Econ 203c, Problem Set 4, Spring 2003

Question 1

$$A_n(\theta_0)A_n(\theta_0)' = \left(1/n\sum \frac{\partial \ln f(y_i,x_i;\theta_0)}{\partial\theta}\right)\left(1/n\sum \frac{\partial \ln f(y_i,x_i;\theta_0)}{\partial\theta'}\right)$$

By Law of Large Numbers, $A_n(\theta_0) \overset{p}{\to} E(\frac{\partial \ln f(y,x;\theta_0}{\partial\theta})$

Because θ_0 is the population parameter, $E(\frac{\partial \ln f(y,x;\theta_0)}{\partial\theta}) = 0$

By Slutsky, $A_n(\theta_0)A_n(\theta_0)' \overset{p}{\to} E(\frac{\partial \ln f(y,x;\theta_0)}{\partial\theta})E(\frac{\partial \ln f(y,x;\theta_0)}{\partial\theta'}) = 0$

Therefore, $A_n(\theta_0)A_n(\theta_0)'$ is not a consistent estimator for $I(\theta_0)$. The outer product of the gradient or hessian forms are consistent estimators.

Question 2

$$f(x,y) = \frac{\theta e^{-(\beta+\theta)y}(\beta y)^x}{x!}$$

Because each of the sub-parts of this question rely on results from other parts, I'll answer it out of order by examining first all of the distribution questions.

Note that this joint density is a mixture of two distinct distributions:

$$f(x,y) = \frac{\theta e^{-(\beta+\theta)y}(\beta y)^x}{x!} = [\theta e^{-\theta y}][\frac{e^{-\beta y}(\beta y)^x}{x!}] = f_1(y) * f_2(x|y)$$

where $f_1(y) = \theta e^{-\theta y}$ is an exponential distribution with parameter θ.

and $f_2(x|y) = \frac{e^{-\beta y}(\beta y)^x}{x!}$ is a poison distribution with parameter βy.

This answers [5)] and [6)].

[3)] Show $f(x) = \gamma(1-\gamma)^x$

You could try to calculate this directly:

$$f(x) = \int_{y=0}^{\infty} \frac{\theta e^{-(\beta+\theta)y}(\beta y)^x}{x!} dy$$

Instead, let's use the other information provided in this problem.

If we assume the condtional distribution from question 4)

$$f(y|x) = \frac{\lambda e^{-\lambda y}(\lambda y)^x}{x!}$$

And use the definition of conditional probability:

$$f(y|x) = \frac{f(x,y)}{f(x)} \Rightarrow f(y|x)f(x) = f(x,y)$$

Then,

$$f(y|x)f(x) = \frac{\lambda e^{-\lambda y}(\lambda y)^x}{x!}\gamma(1-\gamma)^x = \frac{\gamma\lambda e^{-\lambda y}(\lambda y(1-\gamma))^x}{x!}$$

$$= \frac{\gamma\lambda e^{-\lambda y}(y(\lambda-\lambda\gamma))^x}{x!}$$

$$= f(x,y)$$

$$\Rightarrow \frac{\gamma\lambda e^{-\lambda y}(y(\lambda-\lambda\gamma))^x}{x!} = \frac{\theta e^{-(\beta+\theta)y}(\beta y)^x}{x!}$$

This implies (1) $\beta = (\lambda - \lambda\gamma)$, (2) $(\beta+\theta) = \lambda$, and (3) $\gamma\lambda = \theta$

(2) and (3) imply: $\beta + \gamma\lambda = \lambda \Rightarrow \beta = \lambda - \gamma\lambda \Rightarrow$(1)

Thus, $f(x) = \gamma(1-\gamma)^x$
and $\gamma = \theta/(\beta+\theta)$

Note that the marginal density for x is negative binomial with parameter γ (one trial) and $E(x) = (1-\gamma)/\gamma$, $V(x) = (1-\gamma)/\gamma^2$.

[4)] By the same argument, we can show that $f(y|x) = \frac{\lambda e^{-\lambda y}(\lambda y)^x}{x!}$.
Next note that $f(y|x)$ is a gamma distribution. The general pdf for a gamma distribution with parameters r and λ is

$$f(z) = \frac{\lambda^r}{(r-1)!}z^{r-1}e^{-\lambda z}$$

Here, define $(r-1) = x$ and $z = y$.

