Statistics Canada Statistique Canada

Methodology Branch

Census & Household Survey Methods Division

Direction de la méthodologie

Division des méthodes de recensement et d'enquêtes ménages

Canada

FITTING BOX-COX TRANSFORMATION MODELS TO LABOUR FORCE SURVEY DATA

S. Kumar and J.N.K. Rao

Statistics Canada and Carleton University

Methodology Branch Working Paper

Number: CHSM 85-080E

FITTING BOX-COX TRANSFORMATION MODELS TO LABOUR FORCE SURVEY DATA

S. Kumar and J.N.K. Rao
Statistics Canada and Carleton University

Box-Cox transformation models are fitted to estimated proportions associated with a binary response variable. The survey design is taken into account by adjusting the standard chisquare (X^2) or the likelihood ratio (G^2) test statistic. The methods are applied to some data from the October 1980 Canadian Labour Force Survey (LFS). Comparisons are made to a previous analysis of the same data utilizing logistic regression models (Kumar and Rao, 1984).

1. INTRODUCTION

The analysis of variation in the estimated proportions associated with a binary response variable is of considerable interest to researchers in social, behavioral and health sciences. Logistic regression (or logit) methods for binomial proportions are inappropriate for analysing sample survey data due to clustering and stratification used in the survey design. Kumar and Rao (1984) utilized an adjustment to standard X^2 or G^2 test, based on certain generalized design effects, that takes account of the survey design and analysed some data from the October 1980 LFS. The sample consisted of males aged 15-64 who were in the labour force and not full-time students. They arrived at a simple logit model explaining the variation in estimated employment rates in the age-by-education cross-classification. Their model is given by:

$$\hat{\nu}_{jk} = \ln[\hat{f}_{jk}/(1-\hat{f}_{jk})] = -3.10 + 0.2111\, A_j - 0.00218\, A_j^2 + 0.1509\, E_k \qquad (1)$$

$$j = 1, .., 10; \quad k = 1, .., 6$$

where $\hat{f}_{jk}$ is the fitted employment rate in the (j,k) -th cell of the two-way table obtained by dividing the age interval [15, 64] into ten groups $[10 + 5j, 14 + 5j]$, $j = 1, .., 10$ and forming the education levels, E_k, by assigning to each person a value based on median years of schooling (E_k = 7, 10, 12, 13, 14 and 16), and A_j is the mid-point of the interval $[10 + 5j, 14 + 5j]$. The fitted unemployment rates $1-\hat{f}_{jk}$ are more precise than the corresponding survey estimates $1-\hat{P}_{jk}$, especially for cells with a small sample. Moreover, the bias of $1-\hat{f}_{jk}$, as an estimate of the population proportion $1-f_{jk}$, should be small since the model (1) provided an adequate fit.

Although the logit model (1) provided an adequate fit, it might be worthwhile to explore the possibility of a non-logit model providing a simpler or better fit. One such model was proposed by Guerrero and Johnson (1982) by utilizing a Box-Cox power transformation of the odds ratio $f_i/(1-f_i)$, $i = 1, .., I$, where I is the total number of cells in a table and f_i is the population proportion of responses in i-th cell. Their model is given by:

$$\nu_i(\lambda) = (f_i/(1-f_i))^{(\lambda)} = \underset{\sim}{x}_i'\underset{\sim}{\beta}, \quad i = 1, .., I \tag{2}$$

$$= \sum_{j=1}^{s} x_{ji}\beta_j,$$

where $\underset{\sim}{\beta}$ is the s-vector of unknown parameters, $\underset{\sim}{x}_i = (x_{1i}, \ldots, x_{si})'$, with $x_{1i}=1$, is the s-vector of known constants derived from the factor levels as in (1), and

$$(f_i/(1-f_i))^{(\lambda)} = \begin{cases} \ln(f_i/(1-f_i)) & \text{if } \lambda = 0 \\ \frac{1}{\lambda}[(f_i/(1-f_i))^{\lambda}-1] & \text{if } \lambda \neq 0 \ . \end{cases} \tag{3}$$

The model (2) includes the logit model as a special case ($\lambda = 0$). Guerrero and Johnson (1982) applied model (2) to some data from the National Survey of Household Income and Expenditures, Mexico to explain the variation in female participation in the Mexican labour force. They found that a value of $\lambda = -6.63$ provided a significantly better fit than the logit model ($\lambda = 0$). However, ,they have ignored the survey design and applied standard methods for binomial proportions.

