4.1 INTRODUCTION

Transformations of variables have vital role in inference in Econometric models. The solutions or the remedies for the various problems in Econometrics such as Heteroscedasticity, Autocorrelation, Multicollinearity, Distributed log models, errors in variables, System of simultaneous structural equations models, Seemingly Unrelated Regression Equations (SURE) models, Nonlinear models etc; can be dealt with the transformations of either dependent Variable or independent Variables or both the variables. Bartlett (1947) used transformations of variables with reference to Analysis of Variance (ANOVA). The main purpose of using transformations in the ANOVA is to change the scale the of measurements in order to make analysis more valid. The transformed dependent variable in the ANOVA model will assumed to be normally distributed. The variance of the transformed variable will assumed to be unaffected by the change of measurements.

In econometrics, the logarithmic transformation of the dependent variable in growth studies leads naturally to the consideration of other transformations. These transformations may be of the dependent variable, the explanatory variable or both. The main use of the logarithmic transformation is to achieve a 'linearizing
transformation' so that the simple techniques may be applied to suitably transformed variables and thus obviate the need to fit more complicated nonlinear relations.

4.2 SOME TRANSFORMATIONS IN ECONOMIC ANALYSIS

(1) Log–Log Transformations:

The growth model has employed a transformation of the dependent variable. Many important Econometric applications involve the logs of both variables.

The functional specification is given by

\[ y = \alpha x^\beta e^x \]  \hspace{1cm} \ldots (4.2.1)

or

\[ \ln y = \ln \alpha + \beta \ln x + \varepsilon \] \hspace{1cm} \ldots (4.2.2)

\[ \Rightarrow Y = A + \beta X + \varepsilon \]

Where

\[ A = \ln \alpha, \ Y = \ln y; \ X = \ln x \]

Here, the slope of the log–log specification gives the elasticity of y with respect to x, i.e;

\[ \beta = \frac{dy}{dx} \frac{x}{y} \]
(2) **Semi log Transformations:**

(i) A widely used specification in human capital models is

\[ y = \alpha \beta^x e^\varepsilon \]  \hspace{1cm} \ldots (4.2.3)

or

\[ \ln y = \ln \alpha + (\ln \beta)x + \varepsilon \]  \hspace{1cm} \ldots (4.2.4)

or

\[ Y = A + Bx + \varepsilon \]

Where,

\[ A = \ln \alpha, B = \ln \beta, Y = \ln y \]

Here, \( y \) denotes earnings and \( X \) years of schooling or work experience. The slope in the semi log specification gives the proportionate change in \( y \) per unit change in \( X \). ie;

\[ B = \frac{1}{y} \frac{dy}{dX} \]  \hspace{1cm} \ldots (4.2.5)

(ii) In a cross-section study of household budgets a semi log specification is given by

\[ y = \alpha + \beta \ln x + \varepsilon \]  \hspace{1cm} \ldots (4.2.6)

Where,

\( y \) denotes expenditure and \( x \) denotes income.

Here, expenditure increases monotonically with income, but at a diminishing rate. The marginal prosperity to consume for this specification is given by \( \frac{\beta}{x} \).
This declines with increasing income, and the elasticity is given by \( \left( \frac{\beta}{y} \right) \) and it also declines as income increases.

(3) \textbf{Reciprocal Transformations}

These transformations have been used in modeling situations where there are asymptotes for one or both Variables.

A rectangular hyperbolic specification with asymptotes at \( y = \alpha_1 \) and \( x = \alpha_2 \) is given by

\[
(y - \alpha_1)(x - \alpha_2) = \alpha_3 \quad \cdots \quad (4.2.7)
\]

This may be written as

\[
y = \alpha_1 + \left( \frac{\alpha_3}{x - \alpha_2} \right) \quad \cdots \quad (4.2.8)
\]

The result of adding a disturbance term to (4.2.8) and attempting to minimize the residual sum of squares gives equations that are nonlinear in the \( \alpha \)'s.

In this case, there is no possible linearizing transformation. However there are two special cases of equation (4.2.8) where linearizing transformations are available.

Setting \( \alpha_1 = 0 \), gives

\[
y = \alpha + \beta \left( \frac{1}{x} \right)
\]
Where,

\[ \alpha_1 = \alpha, \beta = \alpha_3. \]

Setting

\[ \alpha_1 = 0 \] gives, \[ y = \frac{\alpha_3}{x - \alpha_2} \]
or

\[ \frac{1}{y} = \left( \frac{-\alpha_2}{\alpha_3} \right) + \left( \frac{1}{\alpha_3} \right) \]

\[ \Rightarrow \frac{1}{y} = \alpha + \beta x + \epsilon \quad \ldots (4.2.9) \]

Where,

\[ \alpha = -\frac{\alpha_2}{\alpha_3} \quad \text{and} \quad \beta = \frac{1}{\alpha_3} \]

These transformed equations have been frequently used in the study of 'Phillips Curves', where \( y \) represents the usage rate or price change and \( x \) represents the unemployment rate.

4.3 BOX AND COX TRANSFORMATIONS: TWO-VARIABLE RELATIONSHIPS

In very rare cases, economic theory may indicate the appropriate transformation of variables. Zarembka (1974) pointed out that constant elasticity of substitution (CES) production function

\[ y = (\eta_1 K^\rho + \eta_2 L^\rho)^{\nu/\rho} \quad \ldots (4.3.1) \]

gives

\[ Y^{\rho/\nu} = \eta_1 K^\rho + \eta_2 L^\rho \quad \ldots (4.3.2) \]
which is an example of power transformations, of variables, and it is linear in the transformed variables. It poses difficult estimation problems. Some special cases of power transformations in two-variable relationship, however, yield simple estimating procedures.

Box and Cox (1964) have proposed a very general form of transformation for two-variable relationships as

\[ Y^{(\lambda_1)} = \frac{Y^{\lambda_1} - 1}{\lambda_1}, \text{ if } \lambda_1 \neq 0 \]
\[ \ln Y, \text{ if } \lambda_1 = 0. \]  \hspace{1cm} \ldots (4.3.3)

and

\[ X^{(\lambda_2)} = \frac{X^{\lambda_2} - 1}{\lambda_2}, \text{ if } \lambda_2 \neq 0 \]
\[ \ln X, \text{ if } \lambda_2 = 0. \]  \hspace{1cm} \ldots (4.3.4)

consider the two-variable linear model with the transformed variables \( Y^{(\lambda_1)} \) and \( X^{(\lambda_2)} \) as

\[ y_i^{(\lambda_1)} = \alpha + \beta x_i^{(\lambda_2)} + \varepsilon_i, i = 1, 2, \ldots, n \]  \hspace{1cm} \ldots (4.3.5)

and \( \varepsilon_i \sim N(0, \sigma^2_e), \quad \forall \ i = 1, 2, \ldots, n \).

