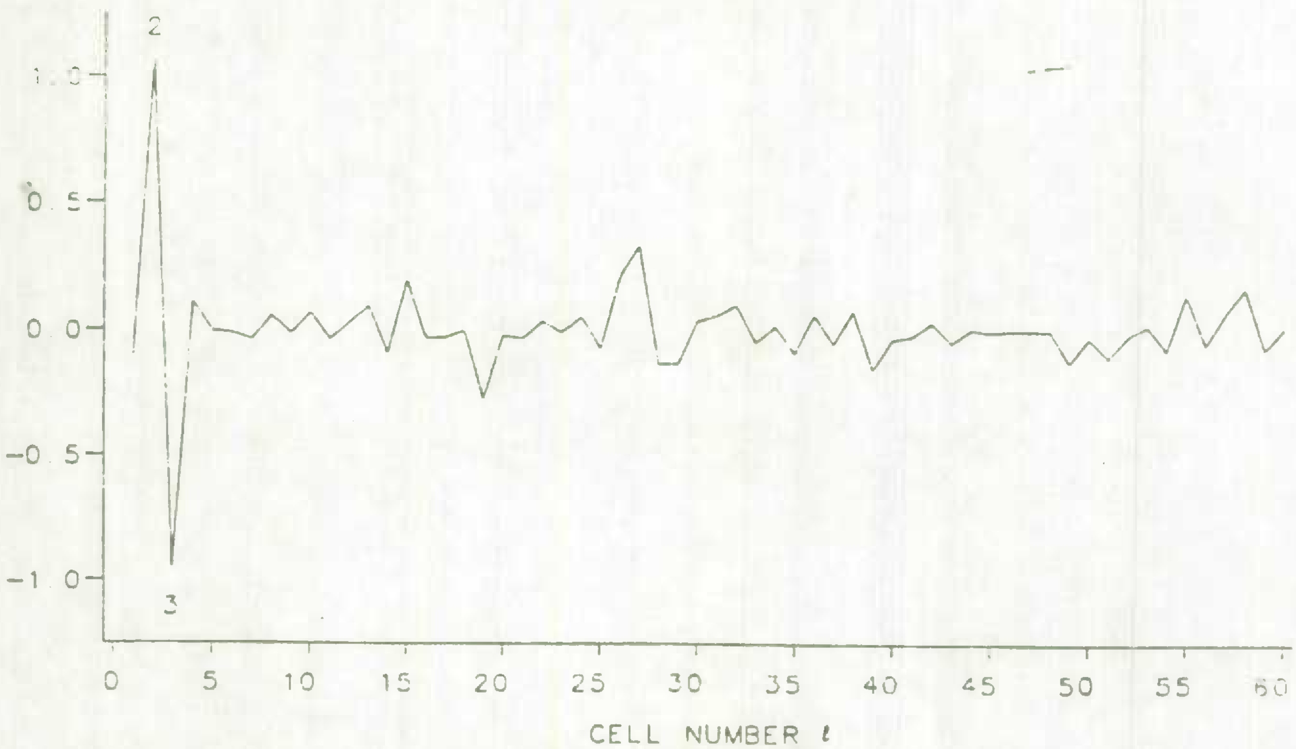

Figure 6: Index Plot of $\{\hat{\beta}_1 - \hat{\beta}_1(-\ell)\}/\text{s.e.}(\hat{\beta}_1)$