In this article, the Box-Cox transformation model is fitted to the 1980 LFS data on unemployment rates, by taking account of the survey design. It is shown that the fitted transformation model is very close to the logit model (1) so that the previous logit analysis cannot be improved upon. We hope to apply the transformation model to Survey of Consumer Finances data on female participation rates, and make analysis similar to that of Guerrero and Johnson (1982), but adjusting the test statistics to account for the survey design.

2. SUMMARY OF THEORY

2.1 Estimates of λ, $\underset{\sim}{\beta}$ and f_i

For general sample designs, it is difficult to obtain appropriate likelihood functions. Hence, it is a common practice to use a "pseudo" maximum likelihood estimate of $\underset{\sim}{\theta}' = (\lambda, \underset{\sim}{\beta}')$ or $\underset{\sim}{f} = (f_1, \ldots, f_I)'$ obtained as the solution of "pseudo" likelihood equations for $\underset{\sim}{\beta}$ and λ given by:

$$\phi_u = \sum_{i=1}^{I} x_{ui} [1 + \lambda(\underset{\sim}{x}_i'\underset{\sim}{\beta})]^{-1} w_i(\hat{P}_i - f_i)=0, \qquad u = 1,\ldots, s$$

$$\phi = -\frac{1}{2} \sum_{i=1}^{I} (\underset{\sim}{x}_i'\underset{\sim}{\beta})^2 w_i (\hat{P}_i - f_i)=0, \quad \lambda = 0 \tag{4}$$

$$\phi = \frac{1}{\lambda}\sum_{i=1}^{I} [(\underset{\sim}{x}_i' \underset{\sim}{\beta})(1 + \lambda \underset{\sim}{x}_i'\underset{\sim}{\beta})^{-1} + \frac{1}{\lambda} \ln(1 + \lambda \underset{\sim}{x}_i'\underset{\sim}{\beta})^{-1}] w_i(\hat{P}_i - f_i) = 0, \quad \lambda \neq 0,$$

where $f_i = f_i(\underset{\sim}{\theta})$. The equations (4) are obtained from the corresponding binomial likelihood equations by replacing n_{i1}/n_i by the survey estimate $\hat{P}_i$ of f_i and n_i/n by the corresponding survey estimate, w_i, of i-th cell proportion, where n_{i1} is the number of sample responses from i-th cell and n_i is the corresponding sample size ($\Sigma n_i = n$). The solution

$\hat{\underset{\sim}{\theta}}' = (\hat{\lambda}, \hat{\underset{\sim}{\beta}}')$ of (4) is obtained iteratively by a quasi-Newton procedure. This would require the evaluation of partial derivatives of ϕ_u and ϕ with respect to $\underset{\sim}{\beta}$ and λ ($\lambda \neq 0$) - see the appendix. The estimate of f_i is given by $\hat{f}_i = f_i(\hat{\underset{\sim}{\theta}})$.

2.2 Estimated Variances and Covariances

Let $\hat{V}$ denote the estimated covariance matrix of the survey estimates $\hat{\underset{\sim}{P}}$, and let B be such that

$$D(\hat{\underset{\sim}{f}})D(\underset{\sim}{1}-\hat{\underset{\sim}{f}})B=(\partial \underset{\sim}{f}/\partial \underset{\sim}{\theta}) \;, \tag{5}$$