This model has five basic parameters namely, \( \alpha, \beta, \lambda_1, \lambda_2 \) and \( \sigma^2_e \).

For particular values of \( \lambda_1 \) and \( \lambda_2 \), one may obtain different forms of two variable regression models.
(1) **Double logarithmic Model**

For \( \lambda_1 = 0, \lambda_2 = 0 \), equation (4.3.5) may be written as

\[
\ln Y = \alpha + \beta \ln X + \epsilon \quad \quad \text{... (4.3.6)}
\]

\[
\Rightarrow \quad Y = a X^\beta e^\epsilon ,
\]

Where,

\[
\ln a = \alpha .
\]

This relationship exists only for positive values of the variables. It is a constant elasticity function and the elasticity is given by \( \beta \). For \( \beta = -1 \), equation (4.3.6) gives \( X Y = a e^\epsilon \), which is a rectangular hyperbola.

(2) **Semi Logarithmic Model**

For \( \lambda_1 = 0, \lambda_2 = 1 \), equation (4.3.5) may be written as

\[
\ln Y = \alpha^* + \beta X + \epsilon ,
\]

Where \( \alpha^* = \alpha - \beta . \)

\[
\Rightarrow \quad Y = e^{\alpha^* + \beta X + \epsilon} \quad \text{... (4.3.7)}
\]

This semi logarithmic function is defined only for positive values of \( Y \). When \( X \) denotes time variable, then \( Y = e^{\alpha^* + \beta X + \epsilon} \) becomes a growth curve and \( \beta = \frac{1}{Y_i} \frac{dY_i}{dt} \) gives instantaneous rate of growth of \( Y \) at time \( t \).
(3) **Linear Model**

For $\lambda_1 = 1$, $\lambda_2 = 1$, equation (4.3.5) may be written as

\[
(Y - 1) = \alpha + \beta(X - 1) + \varepsilon
\]

\[Y = (\alpha - \beta + 1) + \beta X + \varepsilon \quad \ldots \quad (4.3.8)
\]
or

\[Y = \alpha^{**} + \beta X + \varepsilon, \quad \ldots \quad (4.3.9)
\]

Where

\[\alpha^{**} = (\alpha - \beta + 1)\]

Equation (4.3.9) gives a simple linear model.

(4) **Reciprocal Model**

For $\lambda_1 = 1$, $\lambda_2 = -1$, the Equation (4.3.5) may be written as

\[
(Y - 1) = \alpha - \beta \left( \frac{1}{X} - 1 \right) + \varepsilon
\]

or

\[Y = (\alpha + \beta + 1) - \beta \left( \frac{1}{X} \right) + \varepsilon \quad \ldots \quad (4.3.10)
\]
or

\[Y = \alpha^{***} + \beta^* \left( \frac{1}{X} \right) + \varepsilon, \quad \ldots \quad (4.3.11)
\]

Where

\[\alpha^{***} = \alpha + \beta + 1 \text{ and } \beta^* = -\beta\]

Equation (4.3.11) is a reciprocal model.
(5) Logarithmic Reciprocal Model

For \( \lambda_1 = 0, \lambda_2 = -1 \), the equation (4.3.5) may be written as

\[
\ln Y = \alpha - \beta \left( \frac{1}{X} - 1 \right) + \epsilon
\]

or

\[
\ln Y = (\alpha + \beta) - \beta \left( \frac{1}{X} \right) + \epsilon \quad \cdots \quad (4.3.12)
\]

or

\[
\ln Y = \alpha^{***} + \beta^* \left( \frac{1}{X} \right) + \epsilon \quad \cdots \quad (4.3.13)
\]

Where

\[
\alpha^{***} = (\alpha + \beta) \text{ and } \beta^* = -\beta
\]

Equation (4.3.13) gives a logarithmic reciprocal model.

Equation (4.3.13) may we written as

\[
Y = e^{\alpha^{***}} + \frac{\beta^*}{X} + \epsilon \quad \cdots \quad (4.3.14)
\]

If \( X \) represents a time variable then equation (4.3.14) gives a growth curve which starts at zero and approaches an asymptotic level.

Now, consider the Box and Cox two variables transformed model as

\[
Y^{(\lambda_1)} = \alpha + \beta X^{(\lambda_2)} + \epsilon \quad \cdots \quad (4.3.15)
\]

Particular values of 0, 1, or -1 for the \( \lambda \) parameters, we have obtained different functional forms and in each case, a simple linear relationship holds between the transformed Variables. If the \( \lambda \)'s are
free to take on any values, then approximation to linearity may well be improved. In this case, one has to estimate five parameters namely, $\alpha$, $\beta$, $\lambda_1$, $\lambda_2$, and $\sigma_\varepsilon^2$. This may lead to nonlinear estimation.

4.4 TRANSFORMATION MATRIX IN ESTIMATION OF LINEAR REGRESSION MODEL UNDER HETEROSEDASTICITY

Consider a linear model under non-spherical disturbances as

$$Y_{n \times 1} = X_{n \times k} \beta_{k \times 1} + \varepsilon_{n \times 1} \quad \ldots (4.4.1)$$

Where,

$X$ is a non-stochastic matrix with full column rank,

Such that (i) $E[\varepsilon] = 0$

and

$$
\begin{pmatrix}
\sigma_1^2 & 0 & \ldots & 0 \\
0 & \sigma_2^2 & \ldots & 0 \\
\end{pmatrix}
$$

(ii) $E[\varepsilon \varepsilon'] = \Sigma =

$$
\begin{bmatrix}
\sigma_1^2 & 0 & \ldots & 0 \\
0 & \sigma_2^2 & \ldots & 0 \\
\end{bmatrix}
$$

Which is the standard case of heteroscedasticity. This specification assumes that the disturbances are pair wise uncorrelated.

The ordinary least squares (OLS) estimator of $\beta$ is given by

$$\hat{\beta} = (X'X)^{-1} X' Y$$

and its dispersion matrix is given by
\[ \text{Var}(\hat{\beta}) = E[\hat{\beta} - \beta][\hat{\beta} - \beta] = E\left[ (X'X)^{-1}X' \epsilon \epsilon' X (X'X)^{-1} \right] \]

\[ \text{Var}(\hat{\beta}) = (X'X)^{-1}(X' \Sigma X')(X'X)^{-1} \quad \ldots (4.4.2) \]

The optimal minimum variance property of OLS estimators no longer holds. Suppose one of the explanatory variables determines the heteroscedasticity.

Let the heteroscedasticity be modeled by

\[ \sigma_i^2 = \sigma^2 X_i^2, \quad i = 1, 2, \ldots, n. \quad \ldots (4.4.3) \]

Where \( X_i \) is the independent variable thought to be the source of the heteroscedasticity. Now, the dispersion matrix of \( \epsilon \) is given by

\[
\text{Var}(\epsilon) = \sigma^2 \begin{bmatrix}
X_{j1}^2 & 0 & \ldots & 0 \\
0 & X_{j2}^2 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & X_{jn}^2
\end{bmatrix} = \sigma^2 \Omega \quad \ldots (4.4.4)
\]

\( \Omega \) is a Symmetric positive definite matrix.