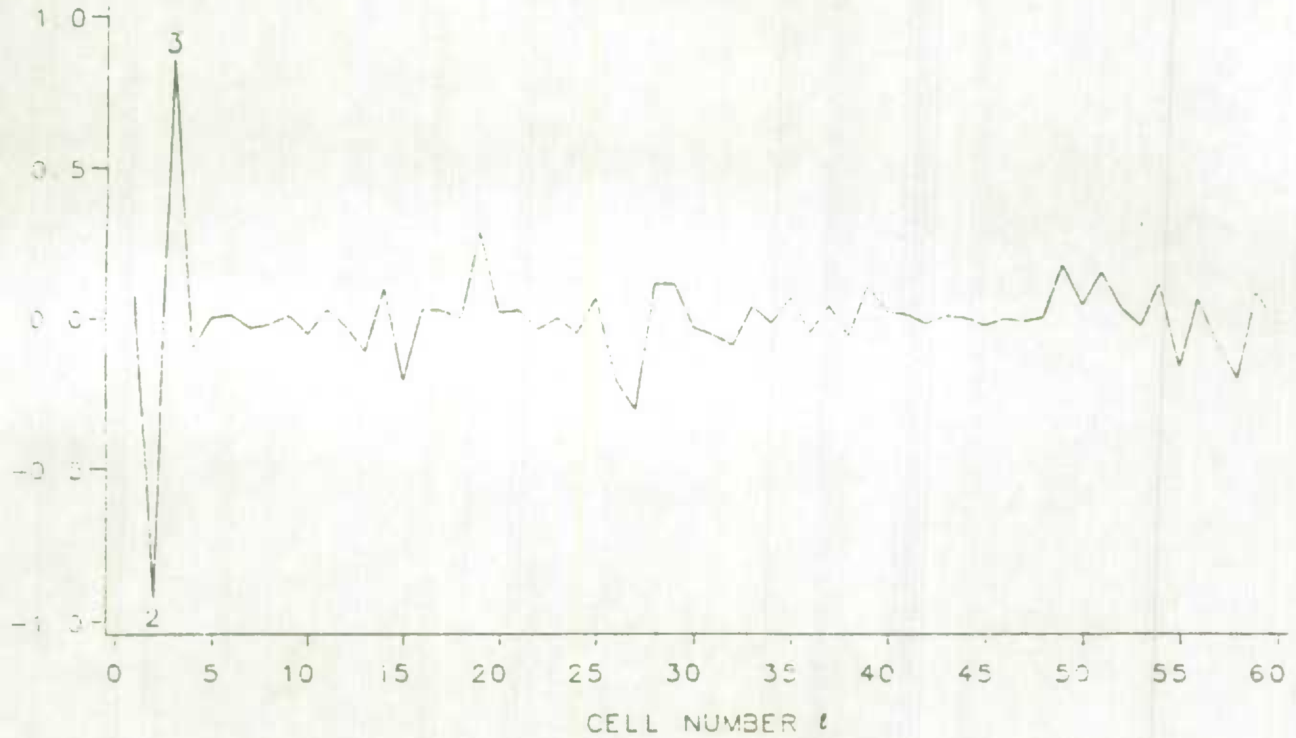

Figure 7: Index Plot of $\{\hat{\beta}_2 - \hat{\beta}_2(-\ell)\}/\text{s.e.}(\hat{\beta}_2)$