where $(\partial \underset{\sim}{f}/\partial \hat{\underset{\sim}{\theta}})$ is the $I\times(s+1)$ matrix of partial derivatives $\partial f_i/\partial \beta_j$ and $\partial f_i/\partial \lambda$ evaluated at $\hat{\underset{\sim}{\theta}}$, and $D(\hat{\underset{\sim}{f}}) = \mathrm{diag}(\hat{f}_1, \ldots, \hat{f}_I)$, $D(\underset{\sim}{1}-\hat{\underset{\sim}{f}}) =$ $\mathrm{diag}(1-\hat{f}_1, \ldots, 1-\hat{f}_I)$. The formula for $(\partial \underset{\sim}{f}/\partial \hat{\underset{\sim}{\theta}})$ is given in the Appendix. The estimated covariance matrix of $\hat{\underset{\sim}{\theta}}$ is given by:

$$\hat{V}(\hat{\underset{\sim}{\theta}}) = (B'\hat{\Delta}\,B)^{-1}(B'D(\underset{\sim}{w})\hat{V}D(\underset{\sim}{w})B)(B'\hat{\Delta}B)^{-1}, \tag{6}$$

where $\hat{\Delta} = \mathrm{diag}(w_1\hat{f}_1(1-\hat{f}_1), \ldots, w_I\hat{f}_I(1-\hat{f}_I))$, $D(\underset{\sim}{w}) = \mathrm{diag}(w_1, \ldots, w_I)$.

The estimated covariance matrix of fitted proportions, $\hat{\underset{\sim}{f}}$, can be obtained from $\hat{V}(\hat{\theta})$ as follows:

$$\hat{V}(\hat{\underset{\sim}{f}}) = D(\hat{\underset{\sim}{f}})D(\underset{\sim}{1}-\hat{\underset{\sim}{f}})B\hat{V}(\hat{\underset{\sim}{\theta}})B'D(\hat{\underset{\sim}{f}})D(\underset{\sim}{1}-\hat{\underset{\sim}{f}}) \tag{7}$$

The diagonal elements of $\hat{V}(\hat{\underset{\sim}{f}})$ provide the variance estimates of $\hat{f}_i$ or $1 - \hat{f}_i$ ($i = , \ldots, I$).

It is also of interest to find the standard errors of the residuals $r_i = \hat{P}_i - \hat{f}_i$ since the standardized residuals $r_i/\text{s.e.}(r_i)$ may be used to detect any outlying cell proportions. The estimated covariance matrix of $\underset{\sim}{r} = (r_1, \ldots, r_I)'$ is given by

$$\hat{V}(\underset{\sim}{r}) = A\hat{V}A', \tag{8}$$

where

$$A = \underset{\sim}{I} - D(\hat{\underset{\sim}{f}})D(\underset{\sim}{1}-\hat{\underset{\sim}{f}})B(B'\hat{\Delta}B)^{-1}B'D(\underset{\sim}{w}) \tag{9}$$

and $\underset{\sim}{I}$ is the identity matrix. The diagonal elements $\hat{V}_{ii}(r)$ of (8) provide the variance estimates of r_i, $i = 1, \ldots, I$.

2.3 Goodness-of-fit-tests

The standard chi-squared test of goodness-of-fit of the model (2) is given by:

$$X^2 = n \sum_{i=1}^{I} \frac{(\hat{P}_i - \hat{f}_i)^2 w_i}{\hat{f}_i(1 - \hat{f}_i)} = \sum_{i=1}^{I} X_i^2 \text{ (say)}. \tag{10}$$

The likelihood ratio test statistic is given by:

$$G^2 = 2n \sum_{i=1}^{I} w_i \left[\hat{P}_i \ln \frac{\hat{P}_i}{\hat{f}_i} + (1-\hat{P}_i) \ln \frac{(1-\hat{P}_i)}{(1-\hat{f}_i)} \right] = \sum_{i=1}^{I} G_i^2 \text{ (say)}, \tag{11}$$

where $G_i^2 = -2nw_i \ln(1-\hat{f}_i)$ at $\hat{P}_i = 0$ and $G_i^2 = -2nw_i \ln \hat{f}_i$ at $\hat{P}_i = 1$.

