

Maximum Likelihood and Examples: Lecture VII

Charles B. Moss

9 September 2014

- 1 Maximum Likelihood
- 2 Estimating with Production Functions and Trends
 - Time Trend
 - Production Function
- 3 Empirical Example
 - Bera-Jarque Test
- 4 Nonlinear Solutions
- 5 Readings

Maximum Likelihood

- An alternative objective approach to estimating the parameters of a distribution function is by maximum likelihood.
 - The argument behind maximum likelihood is to choose those parameters that maximize the likelihood or relative probability of drawing a particular sample.
 - The likelihood function (or the probability of a particular sample) can then be written as

$$L = \prod_{i=1}^N f(x_i | \mu, \sigma^2) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^N (x_i - \mu)^2 \right] \quad (1)$$

- Maximizing this function with respect to the parameters μ and σ^2 implies

$$\max_{\mu, \sigma^2} L = (2\pi\sigma^2)^{-\frac{N}{2}} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^N (x_i - \mu)^2 \right] \quad (2)$$

- Taking the first-order conditions with respect to μ first

$$\begin{aligned} \frac{\partial L}{\partial \mu} &= (2\pi\sigma^2)^{-\frac{N}{2}} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^N (x_i - \mu)^2 \right] \\ &\quad \times \left(-\frac{1}{2\sigma^2} \sum_{i=1}^N [-2(x_i - \mu)] \right) = 0 \quad (3) \\ \Rightarrow \frac{1}{\sigma^2} \left(\sum_{i=1}^N x_i - N\mu \right) &= 0 \Rightarrow \hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i \end{aligned}$$

- In order to solve for the first-order conditions with respect to the variance, we treat σ^2 as a single variable

$$\begin{aligned}
 \frac{\partial L}{\partial \sigma^2} &= -\frac{N}{2} (2\pi)^{-\frac{N}{2}} (\sigma^2)^{-\frac{N}{2}-1} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^N (x_i - \mu)^2 \right] + \\
 & (2\pi\sigma^2)^{-\frac{N}{2}} \left(\frac{(\sigma^2)^{-2}}{2} \sum_{i=1}^N (x_i - \mu)^2 \right) \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^N (x_i - \mu)^2 \right] = 0 \\
 &= -\frac{N (\sigma^2)^{-1}}{2} + \frac{(\sigma^2)^{-2}}{2} \sum_{i=1}^N (x_i - \mu)^2 = 0 \\
 &= -N\sigma^2 \sum_{i=1}^N (x_i - \mu)^2 = 0 \Rightarrow \hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2
 \end{aligned}$$

- The derivation of the maximum likelihood estimates can be simplified by maximizing the logarithm of the likelihood function.

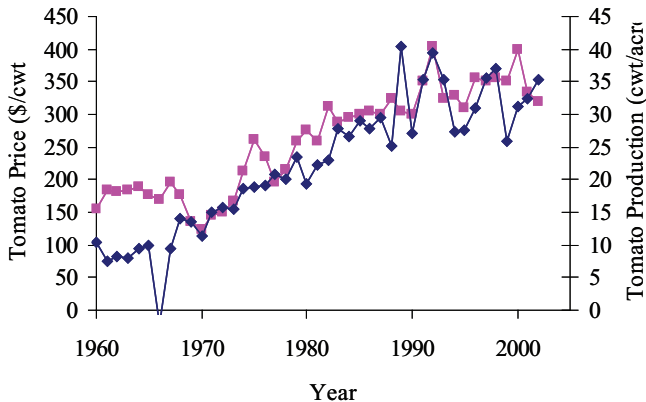
$$\begin{aligned}
 \ln(L) &= -\frac{N}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^N (x_i - \mu)^2 \\
 \frac{\partial \ln(L)}{\partial \mu} &= -\frac{1}{2\sigma^2} \sum_{i=1}^N [-2(x_i - \mu)] = 0 \\
 \Rightarrow \sum_{i=1}^N x_i - N\mu &= 0 \Rightarrow \hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i \\
 \frac{\partial \ln(L)}{\partial \sigma^2} &= -\frac{N}{2} \frac{1}{\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^N (x_i - \mu)^2 = 0 \\
 \Rightarrow -N\sigma^2 + \sum_{i=1}^N (x_i - \mu)^2 &= 0 \Rightarrow \hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2
 \end{aligned} \tag{5}$$

- The method of moments estimator and maximum likelihood estimator of the parameters of the normal distribution are the same.
 - The primary difference is the ability to make statistical inferences under the maximum likelihood estimator.
 - Specifically, the variance of the estimated parameters obtain the Cramer-Rao lower bound

$$V(\hat{\theta}) \geq -\frac{1}{E\left[\frac{\partial^2 \ln(L)}{\partial^2 \theta}\right]} \quad (6)$$

Estimating Probability Functions with Production Functions and Trends

- An additional complication in our estimation of distribution functions for use in decision making under risk and uncertainty is the effect of deterministic factors such as time trends and production functions which pervade economic analysis.