Then,

$$f(y|x,\lambda) = \frac{\lambda^{x+1}}{x!}y^x e^{-\lambda y} = \frac{\lambda(\lambda y)^x e^{-\lambda y}}{x!}.$$

Therefore, $f(y|x)$ integrates to 1 as this is a pdf.

From the gamma distribution, $E(y|x) = r/\lambda = \frac{x+1}{\lambda}$ and $V(y|x) = r/\lambda^2 = \frac{x+1}{\lambda^2}$

Now, I'll answer the MLE questions in order.

1)
a) MLE

$$\ln L(\theta, \beta, x, y) = \ln(\theta) - (\beta + \theta)y + x(\ln\beta + \ln y) + g(x)$$
$$\frac{\partial \ln L(\theta, \beta, x, y)}{\partial \theta} = 1/\theta - y$$

For a random sample of n observations,

$$\frac{\partial \ln L(\theta, \beta, x_1, \ldots, x_n, y_1, \ldots, y_n)}{\partial \theta} = n/\theta - \sum y_i$$
$$\Rightarrow \hat{\theta}_{ml} = \frac{n}{\sum y_i}$$

$$\frac{\partial \ln L(\theta, \beta, x, y)}{\partial \beta} = -y + x/\beta$$

For a random sample of n observations,

$$\frac{\partial \ln L(\theta, \beta, x_1, \ldots, x_n, y_1, \ldots, y_n)}{\partial \beta} = -\sum y_i + 1/\beta \sum x_i$$
$$\Rightarrow \hat{\beta}_{ml} = \frac{\sum x_i}{\sum y_i}$$

b) Asymptotic joint distribution

$$\sqrt{n}\left(\begin{bmatrix} \hat{\theta}_{ml} \\ \hat{\beta}_{ml} \end{bmatrix} - \begin{bmatrix} \theta \\ \beta \end{bmatrix}\right) \xrightarrow{d} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, I(\theta, \beta)^{-1}\right)$$

where $I(\theta, \beta) = \begin{bmatrix} -E\left(\frac{\partial^2 \ln L(\theta, \beta, x, y)}{\partial \theta^2}\right) & -E\left(\frac{\partial^2 \ln L(\theta, \beta, x, y)}{\partial \beta \partial \theta}\right) \\ -E\left(\frac{\partial^2 \ln L(\theta, \beta, x, y)}{\partial \theta \partial \beta}\right) & -E\left(\frac{\partial^2 \ln L(\theta, \beta, x, y)}{\partial \beta \partial \beta}\right) \end{bmatrix}$

$$-E\left(\frac{\partial^2 \ln L(\theta, \beta, x, y)}{\partial \theta^2}\right) = -E(-1/\theta^2) = 1/\theta^2$$

$$-E\left(\frac{\partial^2 \ln L(\theta,\beta,x,y)}{\partial\theta\partial\beta}\right) = -E\left(\frac{\partial^2 \ln L(\theta,\beta,x,y)}{\partial\beta\partial\theta}\right) = 0$$
$$-E\left(\frac{\partial^2 \ln L(\theta,\beta,x,y)}{\partial\beta\partial\beta}\right) = -E(-x/\beta^2) = E(x)/\beta^2$$

Since $I(\theta,\beta)$ is block diagonal:

$$I(\theta,\beta)^{-1} = \begin{bmatrix} \theta^2 & 0 \\ 0 & \beta^2/E(x) \end{bmatrix}$$

Thus,

$$\sqrt{n}\left(\begin{bmatrix} \widehat{\theta}_{ml} \\ \widehat{\beta}_{ml} \end{bmatrix} - \begin{bmatrix} \theta \\ \beta \end{bmatrix}\right) \overset{d}{\to} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \theta^2 & 0 \\ 0 & \theta\beta \end{bmatrix}\right)$$

since $E(x) = (1-\gamma)/\gamma = \dfrac{1-\theta/(\beta+\theta)}{\theta/(\beta+\theta)} = \dfrac{\beta}{\theta}$

2)
Let $\alpha = \theta/(\beta+\theta) = g(\theta,\beta)$

If $g(\theta,\beta)$ is continuous and continuously differentiable then,
by the invariance property (see Greene, p.473),

$$\widehat{\alpha}_{ml} = \widehat{\theta}_{ml}/(\widehat{\beta}_{ml} + \widehat{\theta}_{ml}) = \frac{\dfrac{n}{\sum y_i}}{\dfrac{\sum x_i}{\sum y_i} + \dfrac{n}{\sum y_i}} = \frac{n}{\sum x_i + n}$$