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 - 1 degrees of freedom, but for general designs this result is no longer valid. In fact, X^2(or G^2) is asymptotically distributed as $\Sigma\delta_i W_i$, where δ_i (i = 1, .., I - s - 1) are certain "generalized design effects" and the W_i are independent χ^2 variables each with 1 degree of freedom (d.f.)(Roberts, 1984). Under binomial sampling, $\delta_i=1$ for all i and $\Sigma\delta_i W_i$ reduces to χ^2 with I - s - 1 degrees of freedom.

A first-order correction to X^2 (or G^2) is obtained by treating $X_c^2 = X^2/\delta_.$ or $G_c^2 = G^2/\delta_.$ as χ^2 with I - s - 1 degrees of freedom under the hypothesis that the model (2) is true, where

$$(I - s - 1)\delta_. = n \sum_{i=1}^{I} \hat{V}_{ii}(r)w_i/[\hat{f}_i\ (1 - \hat{f}_i)]. \qquad (12)$$

A better approximation is obtained by treating

$$X_S^2 = \frac{X_c^2}{1+a^2} \text{ or } G_S^2 = \frac{G_c^2}{1+a^2} \text{ as } \chi^2 \text{ with } \nu = \frac{I-s-1}{1+a^2} \text{ d.f.,} \qquad (13)$$

where

$$a^2 = \left[\sum_{i=1}^{I-s-1} \delta_i^2 - (I-s-1)\delta_.^2\right]/(I-s-1)\delta_.^2$$

is the square of coefficient of variation (C.V.) of the δ_i and

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

where $\hat{V}_{ij}(r)$ is the (i,j) the element of the covariance matrix of $\underset{\sim}{r}$, $\hat{V}(\underset{\sim}{r})$.

In the logit case ($\lambda = 0$), Kumar and Rao (1984) have also given a Wald Statistic for goodness-of-fit of the model which is asymptotically χ^2 with $(I-s)$ d.f. It seems more complex to construct a similar Wald Statistic for the goodness-of-fit of the transformation model (2).

2.4 Nested Hypothesis

We are often interested in testing the nested hypothesis $H: \underset{\sim}{\beta}_2 = \underset{\sim}{0}$ given the model (2): $\nu_i(\lambda) = \underset{\sim}{x}_i' \underset{\sim}{\beta} = \underset{\sim}{x}_{i(1)}' \underset{\sim}{\beta}_1 + \underset{\sim}{x}_{i(2)}' \underset{\sim}{\beta}_2$, where $\underset{\sim}{\beta}_1$ is $r \times 1$ and $\underset{\sim}{\beta}_2$ is $q \times 1$ and $\underset{\sim}{x}_{i(1)}' = (x_{1i}, \ldots, x_{ri})$, $x_{i(2)}' = (x_{r+1,i}, \ldots, x_{si})$, $r+q = s$. The "pseudo m.l.e." under H are obtained from (4) by replacing $\underset{\sim}{\beta}$ by $\underset{\sim}{\beta}_1$, $\underset{\sim}{x}_i'$ by $\underset{\sim}{x}_{i(1)}'$ and s by r, and using iterative calcualtions. The chisquare and likelihood ratio tests of H are respectively given by

$$X^2(2|1) = n \sum_{i=1}^{I} w_i \frac{(\hat{f}_i - \hat{\hat{f}}_i)^2}{\hat{\hat{f}}_i(1-\hat{\hat{f}}_i)} \tag{14}$$

and

$$G^2(2|1) = 2n \sum_{i=1}^{I} w_i [\hat{f}_i \ln \frac{\hat{f}_i}{\hat{\hat{f}}_i} + (1-\hat{f}_i) \ln \frac{1-\hat{f}_i}{1-\hat{\hat{f}}_i}], \tag{15}$$

where $\hat{\hat{f}}_i = f_i(\hat{\hat{\lambda}}, \hat{\hat{\underset{\sim}{\beta}}}_1')$ and $\hat{\hat{\underset{\sim}{\beta}}}_1$ and $\hat{\hat{\lambda}}$ are the pseudo m.l.e. of $\underset{\sim}{\beta}_1$ and λ under H. Under independent binomial sampling, both $X^2(2|1)$ and $G^2(2|1)$ are asymptotically distributed as χ^2 with q degrees of freedom, but for general designs this result is no longer valid (Roberts, 1984).