The appropriate non singular transformation matrix is given by
\[
P^{-1} = \begin{pmatrix}
0 & 0 \\
0 & 1 \\
\end{pmatrix}
\]
\[(4.4.5)\]

Such that \([P^{-1} \Omega P^T] = (P^\dagger \Omega^\dagger P)^{-1} = I\]

Consider the original multiple linear regression model as

\[Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_j X_{ji} + \ldots + \beta_k X_{ki} + \epsilon_i,\]

\[i = 1, 2, \ldots, n. \quad (4.4.6)\]

The transformed model can be written as

\[\frac{Y_i}{X_{ji}} = \beta_1 \left( \frac{1}{X_{ji}} \right) + \beta_2 \left( \frac{X_{2i}}{X_{ji}} \right) + \ldots + \beta_j \left( \frac{X_{ji}}{X_{ji}} \right) + \ldots + \beta_k \left( \frac{X_{ki}}{X_{ji}} \right) + \left( \frac{\epsilon_i}{X_{ji}} \right)\]

\[\quad \ldots (4.4.7)\]

The original model may be written in matrix notation as

\[Y_{nk} = X_{nk} \beta_{kx} + \epsilon_{nk} \quad (4.4.8)\]

Pre multiplying both sides of (4.4.8) by non singular transformation matrix \(P^\dagger\) gives

\[P^\dagger Y = P^\dagger X \beta + P^\dagger \epsilon \quad (4.4.9)\]
or

\[ Y^* = X^* \beta + \varepsilon^* \]

Such that

\[ (P^l \Omega^{-l} P)^{-l} = I \text{ or } \Omega^{-l} = (P P^l)^{-l} \]

Where,

\[ Y^* = P^{-l} Y, \quad X^* = P^{-l} X \quad \text{and} \quad \varepsilon^* = P^{-l} \varepsilon \]

One may have

\[ E[\varepsilon^*] = P^{-l} E(\varepsilon) = 0 \]

and

\[ E[\varepsilon^* \varepsilon^*'] = \begin{pmatrix} P^{-l} E(\varepsilon \varepsilon') P^{-l} \\ \sigma^2 P^{-l} \Omega P^{-l} \end{pmatrix} \]

\[ E[\varepsilon^* \varepsilon^*'] = \begin{pmatrix} \sigma^2 (P^{-l} \Omega^{-l} P)^{-l} \\ \sigma^2 \end{pmatrix} = \sigma^2 I \]

Applying the OLS estimation gives

\[ \tilde{\beta} = \left( X'^{-l} X^* \right)^{-l} \left( X'^{-l} Y^* \right) = \left[ X^l \left( P^{-l} P^l \right) X \right]^{-l} \left[ X^l \left( P^{-l} P^l \right) Y \right] \]

\[ \Rightarrow \quad \tilde{\beta} = \left( X^l \Omega^{-l} X \right)^{-l} \left( X^l \Omega^{-l} Y \right) \quad \text{...(4.4.10)} \]

and

\[ \text{Var}(\tilde{\beta}) = \sigma^2 \left( X'^{-l} X^* \right)^{-l} = \sigma^2 \left( X^l \Omega^{-l} X \right)^{-l} \quad \text{...(4.4.11)} \]

The estimator \( \tilde{\beta} \) is the generalized least squares (GLS) estimator of \( \beta \) and the best linear unbiased estimator of \( \beta \) in the linear model with heteroscedastic disturbances.
An unbiased estimator of error variance $\sigma^2$ is given by

$$\hat{\sigma}^2 = \frac{(Y^* - X^* \hat{\beta})(Y^* - X^* \hat{\beta})}{n - k}$$

or

$$\hat{\sigma}^2 = \frac{(Y - X \hat{\beta})' \Omega^{-1} (Y - X \hat{\beta})}{n - k} \quad \text{... (4.4.12)}$$

Here,

$$\Omega^{-1} = \text{diag}\left\{ \frac{1}{X^2_{j1}}, \frac{1}{X^2_{j2}}, ..., \frac{1}{X^2_{jn}} \right\}$$

Also, the elements in the $\Omega$ assumed to be known exactly. In many cases, the elements in $\Omega$ have to be estimated and then substituted in the GLS estimator formulae. The resultant estimator will be known as the Feasible GLS estimator.

4.5 TRANSFORMED ESTIMATION OF LINEAR REGRESSION MODEL WITH FIRST ORDER AUTOREGRESSIVE ERRORS

Consider the linear regression model with first order autoregressive disturbances as

$$y_i = X_i' \beta + \epsilon_i, \quad i = 1, 2, ..., n \quad \text{... (4.5.1)}$$

Such that

$$\epsilon_i = \rho \epsilon_{i-1} + u_i \quad \text{... (4.5.2)}$$
Where,

\[ Y_i \] is the \( i^{th} \) observation on dependent variable \( Y \),

\[ X_i \] is \((k \times 1)\) vector consists of \( i^{th} \) observation on \( k \) explanatory variables,

\( \beta \) is \((k \times 1)\) vector of unknown parameters,

\( \varepsilon_i \) is \( i^{th} \) observation on disturbance variable \( \varepsilon \),

\( \rho \) is first order autocorrelation coefficient.

Here,

\[ E[u_i] = 0, \quad E[u_i^2] = \sigma_u^2, \quad E[u_i u_j] = 0, \quad \forall i \neq j, |\rho| < 1. \]

The usual dispersion matrix of errors is given by

\[
\begin{pmatrix}
1 & \rho_1 & \rho_2 & \ldots & \rho_{n-1} \\
\rho_1 & 1 & \rho_1 & \ldots & \rho_{n-2} \\
\rho_2 & \rho_1 & 1 & \ldots & \rho_{n-3} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \ldots & 1
\end{pmatrix} = \begin{pmatrix}
\varepsilon_i \varepsilon_i' \end{pmatrix} = \begin{pmatrix}
\sigma^2 \\
\sigma^2 \\
\sigma^2 \\
\vdots \\
\sigma^2 \\
\end{pmatrix} = \sigma^2 \Psi = \sigma, \quad \rho_1 = 1 \ldots \rho_{n-1} \rho_{n-2} \rho_{n-3} \ldots 1
\]

... (4.5.3)

