# Chapter 15

# Minimum Distance Estimation and The Generalized Method of Moments

## Exercises

1. The elements of **J** are

$$\frac{\partial\sqrt{b_1}}{\partial m_2} = m_3(-3/2)m_2^{-5/2} \quad \frac{\partial\sqrt{b_1}}{\partial m_3} = m_2^{-3/2} \quad \frac{\partial\sqrt{b_1}}{\partial m_4} = 0$$

$$\frac{\partial b_2}{\partial m_2} = m_4(-2)m_2^{-3} \quad \frac{\partial b_2}{\partial m_3} = 0 \quad \frac{\partial b_2}{\partial m_4} = m_2^{-2}$$

Using the formula given for the moments, we obtain, $\mu_2 = \sigma^2$, $\mu_3 = 0$, $\mu_4 = 3\sigma_4$. Insert these in the derivatives above to obtain

$$\mathbf{J} = \begin{bmatrix} 0 & \sigma^{-3} & 0 \\ -6\sigma^{-2} & 0 & \sigma^{-4} \end{bmatrix}.$$

Since the rows of J are orthogonal, we know that the off diagonal term in **JVJ′** will be zero, which simplifies things a bit. Taking the parts directly, we can see that the asymptotic variance of $\sqrt{b_1}$ will be $\sigma^{-6}$ Asy.Var[$m_3$], which will be

Asy.Var[ $\sqrt{b_1}$ ] = $\sigma^{-6}(\mu_6 - \mu_3^2 + 9\mu_2^3 - 3\mu_2\mu_4 - 3\mu_2\mu_4)$.

The parts needed, using the general result given earlier, are $\mu_6 = 15\sigma^6$, $\mu_3 = 0$, $\mu_2 = \sigma^2$, $\mu_4 = 3\sigma^4$. Inserting these in the parentheses and multiplying it out and collecting terms produces the upper left element of JVJ′ equal to 6, which is the desired result. The lower right element will be

Asy.Var[$b_2$] = $36\sigma^{-4}$ Asy.Var[$m_2$] + $\sigma^{-8}$Asy.Var[$m_4$] - $2(6)\sigma^{-6}$Asy.Cov[$m_2$,$m_4$].

The needed parts are

Asy.Var[$m_2$] = $2\sigma^4$

Asy.Var[$m_4$] = $\mu_8 - \mu_4^2 = 105\sigma^8 - (3\sigma^4)^2$

Asy.Cov[$m_2$,$m_4$] = $\mu_6 - \mu_2\mu_4$ = $15\sigma^6 - \sigma^2(3\sigma^4)$.

Inserting these parts in the expansion, multiplying it out and collecting terms produces the lower right element equal to 24, as expected.

2. The necessary data are given in Examples 15.5. The two moments are $m_1'$=31.278 and $m_2'$.=1453.96. Based on the theoretical results $m_1' = P/\lambda$ and $m_2' = P(P+1)/\lambda^2$, the solutions are $P = \mu_1'^2/(\mu_2' - \mu_1'^2)$ and $\lambda = \mu_1'/(\mu_2' - \mu_1'^2)$. Using the sample moments produces estimates P = 2.05682 and λ = 0.065759. The matrix of derivatives is

$$\mathbf{G} = \begin{bmatrix} \partial\mu_1'/\partial P & \partial\mu_1'/\partial\lambda \\ \partial\mu_2'/\partial P & \partial\mu_2'/\partial\lambda \end{bmatrix} = \begin{bmatrix} 1/\lambda & -P/\lambda^2 \\ (2P+1)/\lambda^2 & -2P(P+1)/\lambda^3 \end{bmatrix} = \begin{bmatrix} 15.207 & -475.648 \\ 1{,}182.54 & -44{,}220.08 \end{bmatrix}.$$

The covariance matrix for the moments is given in Example 18.7;

$$\Phi = \begin{bmatrix} 24.7051 & 2307.126 \\ 2307.126 & 229{,}609.5 \end{bmatrix}$$

3. a. The log likelihood for sampling from the normal distribution is

$$\log L = (-1/2)[n\log 2\pi + n\log\sigma^2 + (1/\sigma^2)\Sigma_i (x_i - \mu)^2]$$

write the summation in the last term as $\Sigma x_i^2 + n\mu^2 - 2\mu\Sigma_i x_i$. Thus, it is clear that the log likelihood is of the form for an exponential family, and the sufficient statistics are the sum and sum of squares of the observations.

b. The log of the density for the Weibull distribution is

$$\log f(x) = \log\alpha + \log\beta + (\beta\text{-}1)\log x_i - \alpha\Sigma_i x_i^{\beta}.$$

The log likelihood is found by summing these functions. The third term does not factor in the fashion needed to produce an exponential family. There are no sufficient statistics for this distribution.

c. The log of the density for the mixture distribution is

$$\log f(x,y) = \log\theta - (\beta+\theta)y_i + x_i\log\beta + x_i\log y_i - \log(x!)$$

This is an exponential family; the sufficient statistics are $\Sigma_i y_i$ and $\Sigma_i x_i$..

4. The question is (deliberately) misleading. We showed in Chapter 8 and in this chapter that in the classical regression model with heteroscedasticity, the OLS estimator is the GMM estimator. The asymptotic covariance matrix of the OLS estimator is given in Section 8.2. The estimator of the asymptotic covariance matrices are $s^2(\mathbf{X'X})^{-1}$ for OLS and the White estimator for GMM.

5. The GMM estimator would be chosen to minimize the criterion

$$q = n\,\mathbf{m'Wm}$$

where $\mathbf{W}$ is the weighting matrix and $\mathbf{m}$ is the empirical moment,

$$\mathbf{m} = (1/n)\Sigma_i (y_i - \Phi(\mathbf{x}_i'\boldsymbol{\beta}))\mathbf{x}_i$$

For the first pass, we'll use $\mathbf{W} = \mathbf{I}$ and just minimize the sumof squares. This provides an initial set of estimates that can be used to compute the optimal weighting matrix. With this first round estimate, we compute

$$\mathbf{W} = [(1/n^2)\,\Sigma_i (y_i - \Phi(\mathbf{x}_i'\boldsymbol{\beta}))^2\,\mathbf{x}_i\,\mathbf{x}_i']^{-1}$$

then return to the optimization problem to find the optimal estimator. The asymptotic covariance matrix is computed from the first order conditions for the optimization. The matrix of derivatives is

$$\mathbf{G} = \partial\mathbf{m}/\partial\boldsymbol{\beta}' = (1/n)\Sigma_i\, -\phi(\mathbf{x}_i'\boldsymbol{\beta})\mathbf{x}_i\mathbf{x}_i'$$

The estimator of the asymptotic covariance matrix will be

$$\mathbf{V} = (1/n)[\mathbf{G'WG}]^{-1}$$

6. This is the comparison between (15-12) and (15-11). The proof can be done by comparing the inverses of the two covariance matrices. Thus, if the claim is correct, the matrix in (15-11) is larger than that in (15-12), or its inverse is smaller. We can ignore the (1/n) as well. We require, then, that

$$\bar{\mathbf{G}}'\boldsymbol{\Phi}^{-1}\bar{\mathbf{G}} > \bar{\mathbf{G}}'\mathbf{W}\bar{\mathbf{G}}[\bar{\mathbf{G}}'\mathbf{W}\boldsymbol{\Phi}\mathbf{W}\bar{\mathbf{G}}]^{-1}\bar{\mathbf{G}}'\mathbf{W}\bar{\mathbf{G}}$$

7. Suppose in a sample of 500 observations from a normal distribution with mean $\mu$ and standard deviation $\sigma$, you are told that 35% of the observations are less than 2.1 and 55% of the observations are less than 3.6. Estimate $\mu$ and $\sigma$.