A simple adjustment to $X^2(2|1)$ or $G^2(2|1)$ is obtained, as in the case of goodness-of-fit, by treating $X^2(2|1)/\delta_.(H)$ or $G^2(2|1)/\delta_.(H)$ as χ^2 with q d.f. under H, where

$$q\delta_.(H) = n \sum_{i=1}^{I} \hat{V}_{ii}(r(H))\, w_i/[\hat{\hat{f}}_i(1-\hat{\hat{f}}_i)] \;. \tag{16}$$

In (16) $\hat{V}_{ii}(\underset{\sim}{r}(H))$, the variance estimate of the residual $r_i(H) = \hat{f}_i - \hat{\hat{f}}_i$, is obtained as the i-th diagonal element of the covariance matrix of the $r_i(H)$:

$$\hat{V}(\underset{\sim}{r}(H)) = D(\hat{\underset{\sim}{f}})D(\underset{\sim}{1}-\hat{\underset{\sim}{f}})(\tilde{B}_2 A_2 \tilde{B}_2')D(\hat{\underset{\sim}{f}})D(\underset{\sim}{1}-\hat{\underset{\sim}{f}}), \tag{17}$$

where

$$\tilde{B}_2 = B_2 - B_1(B_1'\hat{\Delta}B_1)^{-1}(B_1'\hat{\Delta}B_2)$$

$$A_2 = (\tilde{B}_2'\hat{\Delta}\tilde{B}_2)^{-1}(\tilde{B}_2'D(\underset{\sim}{w})\hat{V}D(\underset{\sim}{w})\tilde{B}_2)^{-1}(\tilde{B}_2'\hat{\Delta}\tilde{B}_2)$$

$$D(\hat{\underset{\sim}{f}})D(\underset{\sim}{1}-\hat{\underset{\sim}{f}})(B_1|B_2) = [(\frac{\partial \underset{\sim}{f}}{\partial \underset{\sim}{\theta}_1}) \mid (\frac{\partial \underset{\sim}{f}}{\partial \underset{\sim}{\theta}_2})]$$

and

$$\hat{\underset{\sim}{\theta}}_1' = (\hat{\lambda}, \hat{\underset{\sim}{\beta}}_1'), \; \hat{\underset{\sim}{\theta}}_2 = \hat{\underset{\sim}{\beta}}_2.$$

Again, a better approximation is obtained by treating

$$X_S^2(2|1) = \frac{X^2(2|1)}{\delta_.(H)[1+a^2(H)]} \text{ or } G_s^2(2|1) = \frac{G^2(2|1)}{\delta_.(H)[1+a^2(H)]} \tag{18}$$

as χ^2 with $\nu(H) = q/[1+a^2(H)]$ d.f. under H, where

$$a^2(H) = [\sum_1^q \delta_i^2(H) - q\delta_.^2(H)]/\, q\delta_.^2(H)$$

and

$$\sum_1^q \delta_i^2(H) = \sum_{i=1}^{I}\sum_{j=1}^{I} \hat{V}_{ij}^2(\underset{\sim}{r}(H))(nw_i)(nw_j)/[\hat{f}_i\hat{\hat{f}}_j(1-\hat{\hat{f}}_i)(1-\hat{\hat{f}}_j)],$$

where $\hat{V}_{ij}(\underset{\sim}{r}(H))$ is the (i,j)-th element of covariance matrix of the $r_i(H)$ given by (17).