—■— Tomato Production (cwt/acre) —◆— Tomato Price (\$/cwt)

- Both series appear to have a significant upward trend.
 - To highlight the problem, we first compute the parameters of the normal distribution without a trend line.
 - The result is a mean price of \$22.28/cwt and a variance of 103.78, and a mean yield of 259.51 cwt/acre with a variance of 6187.45.
 - Next, we remove the trend for each data series by the result that

$$\hat{\epsilon} = [I - x(x'x)^{-1}x']y \quad (7)$$

- The standard result from regression analysis is that

$$\hat{\beta} = (x'x)^{-1}(x'y) \quad (8)$$

- Substituting this result into the regression equation

$$\begin{aligned}\hat{\epsilon} &= y - x\beta = y - x \left((x'x)^{-1} (x'y) \right) \\ &= y - x (x'x)^{-1} x'y = [I - x (x'x)^{-1} x'] y\end{aligned}\tag{9}$$

- The mean of each series is then zero under the assumptions of ordinary least squares, but in each case the variance is reduced.
 - In the case of tomato prices the variance falls to 18.13 while the variance in tomato yields falls to 1092.07.
 - The reduction in variance is related to the explanatory power of the model.

$$R_P^2 + \frac{18.13}{103.78} = 1 \quad (10)$$

$$R_Y^2 + \frac{1092.07}{6187.45} = 1$$

Estimating Production Functions

- Another example can be seen in the estimation of production functions.

$$y = x_1^\alpha x_2^\beta x_3^\gamma \quad (11)$$

- To estimate the parameters in the Cobb-Douglas production function most researchers take the logarithm of both sides of the equation and append a constant and an error term yielding

$$\ln(y) = \alpha_0 + \alpha \ln(x_1) + \beta \ln(x_2) + \gamma \ln(x_3) + \epsilon \quad (12)$$

- This procedure leads to several additional considerations in modeling the effect of risk on the agricultural firm.
 - By the Gauss-Markov theorem given the typical conditions on error term the parameter estimates are best linear unbiased.
 - Note that the distributional assumptions required by Gauss-Markov only require that the individual errors be independently and identically distributed (or homoscedasticity); hence, the efficiency of the estimates is not dependent on the shape of the distribution function itself (as long as the distribution function does not violate homoscedasticity).
 - The distribution of the errors is not the distribution of risk in production space.
- Taking the exponential of each side yields

$$y = x_1^\alpha x_2^\beta x_3^\gamma e^{\alpha_0 + \epsilon} \ni \epsilon \sim N(0, \sigma_\epsilon^2) \quad (13)$$

- One difference is that while the normal distribution (or the transformed distribution) is symmetric, the log-normal distribution is positively skewed.

$$\mu_3^*(Y) = \exp \left[3\mu + \frac{3\sigma^2}{2} \right] (\exp[\sigma^2] - 1)^2 (\exp[\sigma^2] + 2) \quad (14)$$
$$\mu_3^*(X) = 0$$

Empirical Example

- We analyze the price and yield data for Florida tomatoes, but start by ignoring the possibility of a trend.
 - Computing the first empirical moment of the distribution along with the second, third, and fourth central moments for the random variables price (P) and yield (Y) are given by

$$\begin{aligned}\mu_1(P) &= 22.2800, \mu_2^*(P) = 103.7818, \\ \mu_3^*(P) &= -185.3151, \mu_4^*(P) = 53,367,955.4 \\ \mu_1(Y) &= 295.5116, \mu_2^*(Y) = 6,187.446, \\ \mu_3^*(Y) &= -35,174.9, \mu_4^*(Y) = 66,173,832.21\end{aligned}\tag{15}$$

- The covariance between the yield and price is 701.3036 implying a correlation coefficient of 0.8752.
- These results suggest that both distributions may be negatively skewed and kurtotic.
- To test for these possibilities, we apply the Bera-Jarque tests.