Where

\[
\rho_s = \frac{E[\varepsilon_i \varepsilon_{i+s}]}{\sigma^2} = \frac{E[\varepsilon_i \varepsilon_{i+s}]}{\sigma^2} \quad \text{for} \quad s = 1, 2, \ldots \quad \ldots (4.5.4)
\]

is the correlation coefficient between two disturbances with \( S \) periods apart and \( \sigma^2 \) is the Variance of \( \varepsilon, \forall i = 1, 2, \ldots \). The number of
Unknown Parameters in the autocorrelation matrix $\psi$ is to be reduced for particular autoregressive process. For the first order autoregressive process, defined in (4.5.2), the elements of $\psi$ are given by

$$\rho_i = \rho^i \quad \ldots \quad (4.5.5)$$

and

$$\sigma^2_e = \frac{\sigma^2_u}{1 - \rho^2} \quad \ldots \quad (4.5.6)$$

Thus,

$$E[\epsilon \epsilon'] = V = \sigma^2_e \psi = \sigma^2_e \cdot W,$$

Where,

$$W = \begin{bmatrix} \begin{array}{cccc} \psi \end{array} \end{bmatrix}$$

$V$ can be written as

$$V = \sigma^2_e \psi = \begin{bmatrix} 1 & \rho & \rho^2 & \ldots & \rho^{n-1} \\ \rho & (1+\rho^2) & -\rho & \ldots & 0 \\ \rho^2 & -\rho & (1+\rho^2) & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \ldots & 1 - \rho^2 \end{bmatrix} \quad \ldots \quad (4.5.7)$$

and

$$V^{-1} = \frac{1 - \rho^2}{\sigma^2_u} \psi^{-1} = \frac{1}{\sigma^2_u} \begin{bmatrix} 1 & -\rho & 0 & \ldots & 0 & 0 \\ -\rho & (1+\rho^2) & -\rho & \ldots & 0 & 0 \\ 0 & -\rho & (1+\rho^2) & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & (1+\rho^2) & -\rho \\ 0 & 0 & 0 & \ldots & -\rho & 1 \end{bmatrix}$$
\[ V^{-1} = \frac{1}{\sigma_a^2} W^{-1} \]  \hspace{1cm} (4.5.8)

Under the application of Generalized least Squares (GLS) estimation, the least squares estimator \( \tilde{\beta} = (X^* X^*)^{-1} X^* Y^* \) is obtained from the transformed model

\[ Y^* = X\beta + \epsilon^* \]

\[ Y^* = PY, \quad X^* = PX, \quad \epsilon^* = P\epsilon \]

\[ \begin{array}{cccccc}
1 - \rho^2 & 0 & 0 & \ldots & 0 & 0 \\
-\rho & 1 & 0 & \ldots & 0 & 0 \\
0 & -\rho & 1 & \ldots & 0 & 0 \\
\end{array} \]

and

\[ \begin{array}{ccc}
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & -\rho & 1 \\
\end{array} \]

... (4.5.9)

Here, \( P \) is the transformation matrix such that

\[ P^tP = W^{-1} = (1 - \rho^2) \psi^{-1} \]  \hspace{1cm} (4.5.10)

The transformed equations are given by

\[ (1 - \rho^2) \gamma_i = \sqrt{1 - \rho^2} X_i^t \beta + \sqrt{1 - \rho^2} \epsilon_i ; \]  \hspace{1cm} (4.5.11)
\[ Y_j - \rho Y_{j-1} = (X_j - \rho X_{j-1}) \beta + e_j - \rho e_{j-1}, \quad j = 2, 3, \ldots, n. \] ... (4.5.12)

If \( \rho \) is known, then the GLS estimator of \( \beta \) is given by

\[ \tilde{\beta} = \left( X \psi^{-1} X \right)^{-1} X \psi^{-1} Y. \] ... (4.5.13)

This GLS estimator \( \tilde{\beta} \) is obtained by finding that \( \beta \) which minimizes the transformed residual sum of squares

\[ (\hat{Y}^* - \hat{X}^* \beta) (\hat{Y}^* - \hat{X}^* \beta) = (Y - X \beta)^{PP} (Y - X \beta) \]
\[ \hat{\epsilon}_1^{*2} + \sum_{j=2}^{n} u_j^2 \] ... (4.5.14)

Where

\[ \hat{\epsilon}_1^* = \sqrt{1 - \rho^2} y_1 - \sqrt{1 - \rho^2} X_1 \beta \] ... (4.5.15)

and

\[ u_j = y_j - \rho y_{j-1} - (X_j - \rho X_{j-1}) \beta, \quad j = 2, 3, \ldots, n. \] ... (4.5.16)

This model was used by Cochrane and Orcutt (1949).

Given an estimate for \( \rho \) say

\[ \hat{\rho} = \frac{\sum_{j=2}^{n} e_j e_{j-1}}{\sum_{j=1}^{n} e_j^2} \] ... (4.5.18)
Where, $\varepsilon_i$'s are the OLS residuals; and the corresponding matrices $\hat{P}$ and $\hat{W}$ such that $\hat{P}^\dagger \hat{P} = \hat{W}^\dagger$.

The estimated GLS (EGLS) estimator is given by

$$\hat{\beta} = \left( X^\dagger \hat{W}^{-1} X \right)^{-1} \left( X^\dagger \hat{W}^{-1} Y \right) \quad (4.5.19)$$

### 4.6 COCHRANE AND ORCUTT TRANSFORMATION FOR THE ESTIMATION OF LINEAR MODEL WITH AUTO CORRELATED DISTURBANCES

Consider the linear regression model

$$Y = X\beta + \varepsilon \quad (4.6.1)$$

Where,

$Y$ is an $n \times 1$ vector of successive observations on the dependent variable

$X$ is an $n \times k$ matrix (rank $k$) of successive observations on the independent variables

and $\varepsilon$ is an $(n \times 1)$ random vector of disturbances with zero means.

The disturbances are assumed to follow an autoregressive process:

$$\varepsilon_t = P \varepsilon_{t-1} + \nu_t \quad (4.6.2)$$

Where

$|P| < 1, \ E(\nu_t) = 0, \ E(\nu_t^2) = \sigma^2, \ E(\nu_t \nu_{t-S}) = 0$ for $S \neq 0$
Given the above set up, it is known that

\[
\begin{bmatrix}
1 & \rho & \rho^2 & \ldots & \rho^{n-1} \\
0 & 1 & \rho & \rho^2 & \ldots & \rho^{n-2} \\
0 & 0 & 1 & \rho & \rho^2 & \ldots & \rho^{n-3} \\
& & & & & & \ddots
\end{bmatrix}
\]

\[
E[\epsilon \epsilon'] = \frac{1}{1 - \rho^2} \begin{bmatrix}
\rho^{n-2} & \rho^{n-3} & \rho^{n-4} & \ldots & \rho \\
\rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \ldots & 1
\end{bmatrix}
\]

\[
W^{-1} = \frac{1}{1 - \rho^2} 
\begin{bmatrix}
1 & -\rho & 0 & 0 & 0 & 0 \\
-\rho & 1 + \rho^2 & -\rho & 0 & 0 & 0 \\
0 & -\rho & 1 + \rho^2 & 0 & 0 & 0 \\
& & & & & \ddots
\end{bmatrix}
\]