Asymptotic distribution:

$$\sqrt{n}(\widehat{\alpha}_{ml} - \alpha) \overset{d}{\to} N(0, I(\alpha)^{-1})$$

where $I(\alpha)^{-1} = \begin{bmatrix} \dfrac{\partial g(\theta,\beta)}{\partial\theta} & \dfrac{\partial g(\theta,\beta)}{\partial\beta} \end{bmatrix} I(\theta,\beta)^{-1} \begin{bmatrix} \dfrac{\partial g(\theta,\beta)}{\partial\theta} & \dfrac{\partial g(\theta,\beta)}{\partial\beta} \end{bmatrix}'$

$$= \begin{bmatrix} \dfrac{\partial g(\theta,\beta)}{\partial\theta} & \dfrac{\partial g(\theta,\beta)}{\partial\beta} \end{bmatrix} \begin{bmatrix} \theta^2 & 0 \\ 0 & \beta^2/E(x) \end{bmatrix} \begin{bmatrix} \dfrac{\partial g(\theta,\beta)}{\partial\theta} \\ \dfrac{\partial g(\theta,\beta)}{\partial\beta} \end{bmatrix}$$

$$\frac{\partial g(\theta,\beta)}{\partial\theta} = \frac{\partial g(\theta,\beta)}{\partial\theta} = -\theta(\beta+\theta)^{-2} + 1(\beta+\theta)^{-1} = \frac{\beta}{(\beta+\theta)^2}$$

$$\frac{\partial g(\theta,\beta)}{\partial\beta} = -\theta(\beta+\theta)^{-2} = -\frac{\theta}{(\beta+\theta)^2}$$

$$I(\alpha)^{-1} = \begin{bmatrix} \frac{\beta}{(\beta+\theta)^2} & -\frac{\theta}{(\beta+\theta)^2} \end{bmatrix} \begin{bmatrix} \theta^2 & 0 \\ 0 & \beta^2/E(x) \end{bmatrix} \begin{bmatrix} \frac{\beta}{(\beta+\theta)^2} \\ -\frac{\theta}{(\beta+\theta)^2} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\beta\theta^2}{(\beta+\theta)^2} & -\frac{\theta\beta^2}{(\beta+\theta)^2 E(x)} \end{bmatrix} \begin{bmatrix} \frac{\beta}{(\beta+\theta)^2} \\ -\frac{\theta}{(\beta+\theta)^2} \end{bmatrix}$$

$$= \frac{\beta\theta^2}{(\beta+\theta)^2}\frac{\beta}{(\beta+\theta)^2} + \frac{\theta\beta^2}{(\beta+\theta)^2 E(x)}\frac{\theta}{(\beta+\theta)^2}$$

$$= \frac{\beta^2\theta^2}{(\beta+\theta)^4} + \frac{\theta^2\beta^2}{(\beta+\theta)^2\beta}\frac{\theta}{(\beta+\theta)^2} \text{ since } E(x) = \frac{\beta}{\theta}$$

$$= \frac{\beta^2\theta^2}{(\beta+\theta)^4} + \frac{\theta^3\beta}{(\beta+\theta)^4}$$

$$= \frac{\beta^2\theta^2 + \theta^3\beta}{(\beta+\theta)^4}$$

Or we can calculate the MLE of this parameter directly from the marginal distribution of x:

$f(x) = \gamma(1-\gamma)^x$ and $\gamma = \theta/(\beta+\theta)$

$$\ln L(x) = \ln(\gamma) + x\ln(1-\gamma)$$
$$\ln L(x_1,\ldots,x_n) = n\ln(\gamma) + \ln(1-\gamma)\sum x_i$$

$$\frac{\partial \ln L(x_1,\ldots,x_n)}{\partial\gamma} = n/\gamma - \frac{\sum x_i}{(1-\gamma)}$$

$$\Rightarrow n/\gamma = \frac{\sum x_i}{(1-\gamma)}$$

$\Rightarrow (1-\gamma)/\gamma = 1/n\sum x_i$
$\Rightarrow 1/\gamma - 1 = 1/n\sum x_i$
$\Rightarrow \widehat{\gamma}_{ml} = \dfrac{1}{1/n\sum x_i + 1}$

$$= \frac{n}{\sum x_i + n} (= \widehat{\theta}_{ml}/(\widehat{\beta}_{ml} + \widehat{\theta}_{ml}))$$

