Assignment #3
Agricultural and Applied Economics 637
Estimation of Econometric Models Using NLS Methods

Due: March 6, 2018
(***Total Points: 150 pts***)

For this assignment, I would like you to estimate some non-linear (in parameters) regression models and to undertake some post-estimation analyses.

As noted in Madsen, Nielson and Tingleff (2014):

> *All methods for non-linear optimization are iterative: From a starting point x_0 the method produces a series of vectors $x_1, x_2, \ldots$, which (hopefully) converges to x^*, a local minimizer for the given function…Most methods have measures which enforce the descending condition $F(x_{k+1}) < F(x_k)$…This prevents convergence to a maximizer and also makes it less probable that we converge towards a saddle point. If the given function has several minimizers, the result will depend on the starting point x_0. We do not know which of the minimizers that will be found; it is not necessarily the minimizer closest to x_0.*

In class we reviewed both the Gauss-Newton (**GN**) and Newton-Raphson (**NR**) nonlinear least squares parameter estimation algorithms which can be represented by the following general equation:

$$\Theta_{n+1} = \Theta_n - \gamma_n P_n \left.\frac{dS(\Theta)}{d\Theta}\right|_{\Theta=\Theta_n}$$

where P_n varies across algorithm.[1] That is, n+1=

$$P_n = \begin{cases} \left(\frac{1}{2}\left(Z(\Theta_n)' Z(\Theta_n)\right)\right)^{-1} & \text{for Gauss-Newton} \\ \left(\left.\frac{d^2S(\Theta)}{d\Theta^2}\right|_{\Theta=\Theta_n}\right)^{-1} & \text{for Newton-Raphson} \end{cases} \quad \text{where } Z(\Theta_n) = \left.\frac{\partial f(X,\Theta)}{\partial\Theta}\right|_{\Theta=\Theta_n}$$

---

[1]The NLS.m file we reviewed in class and what you have access to, uses the alternative step length method to obtain GN based parameters. That is: $\Theta_{n+1} = \Theta_n + (Z'Z)^{-1}Z'\varepsilon$. Either GN step method is correct. The easiest method would to use the current GN code and change to incorporate the NR method.

$S(\Theta_n)=SSE(\Theta)|_{\Theta=\Theta n}$ and γ_n is the iteration step length adjuster defined to ensure that the SSE does not increase across iterations. Take a look at Madsen, Nielson and Tingleff(2004) for a good review as well as Appendix E in Greene, p. 1097- 1099.

1. (***25 pts***) Most NLS estimation software provide you with a choice of methods for obtaining parameter estimates. That is, the user specifies the desired estimation method via the command file that controls estimation. So far, in class we have reviewed two methods for obtaining NLS parameter estimates, GN and NR.

   You may have noticed at this point in the class we have not used the NR method in a ***generic*** NLS program. We also have not checked at the optimal solution whether the SSE function is convex or concave. To make your life easy, I would like you to modify the generic GN algorithm (i.e., nls.m) we used in class to overcome these shortcomings.

   First: Develop a software system where, from the same command file, you can choose which algorithm to use to estimate the parameters via a nonlinear (in parameters) regression model. You will design your estimation function such that both algorithms will be contained in the same estimation function. That is, you make a choice via setting a parameter in your command file that tells the code which algorithm to use (e.g., alg = ***NR*** or ***GN***).[2] The file Program_FlowChart_V2.pptx provides one take on the structure of a two-algorithm system.

   Note, the estimation function should be able to accommodate any functional form, any number of parameters and, as noted above, contains both estimation algorithms. Use the NLS ***GN*** code distributed in class designed to handle any functional form and model size as a starting point to develop your combined MATLAB estimation function.

[2] Use your own method for allowing the user to select which estimation algorithm to use. For example you could use 0 for the GN algorithm and 1 for the NR.

The one difficulty with putting the GN and NR estimation systems into one package is that we need to extend the NR code we developed in class specifically to estimate the 1 variable/2 parameter nonlinear regression model to a more general application that can handle M exogenous variables and K parameters.