3. APPLICATION TO LFS

We now apply the results of Section 2 to LFS data described in Section 1 and previously analysed by fitting the logit model (1). Prompted by the model (1), we consider the following transformation model (2):

$$\nu_{jk}(\lambda) = \beta_o + \beta_1 A_j + \beta_2 A_j^2 + \beta_3 E_k, \qquad j = 1, \ldots, 10; \quad k = 1, \ldots, 6. \qquad (19)$$

Table 1 provides the pseudo m.l.e. of $\tilde{\theta}' = (\lambda, \tilde{\beta}')$ and the test statistics X^2, $X^2/\delta.$, X_S^2, G^2, $G^2/\delta.$ and G_S^2 under the model (19). The corresponding values under the logit model ($\lambda = 0$) are also given for comparison.

It is clear from Table 1 that the value of X^2 (or G^2) or the value of the adjusted statistic $X^2/\delta_.$ (or $G^2/\delta_.$) for the transformation model is essentially equal to the corresponding value under the logit model. Thus the transformation model provides no improvement in fit over the logit model. This is also clear from the value of $\hat{\lambda}$ (= 0.016) which is not significantly different from $\lambda = 0$ when compared to its standard error (0.085). The estimates of regression coefficients are essentially equal under the two models, but the C.V. of $\hat{\beta}_i$ is much larger than the corresponding C.V. under the logit model, due to the large C.V. associated with $\hat{\lambda}$ and the fact that $\hat{\beta}_i$ depends on $\hat{\lambda}$.

As in the case of the logit model, we would reject the model (19) if the survey design is ignored and the value of X^2 (or G^2) is referred to $\chi^2_{0.05}(55) = 73.3$, the upper 5% point of χ^2 variable with I-s-1 = 55 d.f. On the other hand, the value of $X^2/\delta_.$ (or $G^2/\delta_.$) when compared to 73.3, indicated that the model is adequate, the significance level (or P-value) being approximately equal to 0.40. Moreover, in the present context with s (=4) relatively small compared to I (=60), the simple correction $X^2/d_.$ (Fellegi, 1980), depending only on the average cell deff $d_.$, is very close to $X^2/\delta_.$ requiring the

Table 1: Pseudo m.l.e. and Test Statistics under the Transformation Model (19) and the corresponding Logit Model ($\lambda = 0$)

	Transformation Model		Logit Model	
		s.e.		s.e.
$\hat{\beta}_0$	-3.28	(0.975)	-3.10	(0.247)
$\hat{\beta}_1$	0.219	(0.0468)	0.211	(0.013)
$\hat{\beta}_2$	-0.00227	(0.00049)	-0.00218	(0.00017)
$\hat{\beta}_3$	0.1579	(0.0385)	0.1509	(0.0115)
$\hat{\lambda}$	0.016	(0.085)	-	
X^2	99.6		99.8	
G^2	102.6		102.5	
$\delta_.$	1.75		1.85	
$X^2/\delta_.$	57.1		53.8	
$G^2/\delta_.$	58.8		55.3	
d.f. for $\frac{X^2}{\delta_.}$ or $\frac{G^2}{\delta_.}$	55		56	
X_S^2	40.7		23.4	
G_S^2	42.0		23.9	
d.f. for X_S^2 or G_S^2	39.2		24.2	
a^2	0.40		1.31	
X_S^2 (0.05)	54.6		47.7	
G_S^2 (0.05)	56.4		48.9	
$d.$	1.905		1.905	
$X^2/d.$	52.3		52.4	
$G^2/d.$	53.9		53.8	

knowledge of estimated covariance matrix of the survey estimates $\hat{P}_{jk}$.

The value of Satterthwaite correction X_S^2 (of G_S^2) is larger than the corresponding value under the logit model due to a smaller C.V. of the δ_i, i.e., the a-value under the transformation model. The value of X_S^2 (or G_S^2) when adjusted to refer to $\chi^2_{0.05}(55)$, denoted as $X_S^2(0.05)$ (or $G_S^2(0.05)$) in Table 1, is also not significant.