Further, it can be verified that transformation that would convert

(4.6.1) to a standard linear regression model with scalar matrix,

When \( \rho \) is known is given by;

\[
Y^* = PY \\
X^* = PX \\
\epsilon^* = P \epsilon
\]
Where,

\[ P \text{ is given by} \]

\[
\begin{pmatrix}
(1 - \rho^2)^{\frac{1}{2}} & 0 & 0 \\
-\rho & 1 & 0 \\
0 & 0 & 0 \\
\end{pmatrix}
\]

\[ P = \]

\[
\begin{pmatrix}
0 & 0 & 0 \\
-\rho & 1 & 0 \\
0 & 0 & 0 \\
\end{pmatrix}
\]

(4.6.6)

and is the unique lower triangular matrix, with positive diagonal elements, such that

\[ P^W P^\dagger = (1 - \rho^2) I \]

This furnishes the best linear unbiased estimator (BLUE) of \( \beta \) by Aitken's well known formulae

\[ \hat{\beta} = \left(X^\dagger X^*\right)^\dagger X^\dagger Y^* = (X^\dagger P^\dagger PX)^\dagger XP^\dagger PY = \left(X^\dagger W^{-1} X\right)^\dagger X^\dagger W^{-1} Y \]

\[ \ldots (4.6.7) \]

But the following transformation, which was originally suggested by Cochrane and Orcutt (1949) in (4.6.2) is recommended very often;

\[ Q_y = QX\beta + Q \epsilon, \quad \ldots (4.6.8) \]

Where,

\( Q \) is the \((n - 1) \times n\) matrix from \( P \) by deleting the top row.
The only possible advantage that (4.6.8) have over (4.6.6) is computational. We will show in what follows that the simple least squares estimator (SLSE) $\hat{\beta} = \left( X'X \right) X'Y$, which is the simplest from a computational point of view which does not require a knowledge of $\rho$, is more efficient then the estimator obtained by using the transformation (4.6.8) in some cases.

**Relative Efficiency Of Cochrane – Orcutt Estimator**

Denoting the SLSE from (4.6.1) and (4.6.8) by $\hat{\beta}$ and $\bar{\beta}$ respectively, we may define the relative efficiency $E$, of $\bar{\beta}$ with respect to $\hat{\beta}$ as

$$E = \frac{|\sum \hat{\beta}|}{|\sum \bar{\beta}|} \quad \ldots \quad (4.6.9)$$

Where,

$$\sum_{\hat{\beta}} \text{ and } \sum_{\bar{\beta}} \text{ are the covariance matrices of } \hat{\beta} \text{ and } \bar{\beta} \text{ respectively.}$$

But

$$\sum_{\hat{\beta}} = \frac{\sigma^2}{1-\rho^2} (X'X) X'WXX'X$$

$$\ldots \quad (4.6.10)$$

and

$$\sum_{\bar{\beta}} = \frac{\sigma^2}{1-\rho^2} (X'O'QX)'X'O'O\bar{W}O\bar{Q}X(X'O'QX)'$$

$$\ldots \quad (4.6.11)$$
Therefore, using the formulae (4.6.10) and (4.6.11) and noting that

\[ Q'Q = (1 - \rho^2)(W^{-1}Q) \quad \text{... (4.6.12)} \]

and

\[ QWQ' = (1 - \rho^2)I, \quad \text{... (4.6.13)} \]

Where \( Q \) is an \( n \times n \) matrix having one in the upper left corner and zeros every where else, the formula for \( E \) reduces to

\[ E = \frac{X'WX\|X'(W^{-1} - \theta)X\|}{X'X} \quad \text{... (4.6.14)} \]

Now let us consider a special case in which \( X = (a \text{ column of } n \text{ ones}). \)

Then we have

\[ X'WX = W'L = n + 2(n - 1)\rho + (n - 2)\rho^2 + \ldots + 2\rho^{n-1} \]

\[ X'WX = W'L = \frac{n(1 - \rho^2) - 2\rho(1 - \rho^n)}{(1 - \rho)^2} \quad \text{... (4.6.15)} \]

and

\[ X'(W^{-1} - \theta)X = l'(R^- - \theta) = \frac{(n - 1)(1 - \rho)}{(1 + \rho)} \quad \text{... (4.6.16)} \]

Therefore, the formula for \( E \) specializes to

\[ E = \frac{n(1 - \rho^2) - 2\rho(1 - \rho^n)(n - 1)(1 - \rho)}{n^2(1 - \rho)^2} \frac{(n - 1)(1 - \rho)}{(1 + \rho)} \quad \text{... (4.6.17)} \]
\[ E = \frac{(n-1)}{n} \left[ 1 - 2\rho \left( 1 - \rho^2 \right) \right] \leq \frac{(n-1)}{n} \] ... (4.6.18)

For \( 0 \leq \rho < 1 \). That is, \( \hat{\beta} \) - which is the SLSE of \( \beta \) is superior to \( \bar{\beta} \) for \( 0 \leq \rho < 1 \).

The efficiency of the estimator obtained by the Cochrane-Orcutt transformation is actually less than that of the simple least squares estimator, when positive auto correlation in the residuals is present. On the other hand, the addition of one weighted observation to the Cochrane-Orcutt procedure yields the best linear unbiased estimator.

4.7 BOX-COX TRANSFORMATION IN A LINEAR REGRESSION WITH A LAGGED DEPENDENT VARIABLE

Box and Cox (1964) proposed a power transformation to select an appropriate form of a dependent variable, the extended Box - Cox model have widely been used in applied researches. Some examples in applied econometrics are Zarembka (1986) and Oxley (1986). Using the Box-Cox transformation in estimation of money demand equations and Khan and Ross (1977) and Boylan (1980, 1982) using it in estimation of import demand equations.
Although a lagged dependent variable is often included among explanatory variables in applied researches using the Box-Cox transformation, the lagged dependent variable has been treated as if it were an exogenous explanatory variable.

If the first observation of a dependent variable is non-random, the usual maximum likelihood (ML) estimation procedure yields a correct ML estimate of the transformation parameter. However, if the first observation of a dependent variable is random, the ML estimate of the transformation parameter obtained by the usual ML estimation procedure is not a correct ML estimate.

Consider a Box-Cox model with the lagged dependent variable as an explanatory variable:

\[ Y_t^{(\lambda)} = \gamma Y_{t-1}^{(\lambda)} + x_t^{(\lambda)'} \beta + u_t, \quad t = 2, 3, \ldots, n \quad \ldots \text{(4.7.1)} \]

\[ Y_t^{(\lambda)} = u_t + \epsilon_t, \quad \ldots \text{(4.7.2)} \]

Where,

\( Y_t \) is the \( t \)-th observation on the dependent variable.