If 35% of the observations are less than 2.1, we would infer that

$$\Phi[(2.1 - \mu)/\sigma] = .35, \text{ or } (2.1 - \mu)/\sigma = -.385 \Rightarrow 2.1 - \mu = -.385\sigma.$$

Likewise,

$$\Phi[(3.6 - \mu)/\sigma] = .55, \text{ or } (3.6 - \mu)/\sigma = .126 \Rightarrow 3.6 - \mu = .126\sigma.$$

The joint solution is $\hat{\mu} = 3.2301$ and $\hat{\sigma} = 2.9354$. It might not seem obvious, but we can also derive asymptotic standard errors for these estimates by constructing them as method of moments estimators. Observe, first, that the two estimates are based on moment estimators of the probabilities. Let $x_i$ denote one of the 500 observations drawn from the normal distribution. Then, the two proportions are obtained as follows: Let $z_i(2.1) = \mathbf{1}[x_i < 2.1]$ and $z_i(3.6) = \mathbf{1}[x_i < 3.6]$ be indicator functions. Then, the proportion of 35% has been obtained as $\bar{z}(2.1)$ and .55 is $\bar{z}(3.6)$. So, the two proportions are simply the means of functions of the sample observations. Each $z_i$ is a draw from a Bernoulli distribution with success probability $\pi(2.1) = \Phi((2.1\text{-}\mu)/\sigma)$ for $z_i(2.1)$ and $\pi(3.6) = \Phi((3.6\text{-}\mu)/\sigma)$ for $z_i(3.6)$. Therefore, $E[\bar{z}(2.1)] = \pi(2.1)$, and $E[\bar{z}(3.6)] = \pi(3.6)$. The

variances in each case are Var[$\bar{z}$(.)] = 1/$n$[π(.)(1-π(.))]. The covariance of the two sample means is a bit trickier, but we can deduce it from the results of random sampling. Cov[$\bar{z}$(2.1), $\bar{z}$(3.6)]]
= 1/$n$ Cov[$z_i$(2.1),$z_i$(3.6)], and, since in random sampling sample moments will converge to their population counterparts, Cov[$z_i$(2.1),$z_i$(3.6)] = plim [{(1/$n$) $\sum_{i=1}^{n} z_i(2.1)z_i(3.6)$} - π(2.1)π(3.6)]. But, $z_i$(2.1)$z_i$(3.6) must equal $[z_i(2.1)]^2$ which, in turn, equals $z_i$(2.1). It follows, then, that
Cov[$z_i$(2.1),$z_i$(3.6)] = π(2.1)[1 - π(3.6)]. Therefore, the asymptotic covariance matrix for the two sample proportions is $Asy.Var[p(2.1), p(3.6)] = \mathbf{\Sigma} = \frac{1}{n}\begin{bmatrix} \pi(2.1)(1-\pi(2.1)) & \pi(2.1)(1-\pi(3.6)) \\ \pi(2.1)(1-\pi(3.6)) & \pi(3.6)(1-\pi(3.6)) \end{bmatrix}$. If we insert our sample estimates, we obtain $Est.Asy.Var[p(2.1), p(3.6)] = \mathbf{S} = \begin{bmatrix} 0.000455 & 0.000315 \\ 0.000315 & 0.000495 \end{bmatrix}$. Now, ultimately, our estimates of μ and σ are found as functions of $p$(2.1) and $p$(3.6), using the method of moments. The moment equations are

$$m_{2.1} = \left[\frac{1}{n}\sum_{i=1}^{n} z_i(2.1)\right] - \Phi\left[\frac{2.1-\mu}{\sigma}\right] = 0,$$

$$m_{3.6} = \left[\frac{1}{n}\sum_{i=1}^{n} z_i(3.6)\right] - \Phi\left[\frac{3.6-\mu}{\sigma}\right] = 0.$$

Now, let $\mathbf{\Gamma} = \begin{bmatrix} \partial m_{2.1}/\partial\mu & \partial m_{2.1}/\partial\sigma \\ \partial m_{3.6}/\partial\mu & \partial m_{3.61}/\partial\sigma \end{bmatrix}$ and let **G** be the sample estimate of Γ. Then, the estimator of the asymptotic covariance matrix of ($\hat{\mu}$, $\hat{\sigma}$) is $[\mathbf{G}\mathbf{S}^{-1}\mathbf{G}']^{-1}$. The remaining detail is the derivatives, which are just $\partial m_{2.1}/\partial\mu$ = (1/σ)ϕ((2.1-μ)/σ) and $\partial m_{2.1}/\partial\sigma$ = (2.1-μ)/σ[$\partial m_{2.1}/\partial\sigma$] and likewise for $m_{3.6}$. Inserting our sample estimates produces **G** = $\begin{bmatrix} 0.37046 & -0.14259 \\ 0.39579 & 0.04987 \end{bmatrix}$. Finally, multiplying the matrices and computing the necessary inverses produces $[\mathbf{G}\mathbf{S}^{-1}\mathbf{G}']^{-1} = \begin{bmatrix} 0.10178 & -0.12492 \\ -0.12492 & 0.16973 \end{bmatrix}$. The asymptotic distribution would be normal, as usual. Based on these results, a 95% confidence interval for μ would be 3.2301 ± 1.96(.10178)$^2$ = 2.6048 to 3.8554.

# Chapter 16

# Maximum Likelihood Estimation

## Exercises

1. The density of the maximum is

$$n[z/\theta]^{n-1}(1/\theta), \; 0 < z < \theta.$$

Therefore, the expected value is $E[z] = \int_0^\theta z^n dz = [\theta^{n+1}/(n+1)][n/\theta^n] = n\theta/(n+1)$. The variance is found likewise, $E[z^2] = \int_0^\theta z^2 n(z/n)^{n-1}(1/\theta)dz = n\theta^2/(n+2)$ so $\text{Var}[z] = E[z^2] - (E[z])^2 = n\theta^2/[(n+1)^2(n+2)]$. Using mean squared convergence we see that $\lim_{n\to\infty} E[z] = \theta$ and $\lim_{n\to\infty} \text{Var}[z] = 0$, so that plim $z = \theta$.

2. The log-likelihood is $\ln L = -n\ln\theta - (1/\theta)\sum_{i=1}^n x_i$. The maximum likelihood estimator is obtained as the solution to $\partial\ln L/\partial\theta = -n/\theta + (1/\theta^2)\sum_{i=1}^n x_i = 0$, or $\hat{\theta}_{ML} = (1/n)\sum_{i=1}^n x_i = \bar{x}$. The asymptotic variance of the MLE is $\{-E[\partial^2\ln L/\partial\theta^2]\}^{-1} = \{-E[n/\theta^2 - (2/\theta^3)\sum_{i=1}^n x_i]\}^{-1}$. To find the expected value of this random variable, we need $E[x_i] = \theta$. Therefore, the asymptotic variance is $\theta^2/n$. The asymptotic distribution is normal with mean $\theta$ and this variance.

3. The log-likelihood is $\ln L = n\ln\theta - (\beta+\theta)\sum_{i=1}^n y_i + \ln\beta\sum_{i=1}^n x_i + \sum_{i=1}^n x_i \ln y_i - \sum_{i=1}^n \ln(x_i!)$

The first and second derivatives are

$$\partial\ln L/\partial\theta = n/\theta - \sum_{i=1}^n y_i$$
$$\partial\ln L/\partial\beta = -\sum_{i=1}^n y_i + \sum_{i=1}^n x_i/\beta$$
$$\partial^2\ln L/\partial\theta^2 = -n/\theta^2$$
$$\partial^2\ln L/\partial\beta^2 = -\sum_{i=1}^n x_i/\beta^2$$
$$\partial^2\ln L/\partial\beta\partial\theta = 0.$$

Therefore, the maximum likelihood estimators are $\hat{\theta}_{ML} = 1/\bar{y}$ and $\hat{\beta} = \bar{x}/\bar{y}$ and the asymptotic covariance matrix is the inverse of $E\begin{bmatrix} n/\theta^2 & 0 \\ 0 & \sum_{i=1}^n x_i/\beta^2 \end{bmatrix}$. In order to complete the derivation, we will require the expected value of $\sum_{i=1}^n x_i = nE[x_i]$. In order to obtain $E[x_i]$, it is necessary to obtain the marginal distribution of $x_i$, which is f(x) $= \int_0^\infty \theta e^{-(\beta+\theta)y}(\beta y)^x/x!\,dy = \beta^x(\theta/x!)\int_0^\infty e^{-(\beta+\theta)y}y^x dy$. This is $\beta^x(\theta/x!)$ times a gamma integral. This is $f(x) = \beta^x(\theta/x!)[\Gamma(x+1)]/(\beta+\theta)^{x+1}$. But, $\Gamma(x+1) = x!$, so the expression reduces to

$$f(x) = [\theta/(\beta+\theta)][\beta/(\beta+\theta)]^x.$$

Thus, $x$ has a geometric distribution with parameter $\pi = \theta/(\beta+\theta)$. (This is the distribution of the number of tries until the first success of independent trials each with success probability $1-\pi$. Finally, we require the expected value of $x_i$, which is $E[x] = [\theta/(\beta+\theta)]\sum_{x=0}^\infty x[\beta/(\beta+\theta)]^x = \beta/\theta$. Then, the required asymptotic covariance matrix is $\begin{bmatrix} n/\theta^2 & 0 \\ 0 & n(\beta/\theta)/\beta^2 \end{bmatrix}^{-1} = \begin{bmatrix} \theta^2/n & 0 \\ 0 & \beta\theta/n \end{bmatrix}$.

The maximum likelihood estimator of $\theta/(\beta+\theta)$ is is

$$\widehat{\theta/(\beta+\theta)} = (1/\overline{y})/[\overline{x}/\overline{y} + 1/\overline{y}] = 1/(1+\overline{x}).$$

Its asymptotic variance is obtained using the variance of a nonlinear function

$$V = [\beta/(\beta+\theta)]^2(\theta^2/n) + [-\theta/(\beta+\theta)]^2(\beta\theta/n) = \beta\theta^2/[n(\beta+\theta)^3].$$

The asymptotic variance could also be obtained as $[-1/(1+E[x])^2]^2\text{Asy.Var}[\overline{x}]$.)