3)

We already found the MLE and asymptotic distribution for this in question 2):
$\widehat{\alpha}_{ml} = \widehat{\gamma}_{ml}$

4)
$f(y|x) = \dfrac{\lambda e^{-\lambda y}(\lambda y)^x}{x!}$
where $\lambda = \beta + \theta$
By the invariance property of MLE, $\widehat{\lambda}_{ml} = \widehat{\beta}_{ml} + \widehat{\theta}_{ml} = \dfrac{\sum x_i + n}{\sum y_i}$

Using the delta method,

$$\sqrt{n}\,(\widehat{\lambda}_{ml} - \lambda) \overset{d}{\to} N(0, \Psi)$$

where $\Psi = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \theta^2 & 0 \\ 0 & \theta\beta \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \theta^2 + \theta\beta$

Or...
You could derive the MLE for λ directly:
$\ln L = \ln\lambda - \lambda y + x\ln(\lambda y) + g(x)$

$$\frac{\partial \ln L}{\partial \lambda} = 1/\lambda - y + \frac{x}{\lambda}$$

MLE for a random sample of n observations:

$$1/\lambda(n + \sum x_i) = \sum y_i$$

$$\Rightarrow \hat{\lambda}_{ml} = \frac{\sum x_i + n}{\sum y_i}$$

5) $f(y) = \theta e^{-\theta y}$ is an exponential distribution with parameter θ.

Since this is the same θ above, we have already calculated its MLE and asymptotic distribution.

$$\hat{\theta}_{ml} = \frac{n}{\sum y_i}$$

$$\sqrt{n}\,(\hat{\theta}_{ml} - \theta) \overset{d}{\to} N(0, \theta^2)$$

6) $f(x|y) = \dfrac{e^{-\beta y}(\beta y)^x}{x!}$ is a poison distribution with parameter βy.

Since this is the same β above, we have already calculated its MLE and

asymptotic distribution.

From above: $\hat{\beta}_{ml} = \dfrac{\sum x_i}{\sum y_i}$

In this case, conditional on y, $\hat{\beta}_{ml} = \dfrac{\sum x_i}{ny} = \dfrac{1/n \sum x_i}{y}$,

which is simply the standard MLE estimator for a possion distribution divided by y. [Recall βy is also the mean of this poisson distribution.]

Asymptotic distribution:

$$\sqrt{n}\,(\hat{\beta}_{ml} - \beta) \overset{d}{\to} N(0, \frac{\beta^2}{E(x|y)})$$

where $\dfrac{\beta^2}{E(x|y)} = \dfrac{\beta^2}{\beta y} = \dfrac{\beta}{y}$

Thus, $\sqrt{n}\,(\widehat{\beta}_{ml} - \beta) \overset{d}{\to} N(0, \frac{\beta}{y})$

Another way to show this...

Again, recall that the asymptotic variance for any poisson distribution is the same as the parameter, which in this case is βy (I showed this during section, but it is easy to prove).

β is a function of this parameter:

$\beta = g(\beta y) = \frac{\beta y}{y}$.

Using the delta method,

$\sqrt{n}\,(\widehat{\beta}_{ml} - \beta) \overset{d}{\to} N(0, (\frac{\partial g(\beta y)}{\partial \beta})^2 \beta y)$

where $\frac{\partial g(\beta y)}{\partial \beta} = 1/y$

and $(\frac{\partial g(\beta y)}{\partial \beta})^2 \beta y = (1/y^2)\beta y = \frac{\beta}{y}$.

Question 3:

$pr(y_i|x_i;\theta) = \Phi(x_i'\gamma)$

1) A problem with non-linear models is that the population parameter vector loses it's simple definition of $\beta = \frac{\partial E(y|X)}{\partial X}$.

Here $\frac{\partial E(y|X)}{\partial X} = \phi(x_i'\gamma)\gamma$, where $\phi(x_i'\gamma)$ is the pdf for the standard normal.

Hence γ is a non-linear function of the data.