As noted above, we can express the NR algorithm via the following iteration method: $\beta_{n+1} = \beta_n - H(\beta_n)^{-1} \left. \frac{dS(\beta)}{d\beta} \right|_{\beta=\beta_n}$ where H(β) is the (K × K) Hessian matrix of the SSE function evaluated at the current parameter estimates, i.e., $H(\beta_n) \equiv \left. \frac{\partial^2 SSE}{\partial \beta} \right|_{\beta=\beta_n}$. Remember that SSE is a scaler and a function of K parameters and that β_n and $\left. \frac{dS(\beta)}{d\beta} \right|_{\beta=\beta_n}$ are (K × 1) vectors. Similar to what we have done in terms of numerical evaluation of gradients of any nonlinear function, the function contained in the MATLAB file *hessian_bwg.m* is used to evaluate the ***numerical*** Hessian of any function.

In addition to the above, in both the GN and NR algorithms make sure you make available a *variable step length method* for estimation.

Second, add a feature to the program where, at the ***optimal parameter values*** obtained via either method, you check for the convexity of the SSE function. If it passes the test: (i) print a statement to your output file to that effect; and (ii) print the usual regression output. If it does not pass the test then have the software (i) print a message to that effect and (ii) do not print your regression results. You can use this combined code for the remainder of the assignment.

Your program output should include a table that shows:

i. The minimum SSE values under both estimators;
ii. The SSE minimizing coefficient vector;
iii. Associated parameter standard errors;

iv. The p-value of the coefficients based on the normal distribution instead of the t-statistic.
v. The number of iterations needed to move from the starting values to final parameter estimates;
vi. The step length associated with the last 5 iterations; and
vii. The results of your SSE function convexity check.

a. (***20 pts***) A problem faced by applied economists wanting to examine the relationship between a dependent variable and a set of explanatory variables is determination of an appropriate functional form. Sometimes economic theory provides some guidance.[3] In contrast, the Box-Cox transformation is one method for letting the data tell you something about appropriate functional form (Greene, p. 296-297). With the general relationship, $y = f(x_1, x_2, \ldots x_K)$, the Box-Cox transformation of the exogenous variables (x) can be represented via the following: $x(\lambda) = \frac{x^\lambda - 1}{\lambda}$ where x is strictly positive, $\lambda \neq 0$, k=(1,…,K), and λ is a parameter to estimate along with the traditional regression coefficients. As noted in Greene (2008) p. 297, when $\lambda = 0$ via L'Hopital's Rule we have: $\lim_{\lambda \to 0} \frac{x^\lambda - 1}{\lambda} = \ln x$. The functional form *determined by the data* is the form defined by the estimated transformation parameter, λ.

Consider the following equation, where the quantity of wool demanded, **Q**, depends on the price of wool ($\mathbf{P_W}$) and on the price of synthetics ($\mathbf{P_S}$). Use the above Box-Cox transformation when estimating the following demand relationship:

[3] A good example of this is evident when estimating a production function. Using a linear, in inputs and coefficients, production function does not allow for diminishing marginal products. Some type of quadratic function of input use may be more appropriate for all inputs as this specification allows the data to show evidence of declining marginal products with increased input use, ceteris paribus. You do not impose constant marginal products on your model.

$$\ln(Q_t) = \beta_1 + \frac{\beta_2\left(P_{Wt}^{\lambda} - 1\right)}{\lambda} + \frac{\beta_3\left(P_{St}^{\lambda} - 1\right)}{\lambda} + \varepsilon_t \qquad [1.1]$$

where β_1, β_2, β_3, λ are unknown parameters and ε_t is an iid random error where $e_t \sim (0, \sigma^2)$.

Using the same data as used in Assignment #2, run your code twice, once where have the set your algorithm choice variable set to run the GN algorithm. What is a natural starting value for λ that could make your estimation run relatively smoothly?

Once you have obtained parameter estimates, rerun your model, this time using the NR algorithm for parameter estimation. Make sure you use the same starting values, same convergence criteria and variable step length in both analyses so you can compare the estimation process.

Give a summary of your experience running both model specifications. Prepare a summary table providing key model characteristics across estimation.

b. (**5 pts**) For the estimation of eq. [1.1] what happens to the estimation process when you use fixed step lengths in your estimation compared to variable step length.