For part (c), we just note that $\gamma = \theta/(\beta+\theta)$. For a sample of observations on $x$, the log-likelihood would be

$$\ln L = n\ln\gamma + \ln(1-\gamma)\sum_{i=1}^{n} x_i$$

$$\partial\ln L/d\gamma = n/\gamma - \sum_{i=1}^{n} x_i/(1-\gamma).$$

A solution is obtained by first noting that at the solution, $(1-\gamma)/\gamma = \overline{x} = 1/\gamma - 1$. The solution for $\gamma$ is, thus, $\hat{\gamma} = 1/(1+\overline{x})$.Of course, this is what we found in part b., which makes sense.

For part (d) $f(y|x) = \dfrac{f(x,y)}{f(x)} = \dfrac{\theta e^{-(\beta+\theta)y}(\beta y)^x(\beta+\theta)^x(\beta+\theta)}{x!\quad\theta\quad\beta x}$. Cancelling terms and gathering the remaining like terms leaves $f(y|x) = (\beta+\theta)[(\beta+\theta)y]^x e^{-(\beta+\theta)y}/x!$ so the density has the required form with $\lambda = (\beta+\theta)$. The integral is $\left\{[\lambda^{x+1}]/x!\right\}\int_0^\infty e^{-\lambda y}y^x dy$. This integral is a Gamma integral which equals $\Gamma(x+1)/\lambda^{x+1}$, which is the reciprocal of the leading scalar, so the product is 1. The log-likelihood function is

$$\ln L = n\ln\lambda - \lambda\sum_{i=1}^{n} y_i + \ln\lambda\sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\ln x_i!$$

$$\partial\ln L/\partial\lambda = (\sum_{i=1}^{n} x_i + n)/\lambda - \sum_{i=1}^{n} y_i.$$

$$\partial^2\ln L/\partial\lambda^2 = -(\sum_{i=1}^{n} x_i + n)/\lambda^2.$$

Therefore, the maximum likelihood estimator of $\lambda$ is $(1+\overline{x})/\overline{y}$ and the asymptotic variance, conditional on the $x$s is $\text{Asy.Var.}\left[\hat{\lambda}\right] = (\lambda^2/n)/(1+\overline{x})$

Part (e.) We can obtain $f(y)$ by summing over $x$ in the joint density. First, we write the joint density as $f(x,y) = \theta e^{-\theta y}e^{-\beta y}(\beta y)^x/x!$. The sum is, therefore, $f(y) = \theta e^{-\theta y}\sum_{x=0}^{\infty} e^{-\beta y}(\beta y)^x/x!$. The sum is that of the probabilities for a Poisson distribution, so it equals 1. This produces the required result. The maximum likelihood estimator of $\theta$ and its asymptotic variance are derived from

$$\ln L = n\ln\theta - \theta\sum_{i=1}^{n} y_i$$

$$\partial\ln L/\partial\theta = n/\theta - \sum_{i=1}^{n} y_i$$

$$\partial^2\ln L/\partial\theta^2 = -n/\theta^2.$$

Therefore, the maximum likelihood estimator is $1/\overline{y}$ and its asymptotic variance is $\theta^2/n$. Since we found $f(y)$ by factoring $f(x,y)$ into $f(y)f(x|y)$ (apparently, given our result), the answer follows immediately. Just divide the expression used in part e. by $f(y)$. This is a Poisson distribution with parameter $\beta y$. The log-likelihood function and its first derivative are

$$\ln L = -\beta\sum_{i=1}^{n} y_i + \ln\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} x_i\ln y_i - \sum_{i=1}^{n}\ln x_i!$$

$$\partial\ln L/\partial\beta = -\sum_{i=1}^{n} y_i + \sum_{i=1}^{n} x_i/\beta,$$

from which it follows that $\hat{\beta} = \overline{x}/\overline{y}$.

4. The log-likelihood and its two first derivatives are

$$\log L = n\log\alpha + n\log\beta + (\beta-1)\sum_{i=1}^{n}\log x_i - \alpha\sum_{i=1}^{n} x_i^\beta$$

$$\partial\log L/\partial\alpha = n/\alpha - \sum_{i=1}^{n} x_i^\beta$$

$$\partial \log L/\partial\beta \;=\; n/\beta + \sum_{i=1}^{n} \log x_i - \alpha \sum_{i=1}^{n} (\log x_i) x_i^{\beta}$$

Since the first likelihood equation implies that at the maximum, $\hat{\alpha} = n / \sum_{i=1}^{n} x_i^{\beta}$ , one approach would be to scan over the range of β and compute the implied value of α. Two practical complications are the allowable range of β and the starting values to use for the search.

The second derivatives are

$$\partial^2 \ln L/\partial\alpha^2 \;=\; -n/\alpha^2$$

$$\partial^2 \ln L/\partial\beta^2 \;=\; -n/\beta^2 - \alpha \sum_{i=1}^{n} (\log x_i)^2 x_i^{\beta}$$

$$\partial^2 \ln L/\partial\alpha\partial\beta = -\sum_{i=1}^{n} (\log x_i) x_i^{\beta} \;.$$

If we had estimates in hand, the simplest way to estimate the expected values of the Hessian would be to evaluate the expressions above at the maximum likelihood estimates, then compute the negative inverse. First, since the expected value of $\partial \ln L/\partial\alpha$ is zero, it follows that $E[x_i^{\beta}] = 1/\alpha$. Now,

$$E[\partial \ln L/\partial\beta] \;=\; n/\beta + E[\sum_{i=1}^{n} \log x_i\;] - \alpha E[\sum_{i=1}^{n} (\log x_i) x_i^{\beta}\;]= 0$$

as well. Divide by $n$, and use the fact that every term in a sum has the same expectation to obtain

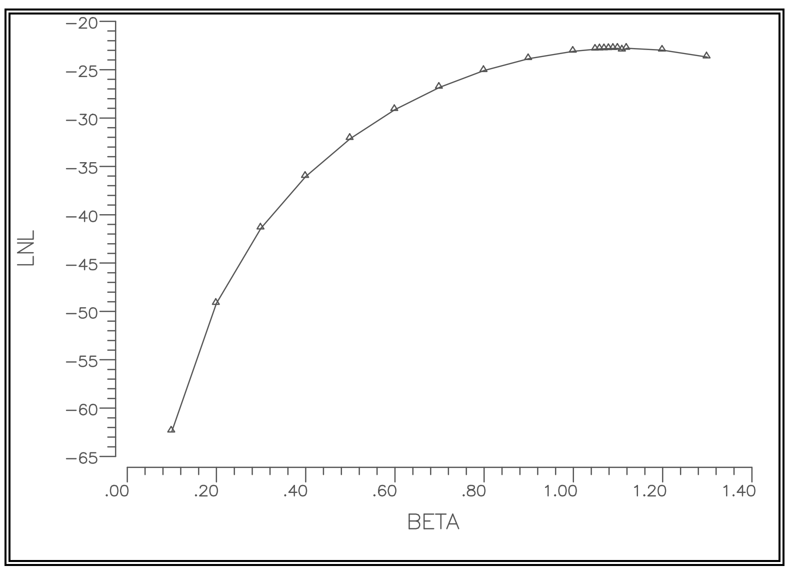
$$1/\beta + E[\ln x_i] - E[(\ln x_i)x_i^{\beta}]/E[x_i^{\beta}] = 0.$$

Now, multiply through by $E[x_i^{\beta}]$ to obtain $E[x_i^{\beta}] \;=\; E[(\ln x_i)x_i^{\beta}] - E[\ln x_i]E[x_i^{\beta}]$

or $\qquad 1/(\alpha\beta) \;=\; \text{Cov}[\ln x_i, x_i^{\beta}]$. ~

5. As suggested in the previous problem, we can concentrate the log-likelihood over α. From $\partial \log L/\partial\alpha = 0$, we find that at the maximum, $\alpha = 1/[(1/n) \sum_{i=1}^{n} x_i^{\beta}\;]$. Thus, we scan over different values of β to seek the value which maximizes log$L$ as given above, where we substitute this expression for each occurrence of α. Values of β and the log-likelihood for a range of values of β are listed and shown in the figure below.

| β | log$L$ |
|---|---|
| 0.1 | -62.386 |
| 0.2 | -49.175 |
| 0.3 | -41.381 |
| 0.4 | -36.051 |
| 0.5 | -32.122 |
| 0.6 | -29.127 |
| 0.7 | -26.829 |
| 0.8 | -25.098 |
| 0.9 | -23.866 |
| 1.0 | -23.101 |
| 1.05 | -22.891 |
| 1.06 | -22.863 |
| 1.07 | -22.841 |
| 1.08 | -22.823 |
| 1.09 | -22.809 |
| 1.10 | -22.800 |
| 1.11 | -22.796 |
| 1.12 | -22.797 |
| 1.2 | -22.984 |
| 1.3 | -23.693 |




The maximum occurs at β = 1.11. The implied value of α is 1.179. The negative of the second derivatives matrix at these values and its inverse are $\mathbf{I}\left(\hat{\alpha},\hat{\beta}\right) = \begin{bmatrix} 25.55 & 9.6506 \\ 9.6506 & 27.7552 \end{bmatrix}$ and $\mathbf{I}^{-1}\left(\hat{\alpha},\hat{\beta}\right) = \begin{bmatrix} .04506 & -.2673 \\ -.2673 & .04148 \end{bmatrix}$.