\( x_t \) is a 1 x \( k \) vector of observations on exogenous explanatory variables.

\( \gamma \) is a scalar parameter for the lagged dependent variable

\( |\gamma| < 1 \)

and \( \beta \) is a \( k \times 1 \) vector of coefficients.
As to the error terms. We assume that \( u_2, u_3, \ldots, u_n \) are normally and independently distributed with mean zero and variance \( \sigma^2 \), and \( u_1 \) is normally and independently of \( u_2, u_3, \ldots, u_n \), distributed with mean zero and variance \( \sigma_1^2 \). The mean and variance of \( Y_1^{(\lambda)} \) are specified later.

The transformation of \( Y_1 \) is defined as

\[
Y_1^{(\lambda)} = \begin{cases} 
\frac{Y_i^\lambda - 1}{\lambda} & \text{for } \lambda \neq 0, \\
\log Y_i & \text{for } \lambda = 0.
\end{cases} \tag{4.7.3}
\]

Where 'log' denotes a natural logarithm. The transformation of the components of \( x_i \) is similarly defined, except for the constant term. The model (4.7.1) reduces to a log-linear form when \( \lambda = 0 \), and to a linear form when \( \lambda = 1 \).

In the model (4.7.1) the same transformation parameter \( \lambda \) is used for all variables. Although it may be possible to use different \( \lambda \) values for each variable, it requires a considerable computational burden. When \( k=2 \) the ML estimation procedure given below requires the grid search over the four dimensional region of parameters \( \lambda_1 \) and \( \lambda_2 \) (for \( x_1 \) and \( x_2 \)) and \( \lambda \) and \( \gamma \). Thus, Ohtani (1990) used the same transformation parameter for all variables. However,
even when different $\lambda$ values are used, the ML estimation procedure is almost the same as the one using a single transformation parameter except for the computational burden.

Since the lagged dependent variable is included among explanatory variables, the model (4.7.1) is a first order autoregressive process of $Y_t^{(\lambda)}$ with the exogenous variables $x_t^{(\lambda)}$. Continuing to substitute in (4.7.1) we obtain

$$Y_t^{(\lambda)} = \gamma^{t-1} Y_1^{(\lambda)} + \sum_{i=0}^{t-2} \gamma^i x_{t-i}^{(\lambda)} \beta + \sum_{i=0}^{t-2} \gamma^i u_{t-i} \quad \ldots \quad (4.7.4)$$

From (4.7.4) the mean and variance of $Y_t^{(\lambda)}$ for $t = 2, 3, \ldots, n$ are

$$E[Y_t^{(\lambda)}] = \gamma^{t-1} E[Y_1^{(\lambda)}] + \sum_{i=0}^{t-2} \gamma^i E[x_{t-i}^{(\lambda)}] \beta \quad \ldots \quad (4.7.5)$$

$$\mu_t = x_t^{(\lambda)} \beta \quad \ldots \quad (4.7.6)$$

$$\sigma_1^2 = \frac{\sigma^2}{(1-\gamma^2)} \quad \ldots \quad (4.7.7)$$

In this case, the mean and variance of $Y_t^{(\lambda)}$ for $t = 1, 2, \ldots, n$ are written as

$$E[Y_t^{(\lambda)}] = \sum_{i=0}^{t-1} \gamma^i x_{t-i}^{(\lambda)} \beta \quad \ldots \quad (4.7.8)$$
\[ V(y_r^{(2)}) = (1 - \gamma^2) \]  

... (4.7.9)

Note that \((1 - \gamma^2)^{1/2} Y_1^{(2)}\) is distributed as \(N\left((1 - \gamma^2)^{1/2} x_1^{(2)} \right) \beta, \sigma^2\):

\[
\begin{pmatrix}
(1 - r^2)^{1/2} Y_1^{(2)} \\
Y_2^{(2)} - rY_1^{(2)}
\end{pmatrix}
\quad \begin{pmatrix}
(1 - r^2)^{1/2} x_1^{(2)} \\
x_2^{(2)} - r x_1^{(2)}
\end{pmatrix}
\]

Denoting \(Y^* = \begin{pmatrix}
Y_1^{(2)} \\
Y_2^{(2)} - rY_1^{(2)}
\end{pmatrix}\) and \(X^* = \begin{pmatrix}
x_1^{(2)} \\
x_2^{(2)} - r x_1^{(2)}
\end{pmatrix}\),

\[ u = (1 - \gamma^2)^{1/2} u_1, \ u_2, \ldots, u_\tau \]

The model (4.7.1) finally written as

\[ Y^* = X^* \beta + u^* \quad \ldots (4.7.10) \]

Since \(u^*\) is distributed as \(N(0, \sigma^2 I_n)\) and the Jacobian of transformation of the dependent variable is

\[ J = \det \frac{\partial Y^*_t}{\partial Y_s} = (1 - \gamma^2)^{1/2} \prod_{t=1}^n Y_t^{1/2} \quad \ldots (4.7.11) \]

\[ \text{det} \quad 310 \]

\[ \nu 519 \]
Where,

\( Y_t^* \) is the \( t \)th component of \( Y^* \), the log–likelihood function is

\[
L\left(\gamma, \beta, \sigma^2, \frac{\lambda}{Y}, X \right) = -\left(\frac{n}{2}\right) \log 2\Pi - \left(\frac{n}{2}\right) \log \sigma^2 - \frac{\left(\gamma^* - X^* \beta\right)}{(2\sigma)^2} \left(\frac{\gamma^* - X^* \beta}{(2\sigma)^2}\right) \\
+ \left(\frac{1}{2}\right) \log(1 - r^2) + (\lambda - 1) \sum_{i=1}^{n} \log y_i \quad \ldots \quad (4.7.12)
\]

The maximum likelihood (ML) estimators of \( \beta \) and \( \sigma^2 \) given both \( \lambda \) and \( \gamma \) are

\[
\hat{\beta}^* (\gamma, \lambda) = \left(X^* X^*\right)^{-1} X^* Y^* \quad \ldots \quad (4.7.13)
\]

\[
\hat{\sigma}^{*2} (\phi, \gamma, \lambda) = \frac{\left[Y^* - X^* \hat{\beta}^* (\gamma, \lambda)\right] \left[Y^* - X^* \hat{\beta}^* (\gamma, \lambda)\right]}{n} \quad \ldots \quad (4.7.14)
\]

Substituting (4.7.13) and (4.7.14) in to (4.7.12), the concentrated log–likelihood function is obtained:

\[
L_{\text{max}} = \left(\gamma, \frac{\lambda}{Y}, X \right) = -\left(\frac{n}{2}\right) (1 + \log 2\Pi) - \left(\frac{n}{2}\right) \log \hat{\sigma}^{*2} (\gamma, \lambda) \quad \ldots \quad (4.7.15)
\]