Given the model (19), it is of interest to test for the possibility of a simpler model, involving only the linear effects of age and education, providing an adequate fit, i.e. test the hypothesis $H: \beta_2 = 0$. In the logit case, the test statistic $X^2(2|1)/\delta_.(H)$ or $G^2(2|1)/\delta_.(H)$ turned out to be highly significant, but the possibility of a transformation model with a λ- value significantly different from zero and providing adequate fit under the simpler model exists. We obtained $\hat{\hat{\lambda}} = 0.223$ under the simpler model, but

$$X^2(2|1)/\delta_.(H) = 208.6 \text{ or } G^2(2|1)/\delta_.(H) = 181.4$$

is highly significant, when referred to $\chi^2_{0.01}(1) = 6.6$, the upper 1% point of χ^2 variable with q = 1 d.f.

4. DIAGNOSTICS

Kumar and Rao (1984) developed diagnostic procedures for the logistic regression to detect any outlying cell proportions and influential points in the factor space, after making necessary adjustments to account for the survey design. The diagnostic procedures for the transformation model are analogous to those for the logitic regression model.

APPENDIX

A.1 Partial Derivatives $\partial f_i/\partial\beta_j$ and $\partial f_i/\partial\lambda$

$$\frac{\partial f_i}{\partial \beta_j} = x_{ji} f_i^2 \ Q_i^{-1-1/\lambda}$$

$$\frac{\partial f_i}{\partial \lambda} = f_i^2 \ (Q_i \ln Q_i - Q_i + 1)\lambda^{-2} \ Q_i^{-1-1/\lambda},$$

where $Q_i = 1 + \lambda \sum_j x_{ji}\beta_j$.

A.2 Partial Derivatives of ϕ_u and ϕ with respect to β_j and λ

$$\frac{\partial \phi_u}{\partial \beta_j} = \sum_{i=1}^{I} x_{ui} \left[\frac{-w_i(\hat{P}_i - f_i)\lambda}{Q_i^2} - \frac{w_i f_i^2}{Q_i^{2+1/\lambda}} \right] x_{ji}$$

$$\frac{\partial \phi_u}{\partial \lambda} = \sum_{i=1}^{I} x_{ui} \left[\frac{w_i(\hat{P}_i - f_i)(1-Q_i)}{\lambda Q_i^2} + \frac{w_i f_i^2 (Q_i - 1 - Q_i \ln Q_i)}{\lambda^2 \ Q_i^{2+1/\lambda}} \right]$$

$$\frac{\partial \phi}{\partial \beta_j} = \frac{1}{\lambda^2} \sum_{i=1}^{I} \left[\frac{\lambda w_i(\hat{P}_i - f_i)(1-Q_i)}{Q_i^2} + \frac{w_i f_i^2 (1 - Q_i + Q_i \ln Q_i)}{Q_i^{2+1/\lambda}} \right] x_{ji}$$

$$\frac{\partial \phi}{\partial \lambda} = -\frac{2}{\lambda^3} \sum_{i=1}^{I} \frac{1}{Q_i} \left[(Q_i - 1 - Q_i \ln Q_i) w_i (\hat{P}_i - f_i) \right]$$
$$- \frac{1}{\lambda^2} \sum_{i=1}^{I} \left[\frac{1}{\lambda Q_i} (1-Q_i)^2 w_i (\hat{P}_i - f_i) + \frac{1}{\lambda^2 Q_i^{2+1/\lambda}} w_i f_i^2 (1 - Q_i + Q_i \ln Q_i)^2 \right]$$

REFERENCES

[1] Fellegi, I.P. (1980). Approximate tests of indenpendence and goodness of fit based on stratified multistage samples, J. Amer. Statist. Assoc., 75, pp. 261-268.

[2] Guerrero, V.M. and Johnson, R.A. (1982). Use of the Box-Cox transformation with binary response models. Biometrika, 69, pp. 309-314.

[3] Kumar, S. and Rao, J.N.K. (1984). Logistic regression analysis of labour force survey data. Survey Methodology, 10, pp. 62-81.

[4] Roberts, G. (1984). On chisquare tests for Box-Cox transformation models with cell proportions estimated from survey data. Unpublished manuscript, Carleton University.