The Wald statistic for the hypothesis that β = 1 is $W \;=\; (1.11 - 1)^2/.041477 = .276$. The critical value for a test of size .05 is 3.84, so we would not reject the hypothesis.

If $\beta = 1$, then $\hat{\alpha} = n / \sum_{i=1}^{n} x_i = 0.88496$. The distribution specializes to the geometric distribution if $\beta = 1$, so the restricted log-likelihood would be

$$\log L_r = n\log\alpha - \alpha \sum_{i=1}^{n} x_i = n(\log\alpha - 1) \text{ at the MLE.}$$

$\log L_r$ at $\alpha = .88496$ is -22.44435. The likelihood ratio statistic is $-2\log\lambda = 2(23.10068 - 22.44435) = 1.3126$. Once again, this is a small value. To obtain the Lagrange multiplier statistic, we would compute

$$\begin{bmatrix} \partial \log L / \partial\alpha & \partial \log L / \partial\beta \end{bmatrix} \begin{bmatrix} -\partial^2 \log L / \partial\alpha^2 & -\partial^2 \log L / \partial\alpha\partial\beta \\ -\partial^2 \log L / \partial\alpha\partial\beta & -\partial^2 \log L / \partial\beta^2 \end{bmatrix}^{-1} \begin{bmatrix} \partial \log L / \partial\alpha \\ \partial \log L / \partial\beta \end{bmatrix}$$

at the restricted estimates of $\alpha = .88496$ and $\beta = 1$. Making the substitutions from above, at these values, we would have

$$\partial\log L/\partial\alpha = 0$$

$$\partial\log L/\partial\beta = n + \sum_{i=1}^{n} \log x_i - \frac{1}{\bar{x}} \sum_{i=1}^{n} x_i \log x_i = 9.400342$$

$$\partial^2\log L/\partial\alpha^2 = -n\bar{x}^{-2} = -25.54955$$

$$\partial^2\log L/\partial\beta^2 = -n - \frac{1}{\bar{x}} \sum_{i=1}^{n} x_i (\log x_i)^2 = -30.79486$$

$$\partial^2\log L/\partial\alpha\partial\beta = -\sum_{i=1}^{n} x_i \log x_i = -8.265.$$

The lower right element in the inverse matrix is .041477. The LM statistic is, therefore, $(9.40032)^2 .041477 = 2.9095$. This is also well under the critical value for the chi-squared distribution, so the hypothesis is not rejected on the basis of any of the three tests.

6. a. The full log likelihood is $\log L = \Sigma \log f_{yx}(y,x|\alpha,\beta)$.
b. By factoring the density, we obtain the equivalent $\log L = \Sigma[\log f_{y|x}(y|x,\alpha,\beta) + \log f_x(x|\alpha)]$
c. We can solve the first order conditions in each case. From the marginal distribution for x,

$$\Sigma\, \partial \log f_x(x|\alpha)/\partial\alpha = 0$$

provides a solution for $\alpha$. From the joint distribution, factored into the conditional plus the marginal, we have

$$\Sigma[\partial\log f_{y|x}(y|x,\alpha,\beta)/\partial\alpha + \partial\log f_x(x|\alpha)/\partial\alpha = 0$$
$$\Sigma[\partial\log f_{y|x}(y|x,\alpha,\beta)/\partial\beta = 0$$

d. The asymptotic variance obtained from the first estimator would be the negative inverse of the expected second derivative, $\text{Asy.Var}[a] = \{[-E[\Sigma^2\partial \log f_x(x|\alpha)/\partial\alpha^2]\}^{-1}$. Denote this $A_{\alpha\alpha}^{-1}$. Now, consider the second estimator for $\alpha$ and $\beta$ jointly. The negative of the expected Hessian is shown below. Note that the $A_{\alpha\alpha}$ from the marginal distribution appears there, as the marginal distribution appears in the factored joint distribution.

$$-E\frac{\partial^2 \ln L}{\partial\begin{pmatrix}\alpha\\ \beta\end{pmatrix}\begin{pmatrix}\alpha\\ \beta\end{pmatrix}'} = \begin{bmatrix} B_{\alpha\alpha} & B_{\alpha\beta} \\ B_{\beta\alpha} & B_{\beta\beta} \end{bmatrix} + \begin{bmatrix} A_{\alpha\alpha} & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} A_{\alpha\alpha} + B_{\alpha\alpha} & B_{\alpha\beta} \\ B_{\beta\alpha} & B_{\beta\beta} \end{bmatrix}$$

The asymptotic covariance matrix for the joint estimator is the inverse of this matrix. To compare this to the asymptotic variance for the marginal estimator of $\alpha$, we need the upper left element of this matrix. Using the formula for the partitioned inverse, we find that this upper left element in the inverse is

$$[(A_{\alpha\alpha}+B_{\alpha\alpha}) - (B_{\alpha\beta}B_{\beta\beta}^{-1}B_{\beta\alpha})]^{-1} = [A_{\alpha\alpha} + (B_{\alpha\alpha} - B_{\alpha\beta}B_{\beta\beta}^{-1}B_{\beta\alpha})]^{-1}$$

which is smaller than $A_{\alpha\alpha}$ as long as the second term is positive.

e. (Unfortunately, this is an error in the text.) In the preceding expression, $B_{\alpha\beta}$ is the cross derivative. Even if it is zero, the asymptotic variance from the joint estimator is still smaller, being $[A_{\alpha\alpha} + B_{\alpha\alpha}]^{-1}$. This makes sense. If $\alpha$ appears in the conditional distribution, then there is additional information in the factored joint likelihood that is not in the marginal distribution, and this produces the smaller asymptotic variance.

7. The log likelihood for the Poisson model is

$$\text{LogL} = -n\lambda + \log\lambda\Sigma_i y_i - \Sigma_i \log y_i!$$

The expected value of 1/n times this function with respect to the true distribution is

$$E[(1/n)\log L] = -\lambda + \log\lambda\, E_0[\,\bar{y}\,] - E_0\,(1/n)\Sigma_i \log y_i!$$

The first expectation is $\lambda_0$. The second expectation can be left implicit since it will not affect the solution for $\lambda$ - it is a function of the true $\lambda_0$. Maximizing this function with respect to $\lambda$ produces the necessary condition

$$\partial E_0\,(1/n)\log L]/\partial\lambda = -1 + \lambda_0/\lambda = 0$$

which has solution $\lambda = \lambda_0$ which was to be shown.

8. The log likelihood for a sample from the normal distribution is

$$\text{LogL} = -(n/2)\log 2\pi - (n/2)\log\sigma^2 - 1/(2\sigma^2)\,\Sigma_i\,(y_i - \mu)^2.$$

$$E_0\,[(1/n)\log L] = -(1/2)\log 2\pi - (1/2)\log\sigma^2 - 1/(2\sigma^2)\,E_0[(1/n)\,\Sigma_i\,(y_i - \mu)^2].$$

The expectation term equals $E_0[(y_i - \mu)^2] = E_0[(y_i - \mu_0)^2] + (\mu_0 - \mu)^2 = \sigma_0^2 + (\mu_0 - \mu)^2$ . Collecting terms,

$$E_0\,[(1/n)\log L] = -(1/2)\log 2\pi - (1/2)\log\sigma^2 - 1/(2\sigma^2)[\,\sigma_0^2 + (\mu_0 - \mu)^2]$$

To see where this is maximized, note first that the term $(\mu_0 - \mu)^2$ enters negatively as a quadratic, so the maximizing value of $\mu$ is obviously $\mu_0$. Since this term is then zero, we can ignore it, and look for the $\sigma^2$ that maximizes $-(1/2)\log 2\pi - (1/2)\log\sigma^2 - \sigma_0^2/(2\sigma^2)$. The –1/2 is irrelevant as is the leading constant, so we wish to minimize (after changing sign) $\log\sigma^2 + \sigma_0^2/\sigma^2$ with respect to $\sigma^2$. Equating the first derivative to zero produces $1/\sigma^2 = \sigma_0^2/(\sigma^2)^2$ or $\sigma^2 = \sigma_0^2$, which gives us the result.