Random, the first rows of \( Y^* \), \( X^* \) and \( u^* \) should be deleted, and the Jacobian should be altered accordingly. In this case, the number of observations reduces to \( n-1 \) and the log–likelihood function becomes
\[
L \left( \frac{a}{y}, \frac{b}{y}, X \right) = \left[ \frac{1}{2} \right] \log 2\pi - \left[ \frac{1}{2} \right] \log \sigma^2 - \frac{\left( \frac{1}{2} - y^{(A)}_i - x^{(A)}_i \right) \left( \frac{1}{2} - y^{(A)}_i - x^{(A)}_i \right)}{2\sigma^2}
\]

\[
\sum_{i=2}^{n} \log y_i
\]

\[\text{... (4.7.16)}\]

Where \( Y^{(A)} = \left( Y_2^{(A)}, Y_3^{(A)}, \ldots, Y_n^{(A)} \right) \quad Y^{(A)}_i = \left( y^{(A)}_1, y^{(A)}_2, \ldots, y^{(A)}_{n-1} \right), \]

and \( X^{(A)} = \left[ x_2^{(A)}, x_3^{(A)}, \ldots, x_n^{(A)} \right]. \) Since the third term in the right hand side of (4.7.16) is rewritten as

\[
\left[ y^{(A)} - \left( y^{(A)}_1, x^{(A)} \right) \right] \left[ y^{(A)} - \left( y^{(A)}_1, x^{(A)} \right) \right] \left[ y^{(A)} - \left( y^{(A)}_1, x^{(A)} \right) \right] \left[ y^{(A)} - \left( y^{(A)}_1, x^{(A)} \right) \right] \frac{1}{2\sigma^2} \]

\[\text{... (4.7.17)}\]

The concentrated log-likelihood function is

\[
L_{\text{max}} = \left( \frac{a}{y}, X \right) = \left[ \frac{1}{2} \right] \log 2\pi - \left[ \frac{1}{2} \right] \log \sigma^2 + \left[ \frac{1}{2} \right] \log \hat{\sigma}^2(\lambda) + (\lambda - 1) \sum_{i=2}^{n} \log y_i,
\]

\[\text{... (4.7.18)}\]

where

\[
\hat{\sigma}^2(\lambda) = \frac{\left( Y - W \hat{\theta}(\lambda) \right) \left[ Y - W \hat{\theta}(\lambda) \right]}{(n-1)}
\]

\[\text{... (4.7.19)}\]

\[
\hat{\theta}(\lambda) = \left( W^T W \right)^{-1} W^T Y \quad \text{and} \quad W = \left( Y_1^{(A)}, X^{(A)} \right).
\]

Thus, as also seen in Anderson (1971), the ML estimate of \( \theta = (y, \beta) \) given \( \lambda \) is the least squares estimate derived from the following linear regression.
\[ Y_{t}^{(\lambda)} = \gamma Y_{t-1}^{(\lambda)} + x_{t}^{(\lambda)} \beta + u_{t}, \quad t = 2, 3, \ldots, n \quad (4.7.20) \]

The above fact means that if the first observation of the dependent variable is non-random, the usual ML estimation procedure on \((4.7.18)\) yields a correct ML estimate of \(\lambda\) even if the lagged dependent variable is included among explanatory variables.

4.8 THE BOX-COX DIFFERENCE TRANSFORMATION IN ECONOMETRICS

The Box-Cox transformation is well known in empirical econometrics and widely used with levels data to select the Correct functional form for a regression. The transformation of a variable \(Y_{t}\) is given by

\[ Y_{t}^{(\lambda)} = \frac{\left(Y_{t}^{(\lambda)} - 1\right)}{\lambda} \text{ for } \lambda \neq 0 \]
\[ \ln Y_{t} \text{ for } \lambda = 0 \quad (4.8.1) \]

Where \(\lambda = 1\) is the linear form and \(\lambda = 0\) is the logarithmic transformation. If \(Y_{t}\) is a time series, then the choice is not necessarily between \(Y_{t}\) and \(\ln Y_{t}\) but often is between \(\Delta Y_{t}\) and the percentage change in \(Y_{t}\). Layson and Seaks (1984) examined the time derivative of equation (4.8.1).
\[
\frac{dY_t^{(\lambda)}}{dt} = \frac{dY_t^{(\lambda)}}{dY_t} \frac{dY_t}{dt}
\]

\[
\frac{dY_t^{(\lambda)}}{dt} = Y_t^{\lambda-1} \left( \frac{dY_t}{dt} \right)
\]

\[\text{... (4.8.2)}\]

For discrete time periods an approximate to (4.8.2) is

\[
\Delta Y_t^{(\lambda)} = Y_{t-1}^{\lambda-1} \Delta Y_t
\]

\[\text{... (4.8.3)}\]

Where \( \Delta Y_t = Y_t - Y_{t-1} \). This is termed the Box–Cox difference (BCD) transformation. It yields the first difference of the variables when \( \lambda = 1 \) or the percentage change when \( \lambda = 0 \).

The BCD transformation can be used to formulate the regression model

\[
\Delta Y_t^{(\lambda)} = \alpha + \beta \Delta X_t^{(\lambda)} + \epsilon_t,
\]

\[\text{... (4.8.4)}\]

Which involves first differences or percentage changes depending on the value of \( \lambda \). We can apply the method of maximum likelihood to estimate \( \lambda \) along with \( \alpha \) and \( \beta \), and thus the data can choose the form of the difference transformation.

Layson and Seaks (1990) considered the likelihood function under normality for the sample of \( n \) observations is
\[ L = -\left( \frac{n}{2} \right) \ln(2\pi) + \ln \sigma^2 - \frac{1}{2} \sigma^2 \left( \sum_{i=1}^{n} \epsilon_i^2 \right) + (\lambda - 1) \sum_{i=1}^{n} \ln Y_{i-1} \] ...

(4.8.5)

Where the leading terms are identical to the log likelihood from a least squares regression on the first differences and the last term is the Jacobian from the transformation.

Although not noted by Layson and Seaks (1990) the estimation of \( \lambda \) and the maximization of equation (4.8.5) can be simplified by using the same scaling device discussed by Spitzer (1982, 1984) for the levels form of the Box–Cox test.

Define the geometric mean of the lagged values of \( Y \) by

\[ g = \exp \left( \frac{1}{n} \sum_{i=1}^{n} \ln Y_{i-1} \right) \] ...