Bera-Jarque Test

- First, we normalize the third moment as

$$\alpha_3 = \frac{\mu_3^*(X)}{(\mu_2^*(X))^{\frac{3}{2}}} \Rightarrow \sqrt{T} \hat{\alpha}_3 \sim N(0, 6) \quad (16)$$

- Second, we normalize the fourth moment of each distribution as

$$\alpha_4 = \frac{\mu_4^*(X)}{(\mu_2^*(X))^2} \Rightarrow \sqrt{T} (\hat{\alpha}_4 - 3) \sim N(0, 24) \quad (17)$$

- The joint test for skewness and kurtosis can then be written as

$$\frac{T}{6} (\hat{\alpha}_3)^2 + \frac{T}{24} (\hat{\alpha}_4 - 3)^2 \sim \chi^2_2 \quad (18)$$

- The result is normalized skewness coefficients of -1.1359 for prices and -0.4684 for yields. *We fail to reject the hypothesis that the distributions are not skewed.*
- The normalized kurtosis coefficients are 38,128.7107 for prices and 11.2018 for yields. *Both empirical distributions appear to be more highly kurtotic than the normal (which has a kurtosis of 3) or leptokurtotic.*

Nonlinear Solutions

- One alternative to the normal is the gamma distribution.
 - The gamma distribution has a closed form solution for the method of moments approach.
 - However, many of the distributions that we use do not have a closed form solution and, hence, must be solved using numerical solution algorithms.
 - We will solve for the parameters of the normal distribution using the Gauss-Sidel approach.

- Gauss-Sidel is an algorithm that is used to solve for the zeros of a multivariate nonlinear function.
 - The multivariate function is a vector function for the difference between the theoretical and empirical moments of a distribution

$$f(\alpha, \beta) = \begin{bmatrix} \alpha\beta - 22.2800 \\ \alpha\beta^2 - 103.7818 \end{bmatrix} \quad (19)$$

- We start with a first-order Taylor series expansion

$$f(\alpha, \beta) \approx f(\alpha_0, \beta_0) + \nabla_{\alpha, \beta} f(\alpha_0, \beta_0) \begin{bmatrix} \alpha - \alpha_0 \\ \beta - \beta_0 \end{bmatrix} \quad (20)$$

- Letting the value of the function go to zero and solving the system for the parameters yields

$$\begin{aligned}
 -\nabla_{\alpha,\beta}^{-1} f(\alpha_0, \beta_0) f(\alpha_0, \beta_0) &= \begin{bmatrix} \alpha - \alpha_0 \\ \beta - \beta_0 \end{bmatrix} \\
 \Rightarrow \begin{bmatrix} \alpha \\ \beta \end{bmatrix} &= \begin{bmatrix} \alpha_0 \\ \beta_0 \end{bmatrix} - \nabla_{\alpha,\beta}^{-1} f(\alpha_0, \beta_0) f(\alpha_0, \beta_0)
 \end{aligned} \tag{21}$$

- For the gamma distribution in this case

$$\begin{aligned}
 \begin{bmatrix} \alpha \\ \beta \end{bmatrix} &= \begin{bmatrix} \alpha_0 \\ \beta_0 \end{bmatrix} - \begin{bmatrix} \beta_0 & \alpha_0 \\ \beta_0^2 & 2\alpha_0\beta_0 \end{bmatrix}^{-1} \begin{bmatrix} \alpha_0\beta_0 - 22.2800 \\ \alpha_0\beta_0^2 - 103.7818 \end{bmatrix} \\
 &= \begin{bmatrix} \alpha_0 \\ \beta_0 \end{bmatrix} - \begin{bmatrix} \frac{2}{\beta_0} & -\frac{1}{\beta_0^2} \\ -\frac{1}{\alpha_0} & \frac{1}{\alpha_0\beta_0} \end{bmatrix} \begin{bmatrix} \alpha_0\beta_0 - 22.2800 \\ \alpha_0\beta_0^2 - 103.7818 \end{bmatrix}
 \end{aligned}$$

- If we start with an initial guess of $\{4,4\}$ the new values of the parameters are $\{4.65364, 4.916366\}$. The system converges to a solution of $\{4.7831, 4.65807\}$ in four iterations.

Readings

- Moss (2010) *Risk, Uncertainty and the Agricultural Firm* Sections 2.6.
- Moss (2014) *Mathematical Statistics for Applied Econometrics* Chapter 7.