9. The log likelihood for the classical normal regression model is

$$\text{LogL} = \Sigma_i -(1/2)[\log 2\pi + \log\sigma^2 + (1/\sigma^2)(y_i - x_i'\beta)^2]$$

If we reparameterize this in terms of $\eta = 1/\sigma$ and $\delta = \beta/\sigma$, then after a bit of manipulation,

$$\text{LogL} = \Sigma_i -(1/2)[\log 2\pi - \log\eta^2 + (\eta y_i - x_i'\delta)^2]$$

The first order conditions for maximizing this with respect to $\eta$ and $\delta$ are

$$\partial\log L/\partial\eta = n/\eta - \Sigma_i\, y_i\,(\eta y_i - x_i'\delta) = 0$$

$$\partial\log L/\partial\delta = \Sigma_i\, x_i\,(\eta y_i - x_i'\delta) = 0$$

Solve the second equation for $\delta$, which produces $\delta = \eta\,(X'X)^{-1}X'y = \eta\, b$. Insert this implicit solution into the first equation to produce $n/\eta = \Sigma_i\, y_i\,(\eta y_i - \eta x_i'b)$. By taking $\eta$ outside the summation and multiplying the entire expression by $\eta$, we obtain $n = \eta^2\,\Sigma_i\, y_i\,(y_i - x_i'b)$ or $\eta^2 = n/[\Sigma_i\, y_i\,(y_i - x_i'b)]$. This is an analytic solution for $\eta$ that is only in terms of the data – b is a sample statistic. Inserting the square root of this result into the solution for $\delta$ produces the second result we need. By pursuing this a bit further, you canshow that the solution for $\eta^2$ is just $n/e'e$ from the original least squares regression, and the solution for $\delta$ is just b times this solution for $\eta$. The second derivatives matrix is

$\partial^2\log L/\partial\eta^2 = -n/\eta^2 - \Sigma_i y_i^2$

$\partial^2\log L/\partial\delta\,\partial\delta' = -\Sigma_i\, x_i x_i'$

$\partial^2\log L/\partial\delta\,\partial\eta = \Sigma_i\, x_i y_i$.

We'll obtain the expectations conditioned on X. $E[y_i|x_i]$ is $x_i'\beta$ from the original model, which equals $x_i'\delta/\eta$. $E[y_i^2|x_i] = 1/\eta^2 (\delta' x_i)^2 + 1/\eta^2$. (The cross term has expectation zero.) Summing over observations and collecting terms, we have, conditioned on X,

$E[\partial^2\log L/\partial\eta^2|X] = -2n/\eta^2 - (1/\eta^2)\delta'X'X\delta$

$E[\partial^2\log L/\partial\delta\,\partial\delta'|X] = -X'X$

$E[\partial^2\log L/\partial\delta\,\partial\eta|X] = (1/\eta)X'X\delta$

The negative inverse of the matrix of expected second derivatives is

$$Asy.Var[\mathbf{d}, h] = \begin{bmatrix} \mathbf{X'X} & -(1/\eta)\mathbf{X'X}\boldsymbol{\delta} \\ -(1/\eta)\boldsymbol{\delta}'\mathbf{X'X} & (1/\eta^2)[2n + \boldsymbol{\delta}\mathbf{X'X}\boldsymbol{\delta} \end{bmatrix}^{-1}$$

(The off diagonal term does not vanish here as it does in the original parameterization.)

10**.** The first derivatives of the log likelihood function are $\partial\log L/\partial\mu = -(1/2\sigma^2)\,\Sigma_i -2(\mathbf{y}_i - \boldsymbol{\mu})$. Equating this to zero produces the vector of means for the estimator of $\boldsymbol{\mu}$. The first derivative with respect to $\sigma^2$ is

$\partial\log L/\partial\sigma^2 = -nM/(2\sigma^2) + 1/(2\sigma^4)\Sigma_i\,(\mathbf{y}_i - \boldsymbol{\mu})'(\mathbf{y}_i - \boldsymbol{\mu})$. Each term in the sum is $\Sigma_m\,(y_{im} - \mu_m)^2$. We already deduced that the estimators of $\mu_m$ are the sample means. Inserting these in the solution for $\sigma^2$ and solving the likelihood equation produces the solution given in the problem. The second derivatives of the log likelihood are

$\partial^2\log L/\partial\boldsymbol{\mu}\partial\boldsymbol{\mu}' = (1/\sigma^2)\Sigma_i\, -\mathbf{I}$

$\partial^2\log L/\partial\boldsymbol{\mu}\partial\sigma^2 = (1/2\sigma^4)\,\Sigma_i -2(\mathbf{y}_i - \boldsymbol{\mu})$

$\partial^2\log L/\partial\sigma^2\partial\sigma^2 = nM/(2\sigma^4) - 1/\sigma^6\,\Sigma_i\,(\mathbf{y}_i - \boldsymbol{\mu})'(\mathbf{y}_i - \boldsymbol{\mu})$

The expected value of the first term is $(-n/\sigma^2)\mathbf{I}$. The second term has expectation zero. Each term in the summation in the third term has expectation $M\sigma^2$, so the summation has expected value $nM\sigma^2$. Adding gives the expectation for the third term of $-nM/(2\sigma^4)$. Assembling these in a block diagonal matrix, then taking the negative inverse produces the result given earlier.

For the Wald test, the restriction is

$H_0$: $\boldsymbol{\mu} - \mu^0\mathbf{i} = \mathbf{0}$.

The unrestricted estimator of $\mu$ is $\overline{\mathbf{x}}$. The variance of $\overline{\mathbf{x}}$ is given above, so the Wald statistic is simply $(\overline{\mathbf{x}} - \mu^0\mathbf{i})'\,Var[(\overline{\mathbf{x}} - \mu^0\mathbf{i})]^{-1}(\overline{\mathbf{x}} - \mu^0\mathbf{i})$. Inserting the covariance matrix given above produces the suggested statistic.

11. The asymptotic variance of the MLE is, in fact, equal to the Cramer-Rao Lower Bound for the variance of a consistent, asymptotically normally distributed estimator, so this completes the argument.

In example 4.9, we proposed a regression with a gamma distributed disturbance,

$$y_i = \alpha + \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$$

where,

$$f(\varepsilon_i) = [\lambda^P/\Gamma(P)]\, \varepsilon_i^{P-1} \exp(-\lambda\varepsilon_i),\ \varepsilon_i \geq 0,\ \lambda > 0,\ P > 2.$$

(The fact that $\varepsilon_i$ is nonnegative will shift the constant term, as shown in Example 4.9. The need for the restriction on $P$ will emerge shortly.) It will be convenient to assume the regressors are measured in deviations from their means, so $\Sigma_i\mathbf{x}_i = \mathbf{0}$. The OLS estimator of $\boldsymbol{\beta}$ remains unbiased and consistent in this model, with variance

$$\text{Var}[\mathbf{b}|\mathbf{X}] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

where $\sigma^2 = \text{Var}[\varepsilon_i|\mathbf{X}] = P/\lambda^2$. [You can show this by using gamma integrals to verify that $E[\varepsilon_i|\mathbf{X}] = P/\lambda$ and $E[\varepsilon_i^2|\mathbf{X}] = P(P+1)/\lambda^2$. See B-39 and (E-1) in Section E2.3. A useful device for obtaining the variance is $\Gamma(P) = (P-1)\Gamma(P-1)$.] We will now show that in this model, there is a more efficient consistent estimator of $\boldsymbol{\beta}$. (As we saw in Example 4.9, the constant term in this regression will be biased because $E[\varepsilon_i|\mathbf{X}] = P/\lambda$; $a$ estimates $\alpha+P/\lambda$. In what follows, we will focus on the slope estimators.

The log likelihood function is

$$\text{Ln } L = \sum_{i=1}^{n} P\ln\lambda - \ln\Gamma(P) + (P-1)\ln\varepsilon_i - \lambda\varepsilon_i$$

The likelihood equations are

$$\begin{aligned}
\partial \ln L/\partial\alpha &= \Sigma_i\,[-(P-1)/\varepsilon_i + \lambda] = 0,\\
\partial \ln L/\partial\beta &= \Sigma_i\,[-(P-1)/\varepsilon_i + \lambda]\mathbf{x}_i = \mathbf{0},\\
\partial \ln L/\partial\lambda &= \Sigma_i\,[P/\lambda - \varepsilon_i] = 0,\\
\partial \ln L/\partial P &= \Sigma_i\,[\ln\lambda - \psi(P) - \varepsilon_i] = 0.
\end{aligned}$$

