

Applying Ordinary Least Squares on the Computer - R: Lecture X

Charles B. Moss ¹

¹University of Florida

January 28, 2012

1 Readings

2 Applied Research Problem

- Econometric Specification

3 Writing R - Code

- Load Subroutines and Libraries
- Data Description
- Load the Data
- Actual Regression
- Post Estimation - Standard Errors
- Post Estimation - R-squared and F
- Run Regression using Canned function

4 Maximum Likelihood

- Compute the Variance

Readings

- * Farnsworth, G.V. 2008. *Econometrics in R*. (<http://cran.r-project.org/doc/contrib/Farnsworth-EconometricsInR.pdf>).
- * Kleiber, C. and A. Zeileis. 2008. *Applied Econometrics with R (Use R!)*. Springer.
- * Laitinen K. and H. Theil. 1978. Supply and Demand of the Multiproduct Firm. *European Economic Review*.
- * Moss, C.B., G. Livanis, and A. Schmitz. 2010. The Effect of Increased Energy Prices on Agriculture: A Differential Approach. *Journal of Agricultural and Applied Economics* 42(4), 711-718.

Applied Research Problem

- The applied research problem which will form the basis of our numerical analysis is taken from Moss, Livanis, and Schmitz (2010).
- This article is available online at <http://ageconsearch.umn.edu/bitstream/100514/2/jaae125.pdf>.
- In our example, we will build the same formulation for the Food and Kindred Products portion of the KLEM data.
- The increase in energy prices since 2004 effected both agricultural production and the manufacture of food and kindred products from agricultural products.

Econometric Specification

- The appendix of Moss, Livanis, and Schmitz presents the derivation of a system of derived demand equations using the approach pioneered by Laitinen and Theil (1978).
- I want to build toward a multiproduct estimation of this formulation, so we will start with a regression that explains the logarithmic change in output level

$$\begin{aligned} d \ln(y_t) = & \alpha_0 + \alpha_1 d \ln(p_t) + \alpha_2 d \ln(w_{1t}) + \alpha_3 d \ln(w_{2t}) \\ & + \alpha_4 d \ln(w_{3t}) + \alpha_5 d \ln(w_{4t}) + \epsilon_t \end{aligned} \quad (1)$$

- The second set of equations can be estimated in a system, but we will start one equation at a time

Writing R - Code

- It is good technique to start code by describing what the code is

AEB 6572 - Econometric Methods II

Charles B. Moss - cbmoss@ufl.edu

R Regression Program Prog02-01-12.R

Load Subroutines and Libraries

- Looking at the first executable line of the program, consider loading any libraries you may need.
- In our program we need to load nlme - Nonlinear Mixed - Effects Models Package so that we can use the function fdHess().

```
library(nlme)
```

Data Description

- Next, describe the data used

Data Set Jorgensen's KLEM dta for Food and Kindred Products

dta[,1] - Year

dta[,2] - log change in output

dta[,3] - log change in quantity of capital

dta[,4] - log change in quantity of labor

dta[,5] - log change in quantity of energy

dta[,6] - log change in quantity of materials

dta[,7] - log change in output price

Load the Data

- Next load the data

```
dta <- read.csv("KLEM-Food01-25-12.csv")
```

Actual Regression

- `y <- as.matrix(dta[,2])`
- `x <-`
`cbind(matrix(1,nrow=nrow(y),ncol=1),as.matrix(dta[,7:11]))`
- `b <- solve(t(x)%*%x)%*%(t(x)%*%y)`

- Compute errors: $\text{err} <- y - x \%*\% b$
- Compute SSE: $\text{sse} <- \text{t}(\text{err}) \%*\% \text{err}$
- Compute MSE: $\text{mse} <- \text{sse} / (\text{nrow}(x) - \text{ncol}(x))$
- Variance of the beta: $\text{vb} <- \text{mse}[1,1] * \text{solve}(\text{t}(x) \%*\% x)$

Post Estimation - R-squared and F

- $qt <- \text{diag}(1, \text{nrow}(x), \text{nrow}(x)) - (\text{matrix}(1, \text{nrow}(x), 1) \%*\% \text{matrix}(1, 1, \text{nrow}(x))) / \text{nrow}(x)$
- Compute Sum Squared Regressor: $ssr <- \text{t}(b) \%*\% \text{t}(x) \%*\% qt \%*\% y$
- Compute Sum Squared Total: $sst <- \text{t}(y) \%*\% qt \%*\% y$
- Compute R-Squared: $r2 <- ssr / sst$
- Compute F: $ff <- (ssr / (\text{ncol}(x) - 1)) / (sse / (\text{nrow}(x) - \text{ncol}(x)))$

Run Regression using Canned function

- `b1 <- lm(y ~ x[,2] + x[,3] + x[,4] + x[,5] + x[,6] )`
- `print(summary.lm(b1))`
- `print(sqrt(diag(vb)))`

Maximum Likelihood

- We start the maximum likelihood by defining the likelihood function.

```
ml <- function(b) {  
  sighat <- t(y-b[1]-b[2]*x[,2]-b[3]*x[,3]-b[4]*x[,4]-b[5]*x[,5]-b[6]*x[,6])%*%  
    (y-b[1]-b[2]*x[,2]-b[3]*x[,3]-b[4]*x[,4]-b[5]*x[,5]-b[6]*x[,6])/nrow(y)  
  nrow(y)/2*log(sighat) }  
}
```

- Next we call the maximization routine using this function.

```
res <- optim(matrix(0,6,1),ml)
```

Compute the Variance

- Again, relying on the Characteristic Function, the variance of the maximum likelihood estimates is defined by the Cramer-Rao lower bound.

$$V_{MLE} = \left[-\nabla_{\beta\beta'}^2 L(\beta) \right]^{-1} = \left(- \begin{bmatrix} \frac{\partial^2 L}{\partial \beta_1 \partial \beta_1} & \cdots & \frac{\partial^2 L}{\partial \beta_1 \partial \beta_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 L}{\partial \beta_k \partial \beta_1} & \cdots & \frac{\partial^2 L}{\partial \beta_k \partial \beta_k} \end{bmatrix} \right)^{-1} \quad (3)$$

- This is implemented in R by

```
resh <- fdHess(res$par, ml)
vest <- solve(resh$Hessian)
```