2. ***(55 pts)*** Use your preferred estimation algorithm results to answer the following questions with respect to the estimation of [1.1]
   a. (**5 pts**) What is the correlation coefficient between the predicted value of Q_t and actual quantity of wool exported, Q_t? (Note: I am not referring to $\ln(Q_t)$!)
   b. (**10 pts**) When determining the average value of Q_t, test the hypothesis that this equals the overall sample average of wool exports. In this calculation, make sure you evaluate Q_t for all observations when accounting for average value variability due to regression coefficient uncertainty. Note: **Do not** use the average prices in your evaluation.

c. (**5 pts**) Given your results, test the hypothesis that wool demand has a semi-log functional form with respect to $\mathbf{P_W}$ and $\mathbf{P_S}$.
d. (**5 pts**) Test the hypothesis that $\boldsymbol{\beta_2} = -\,\boldsymbol{\beta_3}$.
e. (**15 pts**) Estimate the own and cross price elasticities of demand (i.e., $\partial \ln \mathbf{Q}/\partial \ln \mathbf{P_W}$, $\partial \ln \mathbf{Q}/\partial \ln \mathbf{P_S}$) at the sample's mean prices.
   i. Test whether the own-price elasticity is *statistically* different from $\mathbf{-1}$. Is the own-price elasticity different from $\mathbf{-1}$ in an *economical* meaningful way?
   ii. Is the cross-price elasticity statistically different from **0**?
   iii. Test whether the own-price elasticity is equal but of opposite sign of the cross-price elasticity.
f. (**10 pts**) Answer the questions associated with part (e) above but instead of using the mean of the data, use observation specific elasticity values.
g. (**10 pts**) The above structure imposed the constraint that the transformation parameter, λ, is the same for both $\mathbf{P_S}$ and $\mathbf{P_W}$. Re-specifying [1.1] via the following where we eliminate this constraint and allow λ possibly varying across price variable. That is:

$$\ln\left(Q_t\right) = \beta_1 + \frac{\beta_2\left(P_{Wt}^{\lambda_W} - 1\right)}{\lambda_W} + \frac{\beta_3\left(P_{St}^{\lambda_S} - 1\right)}{\lambda_S} + e_t \qquad [2.1]$$

where β_1, β_2, β_3, λ_W, and λ_S are unknown parameters to be estimated. Provide the results of your estimation of equation [2.1]. Does the data provide statistical evidence that the transformation parameters are indeed different from one another? Provide two statistical tests of the null hypothesis that $\lambda_W = \lambda_S$. One of these tests should be directly SSE based.

3. (***70 pts***) For this exercise I would like you to use Mizon (1977)'s data set. He was one of the first authors to provide an overview of statistical inference under nonlinear specifications. In the 1977 paper he estimated a variety of specifications for production functions including the Cobb-Douglas (C-D) and Constant Elasticity of Substitution (CES) where they were allowed to have additive as well as multiplicative error terms. In

his study he used U.K. data on capital, labor use and a common output measure for 24 industries encompassing the years 1954, 1957 and 1960. The following table provides a summary of the variables contained in the ***mizon_1977*** dataset.

Table 1: Variables Contained in the Mizon (1977) dataset

Variable	Description	Units
Quant	Gross value-added at factor cost	Mil. $
Capital	Value of the stock of plant and machinery	Mil. $.
LF	Labor force available for work in the industry	1,000
Unemploy	Number of workers unemployed in the industry	1,000
Hour	Average hours per week worked by those employed	Hours
Year	Survey data year	Year
Industry	Industry ID Number	#

Note: In order to use this data you need to divide the ***LF*** and ***Unemploy*** variables by 1,000 for scaling purposes.

Below are representations of two CES production functions, one with a multiplicative error term, i.e., eq. [3.1], and the 2nd with an addititive error term via eq. [3.2].[4]

CES: *Multiplicative Disturbance*

$$\text{Quant}_t = \alpha\left[\delta \text{Capital}_t^{-\rho} + (1-\delta)\text{Labor}_t^{-\rho}\right]^{-\eta/\rho} \exp(\varepsilon_t) \qquad [3.1]$$

CES: *Additive Disturbance*

$$\text{Quant}_t = \alpha\left[\delta \text{Capital}_t^{-\rho} + (1-\delta)\text{Labor}_t^{-\rho}\right]^{-\eta/\rho} + \varepsilon_t \qquad [3.2]$$

where: $\alpha \equiv$ efficiency parameter, $\alpha > 0$;
$\eta \equiv$ the degree of homogeneity (scale parameter), $\eta > 0$;
$\delta \equiv$ distribution parameter, $0 < \delta < 1$;
$\rho \equiv$ the substitution parameter, $-1 < \rho < \infty$, $\rho \neq 0$;

The input *Capital* is identified in the above table. The amount of *Labor* used is defined as the number of ***annual hours worked***.