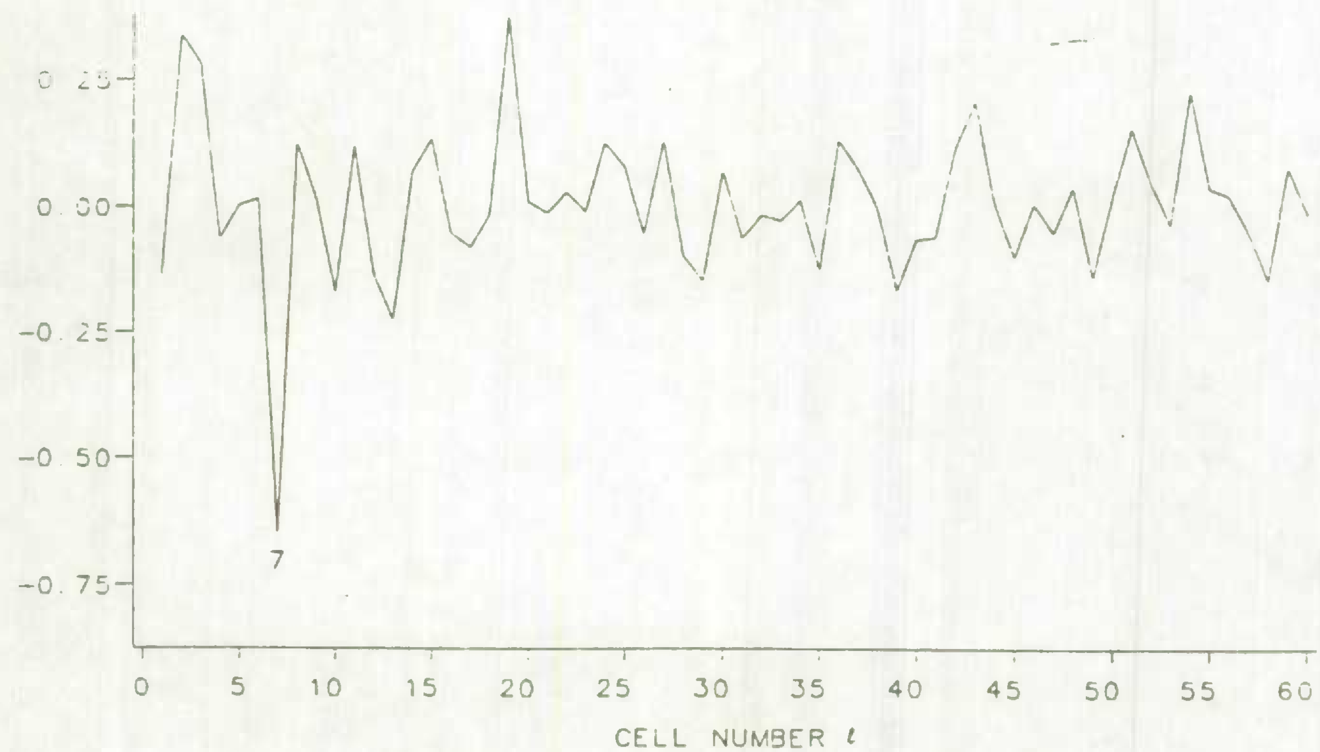

Figure 8: Index Plot of $\{\hat{\beta}_3 - \hat{\beta}_3(-\ell)\}/\text{s.e.}(\hat{\beta}_3)$