2)

$L(\gamma|y_1,\ldots,y_n;x_1,\ldots,x_n) = \prod_{i=1}^{n} \Phi(x_i'\gamma)^{y_i}(1 - \Phi(x_i'\gamma))^{1-y_i}$

$$\ln L = \sum y_i \ln \Phi(x_i'\gamma) + (1-y_i)\ln(1-\Phi(x_i'\gamma))$$

or

$$\ln L = \sum_{y_i=1} \ln \Phi(x_i'\gamma) + \sum_{y_i=0} \ln(1-\Phi(x_i'\gamma))$$

FOC: $\frac{\partial \ln L}{\partial \gamma} = \sum_{y_i=1} \frac{\phi(x_i'\widehat{\gamma}_n)x_i}{\Phi(x_i'\widehat{\gamma}_n)} - \sum_{y_i=0} \frac{\phi(x_i'\widehat{\gamma}_n)x_i}{1-\Phi(x_i'\widehat{\gamma}_n)} = 0$

3) Asymptotic Distribution:

$$\sqrt{n}\,(\widehat{\gamma}_n - \gamma_0) \xrightarrow{d} N(0, \Lambda_0)$$

where $\Lambda_0 = I(\gamma_0)^{-1}$
and $I(\gamma_0) = -E_0(\frac{\partial^2 \ln L(\gamma_0|y,x)}{\partial\gamma_0\partial\gamma_0'}) = E_0(\frac{\partial \ln L(\gamma_0|y,x)}{\partial\gamma_0}\frac{\partial \ln L(\gamma_0|y,x)}{\partial\gamma_0})$,

which is the negative of the expectation of the Hessian matrix of second derivatives of the log likelihood function for a single observation.

4) Estimators for the asymptotic distribution.
The two estimators: $\widehat{\Lambda}_1 = (\widehat{I(\widehat{\gamma}_n)}_1)^{-1}$ and $\widehat{\Lambda}_2 = (\widehat{I(\widehat{\gamma}_n)}_2)^{-1}$

where
$\widehat{I(\widehat{\gamma}_n)}_1 = -1/n \sum \frac{\partial^2 \ln L(\gamma|y_i,x_i)}{\partial\gamma\partial\gamma'}|_{\gamma=\widehat{\gamma}_n}$ (based on the Hessian matrix)
and
$\widehat{I(\widehat{\gamma}_n)}_2 = 1/n \sum \frac{\partial \ln L(\gamma|y_i,x_i)}{\partial\gamma}\frac{\partial \ln L(\gamma|y_i,x_i)}{\partial\gamma'}|_{\gamma=\widehat{\gamma}_n}$ (based on outer-products of gradient).

Show consistency: $\widehat{\Lambda}_1 \xrightarrow{p} \Lambda_0$ and $\widehat{\Lambda}_2 \xrightarrow{p} \Lambda_0$
Assuming these functions are finite, by Law of Large Numbers,
$-1/n \sum \frac{\partial^2 \ln L(\gamma|y_i,x_i)}{\partial\gamma\partial\gamma'}|_{\gamma=\gamma_0} \xrightarrow{p} -E_0(\frac{\partial^2 \ln L(\gamma|y,x)}{\partial\gamma\partial\gamma'})|_{\gamma=\gamma_0}$

and $1/n\sum \frac{\partial \ln L(\gamma|y_i,x_i)}{\partial \gamma}\frac{\partial \ln L(\gamma|y_i,x_i)}{\partial \gamma'}|_{\gamma=\gamma_0} \overset{p}{\to} E_0(\frac{\partial \ln L(\gamma|y,x)}{\partial \gamma}\frac{\partial \ln L(\gamma|y,x)}{\partial \gamma})|_{\gamma=\gamma_0}$

(Notice that I'm evaluating these functions at the true parameter, γ_0. This notation is the same as, e.g.$-E_0(\frac{\partial^2 \ln L(\gamma|y,x)}{\partial\gamma\partial\gamma'})|_{\gamma=\gamma_0} = -E_0(\frac{\partial^2 \ln L(\gamma_0|y,x)}{\partial\gamma\partial\gamma'}))$

Assume: $\hat{\gamma}_n \overset{p}{\to} \gamma_0$

Then by Slutsky,

$$\widehat{I(\hat{\gamma}_n)}_1 = -1/n\sum \frac{\partial^2 \ln L(\gamma|y_i,x_i)}{\partial\gamma\partial\gamma'}|_{\gamma=\hat{\gamma}_n} \overset{p}{\to} -E_0(\frac{\partial^2 \ln L(\gamma_0|y,x)}{\partial\gamma_0\partial\gamma_0'})$$

and

$$\widehat{I(\hat{\gamma}_n)}_2 = 1/n\sum \frac{\partial \ln L(\gamma|y_i,x_i)}{\partial \gamma}\frac{\partial \ln L(\gamma|y_i,x_i)}{\partial \gamma'}|_{\gamma=\hat{\gamma}_n} \overset{p}{\to} E_0(\frac{\partial \ln L(\gamma_0|y,x)}{\partial \gamma_0}\frac{\partial \ln L(\gamma_0|y,x)}{\partial \gamma_0'})$$