The function $\psi(P) = d\ln\Gamma(P)/dP$ is defined in Section E2.3.) To show that these expressions have expectation zero, we use the gamma integral once again to show that $E[1/\varepsilon_i] = \lambda/(P-1)$. We used the result $E[\ln\varepsilon_i] = \psi(P)-\lambda$ in Example 15.5. So show that $E[\partial\ln L/\partial\beta] = \mathbf{0}$, we only require $E[1/\varepsilon_i] = \lambda/(P-1)$ because $\mathbf{x}_i$ and $\varepsilon_i$ are independent. The second derivatives and their expectations are found as follows: Using the gamma integral once again, we find $E[1/\varepsilon_i^2] = \lambda^2/[(P-1)(P-2)]$. And, recall that $\Sigma_i\mathbf{x}_i = \mathbf{0}$. Thus, conditioned on $\mathbf{X}$, we have

$$\begin{aligned}
-E[\partial^2\ln L/\partial\alpha^2] &= E[\Sigma_i\,(P-1)(1/\varepsilon_i^2)] &&= n\lambda^2/(P-2),\\
-E[\partial^2\ln L/\partial\alpha\partial\boldsymbol{\beta}] &= E[\Sigma_i\,(P-1)(1/\varepsilon_i^2)\mathbf{x}_i] &&= \mathbf{0},\\
-E[\partial^2\ln L/\partial\alpha\partial\lambda] &= E[\Sigma_i\,(-1)] &&= -n,\\
-E[\partial^2\ln L/\partial\alpha\partial P] &= E[\Sigma_i\,(1/\varepsilon_i)] &&= n\lambda/(P-1),\\
-E[\partial^2\ln L/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'] &= E[\Sigma_i\,(P-1)(1/\varepsilon_i^2)\mathbf{x}_i\mathbf{x}_i'] &&= \Sigma_i\,[\lambda^2/(P-2)]\mathbf{x}_i\mathbf{x}_i' = [\lambda^2/(P-2)](\mathbf{X}'\mathbf{X}),\\
-E[\partial^2\ln L/\partial\lambda\partial\boldsymbol{\beta}] &= E[\Sigma_i\,(-1)\mathbf{x}_i] &&= \mathbf{0},\\
-E[\partial^2\ln L/\partial P\partial\boldsymbol{\beta}] &= E[\Sigma_i\,(1/\varepsilon_i)\mathbf{x}_i] &&= \mathbf{0},\\
-E[\partial^2\ln L/\partial\lambda^2] &= E[\Sigma_i\,(P/\lambda^2)] &&= nP/\lambda^2,\\
-E[\partial^2\ln L/\partial\lambda\partial P] &= E[\Sigma_i\,(1/\lambda)] &&= n/\lambda,\\
-E[\partial^2\ln L/\partial P^2] &= E[\Sigma_i\,\psi'(P)] &&= n\psi'(P).
\end{aligned}$$

Since the expectations of the cross partials with respect to $\boldsymbol{\beta}$ and the other parameters are all zero, it follows that the asymptotic covariance matrix for the MLE of $\boldsymbol{\beta}$ is simply

$$\text{Asy.Var}[\hat{\boldsymbol{\beta}}_{MLE}] = \{-E[\partial^2\ln L/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}']\}^{-1} = [(P-2)/\lambda^2](\mathbf{X}'\mathbf{X})^{-1}.$$

Recall, the asymptotic covariance matrix of the ordinary least squares estimator is

$$\text{Asy.Var}[\mathbf{b}] = [P/\lambda^2](\mathbf{X'X})^{-1}.$$

(Note that the MLE is ill defined if $P$ is less than 2.) Thus, the ratio of the variance of the MLE of any element of $\boldsymbol{\beta}$ to that of the corresponding element of **b** is $(P-2)/P$ which is the result claimed in Example 4.9.

**Applications**

1. a. For both probabilities, the symmetry implies that $1 - F(t) = F(-t)$. In either model, then,

$$\text{Prob}(y=1) = F(\text{t}) \text{ and } \text{Prob}(y=0) = 1 - F(\text{t}) = F(-t).$$

These are combined in $\text{Prob}(Y=y) = \text{F}[(2y_i-1)t_i]$ where $t_i = \mathbf{x}_i'\boldsymbol{\beta}$. Therefore,

$$\ln L = \Sigma_i \ln F[(2y_i-1)\mathbf{x}_i'\boldsymbol{\beta}]$$

b.
$$\partial \text{lnL}/\partial\boldsymbol{\beta} = \sum_{i=1}^{n} \frac{f[(2y_i-1)\mathbf{x}_i'\boldsymbol{\beta}]}{F[(2y_i-1)\mathbf{x}_i'\boldsymbol{\beta}]}(2y_i-1)\mathbf{x}_i = \mathbf{0}$$

where $f[(2y_i-1)\mathbf{x}_i'\boldsymbol{\beta}]$ is the density function. For the logit model, $f = F(1-F)$. So, for the logit model,

$$\partial \text{lnL}/\partial\boldsymbol{\beta} = \sum_{i=1}^{n} \{1 - F[(2y_i-1)\mathbf{x}_i'\boldsymbol{\beta}]\}(2y_i-1)\mathbf{x}_i = \mathbf{0}$$

Evaluating this expression for $y_i = 0$, we get simply $-F(\mathbf{x}_i'\boldsymbol{\beta})\mathbf{x}_i$. When $y_i = 1$, the term is $[1- F(\mathbf{x}_i'\boldsymbol{\beta})]\mathbf{x}_i$. It follows that both cases are $[y_i - F(\mathbf{x}_i'\boldsymbol{\beta})]\mathbf{x}_i$, so the likelihood equations for the logit model are

$$\partial \text{lnL}/\partial\boldsymbol{\beta} = \sum_{i=1}^{n} [y_i - \Lambda(\mathbf{x}_i'\boldsymbol{\beta})]\mathbf{x}_i = \mathbf{0}.$$

For the probit model, $F[(2y_i-1)\mathbf{x}_i'\boldsymbol{\beta}] = \Phi[(2y_i-1)\mathbf{x}_i'\boldsymbol{\beta}]$ and $\text{f}[(2y_i-1)\mathbf{x}_i'\boldsymbol{\beta}] = \phi[(2y_i-1)\mathbf{x}_i'\boldsymbol{\beta}]$, which does not simplify further, save for that the term $2y_i$ inside may be dropped since $\phi(\text{t}) = \phi(\text{-t})$. Therefore,

$$\partial \text{lnL}/\partial\boldsymbol{\beta} = \sum_{i=1}^{n} \frac{\phi[(2y_i-1)\mathbf{x}_i'\boldsymbol{\beta}]}{\Phi[(2y_i-1)\mathbf{x}_i'\boldsymbol{\beta}]}(2y_i-1)\mathbf{x}_i = \mathbf{0}$$

c. For the logit model, the result is very simple.

$$\partial^2 \text{lnL}/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}' = \sum_{i=1}^{n} -\Lambda(\mathbf{x}_i'\boldsymbol{\beta})[1-\Lambda(\boldsymbol{\beta})]\mathbf{x}_i\mathbf{x}_i'.$$

For the probit model, the result is more complicated. We will use the result that

$$\text{d}\phi(\text{t})/\text{dt} = -\text{t}\phi(\text{t}).$$

It follows, then, that $\text{d}[\phi(\text{t})/\Phi(\text{t})]/\text{dt} = [-\phi(\text{t})/\Phi(\text{t})][\text{t} + \phi(\text{t})/\Phi(\text{t})]$. Using this result directly, it follows that

$$\partial^2 \text{lnL}/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}' = \sum_{i=1}^{n} -\left(\frac{\phi[(2y_i-1)\mathbf{x}_i'\boldsymbol{\beta}]}{\Phi[(2y_i-1)\mathbf{x}_i'\boldsymbol{\beta}]}\right)\left((2y_i-1)\mathbf{x}_i'\boldsymbol{\beta} + \frac{\phi[(2y_i-1)\mathbf{x}_i'\boldsymbol{\beta}]}{\Phi[(2y_i-1)\mathbf{x}_i'\boldsymbol{\beta}]}\right)(2y_i-1)^2\mathbf{x}_i\mathbf{x}_i' = \mathbf{0}$$

This actually simplifies somewhat because $(2y_i-1)^2 = 1$ for both values of $y_i$ and $\phi[(2y_i-1)\mathbf{x}_i'\boldsymbol{\beta}] = \phi(\mathbf{x}_i'\boldsymbol{\beta})$

d. Denote by **H** the actual second derivatives matrix derived in the previous part. Then, Newton's method is

$$\hat{\beta}(j+1) = \hat{\beta}(j) - \left\{\mathbf{H}\left[\hat{\beta}(j)\right]\right\}^{-1}\left[\frac{\partial \ln L[\hat{\beta}(j)]}{\partial \hat{\beta}(j)}\right]$$

where the terms on the right hand side indicate first and second derivatives evaluated at the "previous" estimate of β.

e. The method of scoring uses the expected Hessian instead of the actual Hessian in the iterations. The methods are the same for the logit model, since the Hessian does not involve $y_i$. The methods are different for the probit model, since the expected Hessian does not equal the actual one. For the logit model

$$-[E(\mathbf{H})]^{-1} = \left\{\sum_{i=1}^{n} \Lambda(\mathbf{x}_i'\beta)[1-\Lambda(\beta)]\mathbf{x}_i\mathbf{x}_i'\right\}^{-1}$$

For the probit model, we need first to obtain the expected value. Do obtain this, we take the expected value, with Prob(y=0) = 1 - Φ and Prob(y=1) = Φ. The expected value of the ith term in the negative hessian is the expected value of the term,

$$\left(\frac{\phi[(2y_i-1)\mathbf{x}_i'\beta]}{\Phi[(2y_i-1)\mathbf{x}_i'\beta]}\right)\left((2y_i-1)\mathbf{x}_i'\beta + \frac{\phi[(2y_i-1)\mathbf{x}_i'\beta]}{\Phi[(2y_i-1)\mathbf{x}_i'\beta]}\right)\mathbf{x}_i\mathbf{x}_i'$$