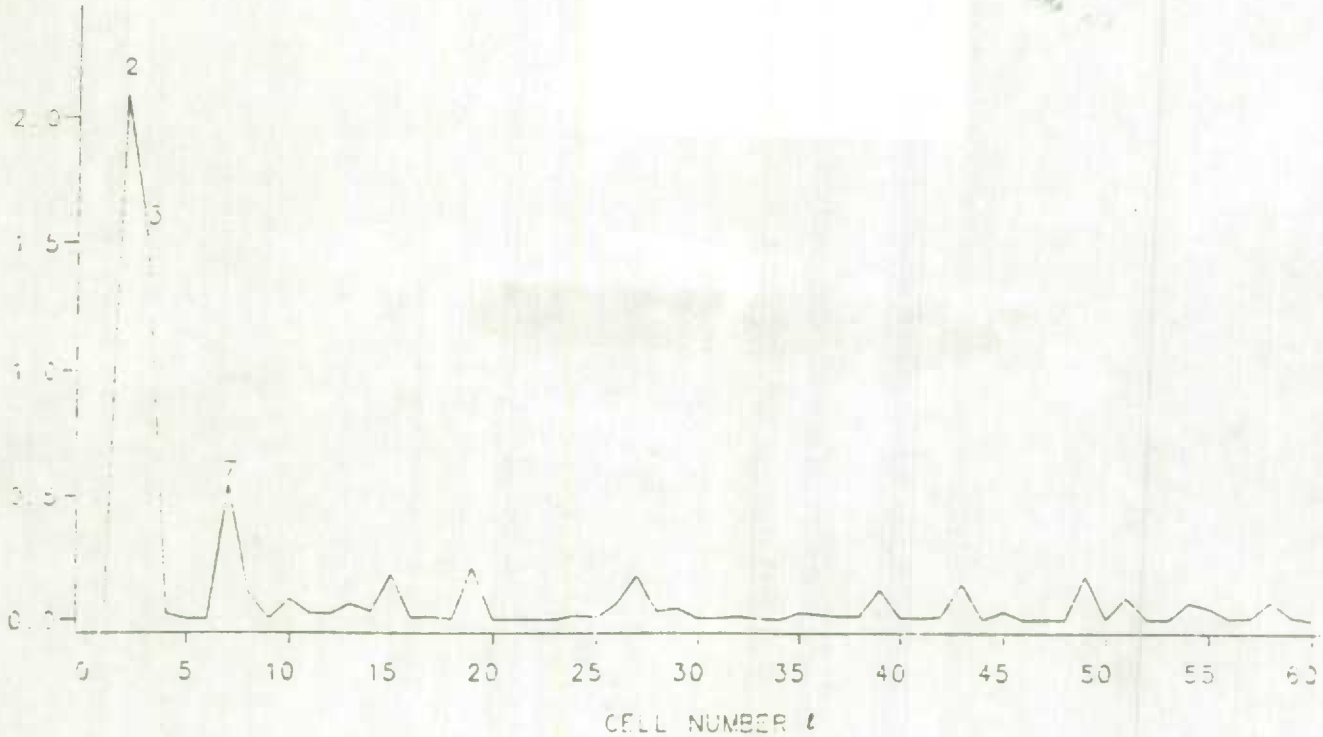

Figure 9: Index Plot of $\{G^2 - \tilde{G}^2(-\ell)\}/\hat{\hat{\delta}}$

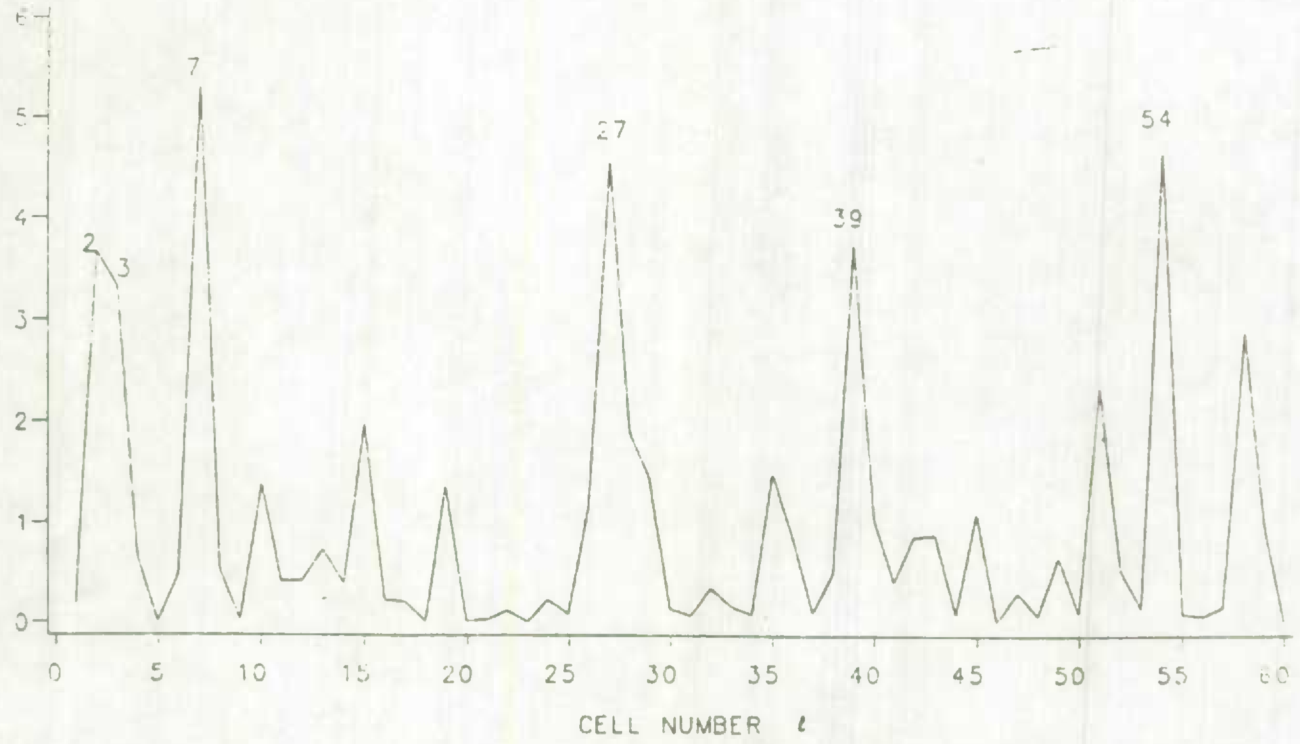

Figure 10: Index Plot of $\{G^2 - G^2(-\ell)\}/\hat{\hat{\delta}}$

Ca 005

STATISTICS CANADA LIBRARY
BIBLIOTHEQUE STATISTIQUE CANADA
1010148753