Again by Slutsky and assuming $\widehat{I(\hat{\gamma}_n)}_1$ and $\widehat{I(\hat{\gamma}_n)}_2$ are invertible,

$$\hat{\Lambda}_1 = \widehat{I(\hat{\gamma}_n)}_1^{-1} \overset{p}{\to} -E_0(\frac{\partial^2 \ln L(\gamma_0|y,x)}{\partial\gamma_0\partial\gamma_0'}) = I(\gamma_0)^{-1} = \Lambda_0$$

and $\hat{\Lambda}_2 = \widehat{I(\hat{\gamma}_n)}_2^{-1} \overset{p}{\to} E_0(\frac{\partial \ln L(\gamma_0|y,x)}{\partial \gamma_0}\frac{\partial \ln L(\gamma_0|y,x)}{\partial \gamma_0'}) = I(\gamma_0)^{-1} = \Lambda_0$

Thus, $\hat{\Lambda}_1 \overset{p}{\to} \Lambda_0$ and $\hat{\Lambda}_2 \overset{p}{\to} \Lambda_0$.

Question 4:

1) Matlab code is attached.
Results:

$(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4) = (0.8228, 0.9289, -0.8624, 0.3941)$

My stata estimates, which are probably "better":

probit y x1 x2 x3 x4, nocons

y	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
x1	.9217028	.2209755	4.17	0.000	.4885988	1.354807
x2	1.020736	.1093184	9.34	0.000	.8064757	1.234996
x3	-.9670356	.1032077	-9.37	0.000	-1.169319	-.7647521
x4	.4484955	.0767641	5.84	0.000	.2980406	.5989504

2) Covariance matrix

The log likelihood function for a single observation:
$\ln L = y\log(\Phi(x_i'\beta)) + (1-y)\log(1-\Phi(x_i'\beta))$
where $x_i = (x_{1i}, x_{2i}, x_{3i}, x_{4i})'$ and $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)'$
(x_i is kx1 and β is kx1)

$$\frac{\partial \ln L}{\partial \beta} = -y_i \frac{1}{\Phi(x_i'\beta)} \frac{\partial \Phi(x_i'\beta)}{\partial x_i'\beta} x_i + (1-y_i) \frac{1}{1-\Phi(x_i'\beta)} \frac{\partial \Phi(x_i'\beta)}{\partial x_i'\beta} X$$
$$= -y_i \frac{\phi(x_i'\beta)}{\Phi(x_i'\beta)} x_i + (1-y_i) \frac{\phi(x_i'\beta)}{1-\Phi(x_i'\beta)} x_i$$

since $\frac{\partial \Phi(x_i'\beta)}{\partial x_i'\beta} = \phi(x_i'\beta),$ where $\phi(\bullet)$ is the normal pdf
($\frac{\partial \ln L}{\partial \beta}$ is kx1)

Using the outer product representation, the information matrix is

$I(\beta) = (\frac{\partial \ln L}{\partial \beta})(\frac{\partial \ln L}{\partial \beta})'$ (kxk)

A consistent estimator for this information matrix is

$$\widehat{I(\widehat{\beta})} = 1/n \sum_{i=1}^{n} (-y_i \frac{\phi(x_i'\widehat{\beta})}{\Phi(x_i'\widehat{\beta})} x_i + (1-y_i) \frac{\phi(x_i'\widehat{\beta})}{1-\Phi(x_i'\widehat{\beta})} x_i)(-y_i \frac{\phi(x_i'\widehat{\beta})}{\Phi(x_i'\widehat{\beta})} x_i + (1-y_i) \frac{\phi(x_i'\widehat{\beta})}{1-\Phi(x_i'\widehat{\beta})} x_i)'$$

Therefore, a consistent estimator for the asymptotic covariance matrix is $\frac{\widehat{I(\hat{\beta})}^{-1}}{n}$