This is

$$\Phi[-\mathbf{x}_i'\beta]\left(\frac{\phi[\mathbf{x}_i'\beta]}{\Phi[-\mathbf{x}_i'\beta]}\right)\left(-\mathbf{x}_i'\beta + \frac{\phi[\mathbf{x}_i'\beta]}{\Phi[-\mathbf{x}_i'\beta]}\right)\mathbf{x}_i\mathbf{x}_i' + \Phi[\mathbf{x}_i'\beta]\left(\frac{\phi[\mathbf{x}_i'\beta]}{\Phi[\mathbf{x}_i'\beta]}\right)\left(\mathbf{x}_i'\beta + \frac{\phi[\mathbf{x}_i'\beta]}{\Phi[\mathbf{x}_i'\beta]}\right)\mathbf{x}_i\mathbf{x}_i'$$

$$= \phi[\mathbf{x}_i'\beta]\left(-\mathbf{x}_i'\beta + \frac{\phi[\mathbf{x}_i'\beta]}{\Phi[-\mathbf{x}_i'\beta]} + \mathbf{x}_i'\beta + \frac{\phi[\mathbf{x}_i'\beta]}{\Phi[\mathbf{x}_i'\beta]}\right)\mathbf{x}_i\mathbf{x}_i'$$

$$= \phi[\mathbf{x}_i'\beta]\left(\frac{\phi[\mathbf{x}_i'\beta]}{\Phi[-\mathbf{x}_i'\beta]} + \frac{\phi[\mathbf{x}_i'\beta]}{\Phi[\mathbf{x}_i'\beta]}\right)\mathbf{x}_i\mathbf{x}_i'$$

$$= \left(\phi[\mathbf{x}_i'\beta]\right)^2\left(\frac{1}{\Phi[-\mathbf{x}_i'\beta]} + \frac{1}{\Phi[\mathbf{x}_i'\beta]}\right)\mathbf{x}_i\mathbf{x}'$$

$$= \left(\phi[\mathbf{x}_i'\beta]\right)^2\left(\frac{\Phi[\mathbf{x}_i'\beta] + \Phi[-\mathbf{x}_i'\beta]}{\Phi[-\mathbf{x}_i'\beta]\Phi[\mathbf{x}_i'\beta]}\right)\mathbf{x}_i\mathbf{x}'$$

$$= \left(\frac{\left(\phi[\mathbf{x}_i'\beta]\right)^2}{[1-\Phi(\mathbf{x}_i'\beta)]\Phi(\mathbf{x}_i'\beta)}\right)\mathbf{x}_i\mathbf{x}'$$

e.

```
?=====================================================
? Application 16.1
?=====================================================
Namelist ; x = one,age,educ,hsat,female,married $
LOGIT ; Lhs = Doctor ; Rhs = X $
Calc ; L1 = logl $
+----------------------------------------------+
```

```
| Binary Logit Model for Binary Choice         |
| Dependent variable               DOCTOR      |
| Number of observations            27326      |
| Log likelihood function       -16405.94      |
| Number of parameters                  6      |
| Info. Criterion: AIC =          1.20120      |
| Info. Criterion: BIC =          1.20300      |
| Restricted log likelihood     -18019.55      |
+----------------------------------------------+
+--------+-------------+---------------+--------+--------+----------+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|
+--------+-------------+---------------+--------+--------+----------+
---------+Characteristics in numerator of Prob[Y = 1]
 Constant|    1.82207669       .10763712    16.928   .0000
 AGE     |     .01235692       .00124643     9.914   .0000   43.5256898
 EDUC    |    -.00569371       .00578743     -.984   .3252   11.3206310
 HSAT    |    -.29276744       .00686076   -42.673   .0000   6.78542607
 FEMALE  |     .58376753       .02717992    21.478   .0000    .47877479
 MARRIED |     .03550015       .03173886     1.119   .2634    .75861817
```

f.

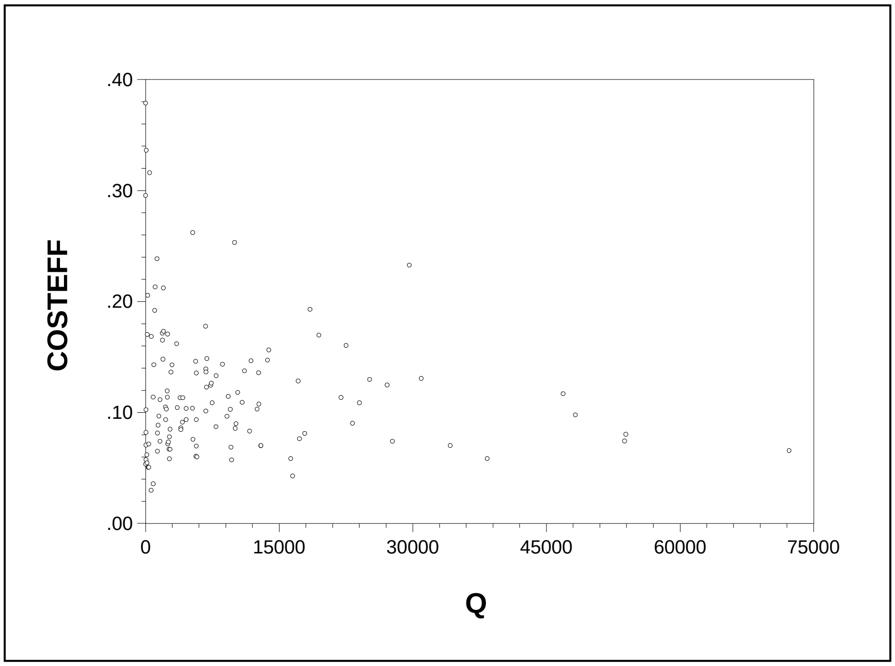
```
Matr ; bw = b(5:6) ; vw = varb(5:6,5:6) $
Matrix ; list ; WaldStat = bw'<vw>bw $
Calc ; list ; ctb(.95,2) $
LOGIT ; Lhs = Doctor ; Rhs = One,age,educ,hsat $
Calc ; L0 = logl $
Calc ; List ; LRStat = 2*(l1-l0) $
Matrix WALDSTAT has  1 rows and  1 columns.
               1
        +--------------
       1|  461.43784
--> Calc ; list ; ctb(.95,2) $
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 Result  =      5.991465
--> Calc ; L0 = logl $
--> Calc ; List ; LRStat = 2*(l1-l0) $
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 LRSTAT  =    467.336374
Logit ; Lhs = Doctor ; Rhs = X ; Start = b,0,0 ; Maxit = 0 $
+----------------------------------------------+
| Binary Logit Model for Binary Choice         |
| Maximum Likelihood Estimates                 |
| Model estimated: May 17, 2007 at 11:49:42PM.|
| Dependent variable               DOCTOR      |
| Weighting variable                 None      |
| Number of observations            27326      |
| Iterations completed                  1      |
| LM Stat. at start values       466.0288      |
| LM statistic kept as scalar    LMSTAT        |
| Log likelihood function       -16639.61      |
| Number of parameters                  6      |
| Info. Criterion: AIC =          1.21830      |
|   Finite Sample: AIC =          1.21830      |
| Info. Criterion: BIC =          1.22010      |
| Info. Criterion:HQIC =          1.21888      |
| Restricted log likelihood     -18019.55      |
| McFadden Pseudo R-squared      .0765802      |
| Chi squared                    2759.883      |
| Degrees of freedom                    5      |
```

```
| Prob[ChiSqd > value] =         .0000000     |
| Hosmer-Lemeshow chi-squared =  23.44388     |
| P-value=  .00284 with deg.fr. =      8      |
+---------------------------------------------+
```

g. The restricted log likelihood given with the initial results equals -18019.55. This is the log likelihood for a model that contains only a constant term. The log likelihood for the model is -16405.94. Twice the difference is about 3,200, which vastly exceeds the critical chi squared with 5 degrees of freedom. The hypothesis would be rejected.