[4] For more detail concerning the CES production function, refer to the following PDF: ces_2.pdf .

a. (***10 pts***) Using the ***Mizon data***, estimate the [3.1] and [3.2] specifications via NLS using your preferred estimation method. For eq. [3.1] estimate the model after taking the natural logarithm of both sides of the equation. Present the typical regression statistics.

b. (***5 pts***) What evidence do you have that the SSE function is *at least a local minimum* at the parameter values you identified as those that minimize the SSE function?

c. (***5 pts***) For the CES model specifications what is the correlation between predicted and actual values of **Quant** (i.e., not **ln[Quant]**)? Why would one be interested in determining such correlations? [***Note****: When evaluating the relationship between predicted and actual values of* ***Quant*** *(versus* ***ln(Quant)****), you should note that* $\boldsymbol{E(\varepsilon)=0}$ *and* $\boldsymbol{E[\exp(\varepsilon_t)]}$ *need not equal 1.0 but in fact* $\boldsymbol{E[\exp(\varepsilon_t)] = \exp(\sigma^2/2)}$]. From these results, what specification would you say is preferred in terms of explaining the variance of **Quant** (not ***ln(Quant)***)?

d. (***5 pts***) Provide statistical evidence as to whether the production technology exhibits constant returns to scale under both specifications.

e. (***15 pts***) Under the logrithmic version of the CES specification in [3.1], evaluate the marginal products of Capital and Labor (i.e., $MP_C = \partial Quant/\partial Capital$ and $MP_L = \partial Quant/\partial Labor$) when these inputs are set at their mean values. Are these marginal products positive from a statistical point of view? [***Note:*** *Make sure that wherever* ***Quant*** *appears in the marginal product expression you use the predicted* ***Quant*** *value [****not ln(Quant)****] of this variable.*

This method will ensure that these marginal products are evaluated at a point associated with the predicted production function. The use of the predicted value ensures that the coefficient variability in this prediction will be accounted for in your estimate of the variance of the marginal products.] The MP_L and MP_C can be represented via the following:

$$MP_L \equiv \frac{\partial \text{Quant}}{\partial \text{Labor}} = \eta\alpha^{-\left(\rho/\eta\right)} \delta \text{Labor}^{-(1+\rho)} \text{Quant}^{\left(1+\rho/\eta\right)}$$

$$MP_C \equiv \frac{\partial \text{Quant}}{\partial \text{Capital}} = \eta\alpha^{-\left(\rho/\eta\right)} \left(1-\delta\right) \text{Capital}^{-(1+\rho)} \text{Quant}^{\left(1+\rho/\eta\right)}$$

f. (**5 pts**) Are the marginal products calculated in (e) equal to one-another

g. (**10 pts**) In contrast to the method used in (e) and (f) to calculate the input marginal products, use the logarithmic functional form of eq. [3.1] to test the null hypotheis that the ***average*** MP_L equals the average MP_K when you average these marginal products across **all observations**? That is, **do not** evaluate your MP's at the mean of the data. Also, remember when evaluating the variance of the MP's it is the impacts of parameter variance you are evaluating. Since you are evaluating the average MP's it is **essential** that the **averaging process is included** in the numerical gradient evaluation. How do your results here compare with those obtained in part (f) above?

h. (**10 pts**) For the CES specification used in (g), evaluate the null hypothesis that the logarithmic intercept varies across industry using a **single joint test**.

i. (**5 pts**) Explain, for the logarithmic version of eq [3.1] and the estimation of eq. [3.2], what happens to the estimation process when you use fixed step lengths in your estimation. Compare the number of iterations and final parameter vector values with the variable step length method estimation of the logarithmic version of eq [3.1] and to eq. [3.2].