Results:

$$\frac{\widehat{I(\hat{\beta})}^{-1}}{n} = \begin{bmatrix} 0.0514 & 0.0060 & -0.0167 & -0.0011 \\ 0.0060 & 0.0149 & -0.0125 & 0.0067 \\ -0.0167 & -0.0125 & 0.0156 & -0.0074 \\ -0.0011 & 0.0067 & -0.0074 & 0.0087 \end{bmatrix}$$

Standard errors:
$SE(\beta) = (SE(\beta_1), SE(\beta_2), SE(\beta_3), SE(\beta_4))$
$= (0.2268, 0.1221, 0.1248, 0.0933)$

3) Hypothesis test:
$H_0 : \beta_2 + \beta_3 = 0$
$H_1 : \beta_2 + \beta_3 \neq 0$

One could use any of the three classical tests: likelihood ratio, wald, or lagrange multiplier, and in practice it is often a good idea to try all three. The simplest in this case is probably the wald test.

Test statistic: $W = [c(\hat{\beta}) - 0]' \frac{\widehat{I(c(\hat{\beta}) - 0)}^{-1}}{n} [c(\hat{\beta}) - 0]$ and $W \sim \chi^2(J)$ under H_0
where $c(\hat{\beta}) = \hat{\beta}_2 + \hat{\beta}_3$

Since there is only one restriction, this simplifies to

$W = \frac{(\hat{\beta}_2 + \hat{\beta}_3)^2}{\widehat{V(\hat{\beta}_2)} + \widehat{V(\hat{\beta}_3)} + 2\widehat{cov(\hat{\beta}_2, \hat{\beta}_3)}}$ and $W \sim \chi^2(1)$ under H_0

The denominator is derived by applying the delta method.

From the results above, $\widehat{V(\hat{\beta}_2)} = 0.0149$, $\widehat{V(\hat{\beta}_3)} = 0.0156$,
and $\widehat{cov(\hat{\beta}_2, \hat{\beta}_3)} = -0.0125$

$$W = \frac{(0.9289 - 0.8624)^2}{0.0149 + 0.0156 - 2 * 0.0125} = 0.8005$$

W critical value at 95% is 3.84

$\Rightarrow$Fail to reject null hypothesis.

4) Elasticities

At the sample mean, $\frac{\partial E(y,x,\beta)}{\partial x} = \phi(\overline{x}\beta)\beta$

Estimated: $\varepsilon_{y,x} = \frac{\partial E(y,x,\widehat{\beta})}{\partial x}\frac{\overline{x}}{\overline{y}} = \phi(\overline{x}\widehat{\beta})\widehat{\beta}\frac{\overline{x}}{\overline{y}}$

Results:

$\varepsilon_{y,x_2} = 0.2295, \varepsilon_{y,x_3} = -0.2115, \varepsilon_{y,x_3} = 0.0906$

5) A new parameter:

$\delta = h(\beta) = \frac{\beta_1\beta_2}{\beta_3^2}$

By invariance property of MLE, the MLE for δ, $\widehat{\delta} = \frac{\widehat{\beta}_1\widehat{\beta}_2}{\widehat{\beta}_3^2} = 1.0277.$

Asymptotic distribution:

$$\sqrt{n}\,(\widehat{\delta} - \delta) \overset{d}{\to} N(0, I(\delta)^{-1})$$

where $I(\delta)^{-1} = \Delta I(\beta)^{-1}\Delta'$ and $\Delta = [\frac{\partial h(\beta)}{\partial \beta_1}\ \frac{\partial h(\beta)}{\partial \beta_2}\ \frac{\partial h(\beta)}{\partial \beta_3}\ \frac{\partial h(\beta)}{\partial \beta_4}]$

$\Delta = [\frac{\beta_2}{\beta_3^2}\ \frac{\beta_1}{\beta_3^2}\ -2\frac{\beta_1\beta_2}{\beta_3^3}\ 0]$

A consistent estimator for Δ is $\widehat{\Delta} = [\frac{\widehat{\beta}_2}{\widehat{\beta}_3^2}\ \frac{\widehat{\beta}_1}{\widehat{\beta}_3^2}\ -2\frac{\widehat{\beta}_1\widehat{\beta}_2}{\widehat{\beta}_3^3}\ 0].$

And a consistent estimator for $I(\beta)$ is given above.

Therefore, the estimate of the standard error is

$$SE(\widehat{\delta}) = (\widehat{\Delta}\frac{\widehat{I(\beta)}^{-1}}{n}\widehat{\Delta}')^{1/2} = 0.1961$$