2. We used LIMDEP to fit the cost frontier. The dependent variable is log(Cost/Pfuel). The regressors are a constant, log(Pcapital/Pfuel), log(Plabor/Pfuel), logQ and $\log^2$Q. The Jondrow measure was then computed and plotted against output. There does not appear to be any relationship, though the weak relationship such as it is, is indeed, negative.

```
+---------------------------------------------+
| Limited Dependent Variable Model - FRONTIER |
| Dependent variable                 LCF      |
| Number of observations             123      |
| Log likelihood function       66.86502      |
| Variances: Sigma-squared(v)=       .01185   |
|            Sigma-squared(u)=       .02233   |
|            Sigma(v)        =       .10884   |
|            Sigma(u)        =       .14944   |
| Sigma = Sqr[(s^2(u)+s^2(v)]=       .18488   |
| Stochastic Cost Frontier, e=v+u.            |
+---------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
          Primary Index Equation for Model
 Constant    -7.494211759       .30737742  -24.381   .0000
 LPK       .5531289074E-01  .70211904E-01     .788   .4308     .88666047
 LPL          .2605889758   .67708437E-01    3.849   .0001     5.5808828
 LQ           .4109789313   .29495035E-01   13.934   .0000     8.1794715
 LQ2       .6058235980E-01  .43732083E-02   13.853   .0000     35.112527
          Variance parameters for compound error
 Lambda       1.373117163       .33353523    4.117   .0000
 Sigma        .1848750589   .28257115E-01    6.543   .0000
```



.40
.30
.20
.10
.00
COSTEFF
0
15000
30000
45000
60000
75000
Q

# Chapter 17

# Simulation Based Estimation and Inference

## Exercises

1. Exponential: The pdf is $f(x) = \theta\exp(-\theta x)$. The CDF is

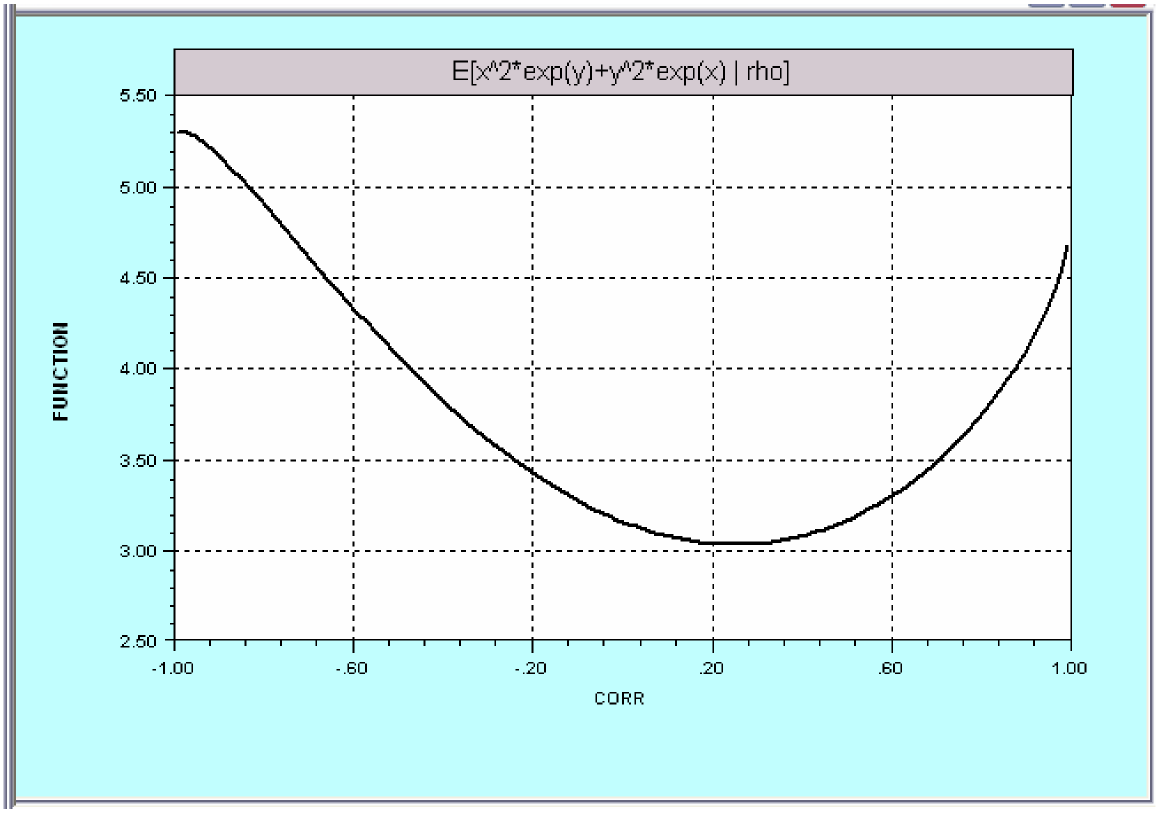
$$F(x) = \int_0^x \theta\exp(-\theta t)dt = \theta\left[-\frac{1}{\theta}\exp(-\theta x) - \left(-\frac{1}{\theta}\exp(-\theta 0)\right)\right] = 1 - \exp(-\theta x).$$

We would draw observations from the U(0,1) population, say $F_i$, and equate these to $F(x_i)$. Inverting the function, we find that $1-F_i = \exp(-\theta x_i)$, or $-(1/\theta)\ln(1-F_i) = x_i$. If $x_i$ has an exponential density, then the density of $y_i = x_i^P$ is
Weibull. If the survival function is $S(x) = \lambda p\exp[-(\lambda x)^p]$, then we may equate random draws from the uniform distribution, $S_i$ to this function (a draw of $S_i$ is the same as a draw of $F_i = 1-S_i$). Solving for $x_i$, we find

$\ln S_i = \ln(\lambda p) - (\lambda x)^p$, so $x_i = (1/\lambda)[\ln(\lambda p) - \ln S_i]^{1/p}$.

2. We will need a bivariate sample on x and y to compute the random variable, then average the draws on it. The precise method of using a Gibbs sampler to draw this bivaraite sample is shown in Example 18.5. Once the bivariate sample of (x,y) is drawn, a large number of observations on $[x^2\exp(y)+y^2\exp(x)]$ is computed and averaged. As noted there, the Gibbs sampler is not much of a simplification for this particular problem. It is simple to draw a sample dircectly from a bivariate normal distribution. Here is a program that does the simulation and plots the estimate of the function

```
Calc   ; Ran(12345) $
Sample ; 1-1000$
Create ; xf=rnn(0,1) ; yfb=rnn(0,1) $
Matrix ; corr=init(100,1,0) ; function=corr $
Calc   ; i=0 $
Proc
Calc   ; i=i+1 $
Matrix ; corr(i)=ro $
Matrix ; c=[1/ro,1] ; c=chol(c) $
Create ; yf = c(2,1)*xf + c(2,2)*yfb $
Create ; fr=xf^2*exp(yf)+yf^2*exp(xf) $
Calc   ; ef = xbr(fr) ; ro=ro+.02 $
Matrix ; function(i)=ef $
Endproc $
Calc   ; ro=-.99 $
Execute; n=100 $
Mplot  ; Lhs = corr ; Rhs = Function ; Fill
       ; Grid ; Endpoints = -1,1
       ; Title=E[x^2*exp(y)+y^2*exp(x) | rho] $
```




# Application

```
?==============================================================
? Application 17.1.  Monte Carlo Simulation
?==============================================================
? Set seed of RNG for replicability
Calc ; Ran(123579) $
? Sample size is 50.  Generate x(i) and z(i) held fixed
Sample ; 1 - 50 $
Create ; xi = rnn(0,1) ; zi = rnn(0,1) $
Namelist ; X = one,xi,zi ; X0 = one,xi $
? Moment Matrices
Matrix ; XXinv = <X'X> ; X0X0inv = <X0'X0> $
Matrix ; Waldi = init(1000,1,0) $
Matrix ; LMi = init(1000,1,0) $

?*************************************************************
? Procedure studies the LM statistic
?*************************************************************
Proc = LM (c) $
? Three kinds of disturbances
Create ?; Eps = Rnt(5) ? Nonnormal distribution
       ; vi=exp(.2*xi) ; eps = vi*rnn(0,1) ? Heteroscedasticity
       ?;eps= Rnn(0,1) ? Standard normal distribution
       ; y = 0 + xi + c*zi +eps $
Matrix ; b0 = X0X0inv*X0'y $
Create ; e0 = y - X0'b0 $
Matrix ; g = X'e0 $
Calc   ; lmstat = qfr(g,xxinv)/(e0'e0/n) ; i = i + 1 $
Matrix ; Lmi (i) = lmstat $
EndProc $
```

```
Calc ; i = 0 ; gamma = -1 $
Exec ; Proc=LM(gamma) ; n = 1000 $
samp;1-1000$
create;LMv=lmi $
create;reject=lmv>3.84$
Calc  ; List ; Type1 = xbr(reject) ; pwr = 1-Type1 $

?****************************************************************
? Procedure studies the Wald statistic
?****************************************************************
Proc = Wald(c) $
Create ; if(type=1)Eps = Rnn(0,1) ? Standard normal distribution
       ; if(type=2)vi=exp(.2*xi)   ? eps = vi*rnn(0,1) ? Heteroscedasticity
       ; if(type=3)eps= Rnt(5)     ? Nonnormal distribution
       ; y = 0 + xi + c*zi +eps $
Matrix ; b0=XXinv*X'y $
Create ; e0=y-X'b0$
Calc   ; ss0 = e0'e0/(47)
       ; v0 = ss0*xxinv(3,3)
       ; wald0=(b0(3))^2/v0
       ; i=i+1 $
Matrix ; Waldi(i)=Wald0 $
EndProc $
? Set the values for the simulation
Calc ; i = 0 ; gamma = 0 ; type=1 $
Sample ; 1-50 $
Exec ; Proc=Wald(gamma) ; n = 1000 $
samp;1-1000$
create;Waldv=Waldi $
create;reject=Waldv > 3.84$
Calc  ; List ; Type1 = xbr(reject) ; pwr = 1-Type1 $
```

To carry out the simulation, execute the procedure for different values of “gamma” and “type.”  Summarize the results with a table or plot of the rejection probabilities as a function of gamma.