(4.8.6)

and then divide the values of \( Y_i \) by \( g \) to obtain \( Y_i^* = \frac{Y_i}{g} \). With this definition of \( Y_i^* \), the Jacobian term in the log likelihood function is zero. Thus it is very simple to maximize the log likelihood function since one need only subject \( Y_i^* \) and the regressors to BCD transformation and perform a no. of least squares regressions for
different values of $\lambda$ in order to maximize equation (4.8.5). This scaling device not only simplifies the computation of $\lambda$ but Spitzer (1982) shows that it is also an important safeguard if the standard errors and t-ratios for the linear coefficients are to be used. It can be shown that the coefficients $\alpha$ and $\beta$ of the scaled model must be multiplied by $g^{\lambda}$ to recover the linear coefficients of the original model.

Any ordinary regression program that permits transformation of the variables can easily obtain the ML estimate of $\lambda$ and the maximum of equation (4.8.5). Then LR tests can be performed to test hypothesis about $\lambda$. For example, an asymptotic test of the first difference model is provided by

$$
\chi^2_1 = 2\{L(\hat{\lambda}) - L(\lambda = 1)\} \quad \ldots \quad (4.8.7)
$$

Colston and Robins (1985) derived an asymptotically equivalent $\chi^2$ statistic for the LM test and point out that it only requires estimation under the null hypothesis and thus avoids the search over $\lambda$ for the required for the LR test. While the LR testing does not require any special software as noted above, it is true that it requires a series of regressions. The LM test requires running only
one artificial regression to obtain the statistic where the following
four derivatives define the regressors.

\[
\frac{\partial L}{\partial \alpha} = \frac{e_i}{\hat{\sigma}^2} 
\]  
\... (4.8.8)

\[
\frac{\partial L}{\partial \beta} = \frac{e_i \Delta X^{(4)}_i}{\hat{\sigma}^2} 
\]  
\(4.8.9\)

\[
\frac{\partial L}{\partial \sigma^2} = \frac{(e_i^2 - \hat{\sigma}^2)}{2\hat{\sigma}^4} 
\]  
\(4.8.10\)

\[
\frac{\partial L}{\partial \lambda} = \ln Y_{i-1} - \left( \frac{e_i}{\hat{\sigma}^2} \right) (\Delta Y^{(4)}_i \ln Y_{i-1} - \hat{\beta} \Delta X^{(4)}_i \ln X_{i-1}) 
\]  
\... (4.8.11)

Thus the LM form of the difference test is given by \(X^2_1 = nR^2\).

Where \(R^2\) is the un-centered \(R^2 = 1 - \frac{\sum e_i^2}{\sum Y_i^2}\) from a regression of a
vector of ones on the four artificial series in equation (4.8.8) to
(4.8.11). The LM test is computed as

\[
X^2_1 = n - \sum e_i^2 
\]  
\... (4.8.12)

Since \(\sum Y_i^2\) is just \(n\) when each \(n\) value is one. Conlson and
Robins (1985) computed this for an illustrative model used by Layson
and Seaks (1990) and shown that it yields results fairly close to those
of the LR test.
4.9 STEIN-LIKE ESTIMATION IN LINEAR REGRESSION USING THE BOX-COX TRANSFORMATION

The general transformation regression model Introduced by Minbokim and R. Carter Hill (1995) is given by

\[ T(Y_t) = \beta_1 + \beta_2 x_{it}^{(4)} + \ldots + \beta_k x_{it}^{(4k)} + \epsilon_t \] \hspace{1cm} (4.9.1)

or

\[ T(Y_t) = u_t + \epsilon_t \] \hspace{1cm} (4.9.2)

\[ \epsilon_t \sim iid N(0, \sigma^2), \hspace{1cm} t = 1, 2, \ldots, T, \]

Where,

\( T(Y_t) \) is a suitable transformation that makes the random Variables

\( \epsilon_t \), normally distributed \( x_{it}^{(4)} \) is a Box-Cox transformation

\[ x_{it}^{(4)} = \begin{cases} \left( \frac{x_{it}^{(4k)} - 1}{\lambda} \right), & \text{if } \lambda_i \neq 0 \\ \ln(x_{it}), & \text{if } \lambda_i = 0 \end{cases} \] \hspace{1cm} (4.9.3)

and

\( x_{it}, \ i = 1, 2, \ldots, k \), are assumed to be non stochastic.

When the Box-Cox transformation is applied to the dependent Variable, the pre transformed dependent variable \( Y_t \) must be truncated (\( Y_t > 0 \)) in order for the Box-Cox transformation to be well-defined.
Thus the transformed variable $Y_t^{(4)}$ is also truncated.

$$T(Y_t) = Y_t^{(4)} \quad \text{if } L < Y_t^{(4)} < R,$$

$$\ldots \ (4.9.4)$$

Where

$$R = -\frac{1}{\lambda_i} \quad \text{and} \quad L = -\alpha \quad \text{if } \lambda_i < 0; \quad R = +\alpha \quad \text{and}$$

$$L = -\frac{1}{\lambda_i} \quad \text{if } \lambda_i > 0. \quad \text{Let} \quad \theta = (\beta_1, \ldots, \beta_k, \lambda_1, \ldots, \lambda_k, \sigma^2) \text{ and}$$

$$K = \dim(\theta). \quad \text{The regular log likelihood function of the Box-Cox model is given by}$$

$$l(\theta) = -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{T} \varepsilon_i^2 - \sum_{i=1}^{T} \ln(\theta(R_i) - \phi(L_i)) + (\lambda_i - 1) \sum_{i=1}^{T} \ln Y_i,$$

$$\ldots \ (4.9.5)$$

Were

$$R_i = \frac{(R - \mu_i)}{\sigma}, \quad L_i = \frac{(L - \mu_i)}{\sigma}, \quad \phi(.) \text{ is the distribution function}$$

of the standard normal variate, and $\varepsilon_i = Y_t^{(4)} - \mu_i.$

The MLE $\hat{\theta}$ is obtained by solving the first order conditions for log-likelihood function (4.9.5).

Consider the linear restriction

$$R\theta = r \quad \ldots \ (4.9.6)$$
Where,

\( R \) is a non stochastic \( J \times K \) matrix of rank \( J \leq K \).

The constrained MLE \( \hat{\theta} \) is obtained by maximizing the Lagrangian function

\[
M(\theta, \mu) = l(\theta) - \mu'(R\theta - r), \quad \ldots (4.9.7)
\]

Where,

\( \mu \) is a \( J \times 1 \) Lagrange multiplier vector. Then the pretest estimator \( \hat{\theta}^* \) is defined as

\[
\hat{\theta}^* = \begin{cases} 
\hat{\theta} & \text{if } u \leq C_\alpha, \\
\hat{\theta} & \text{if } u > C_\alpha 
\end{cases} \quad \ldots (4.9.8)
\]

Where,

\( u \) is an asymptotic \( \chi^2_{(J)} \) test statistic.

Using indicator functions, the pretest estimator can be written

\[
\hat{\theta}^* = I_{[0,C_\alpha]}(u)\hat{\theta} + I_{(C_\alpha,\infty)}(u)\hat{\theta}. \quad \ldots (4.9.9)
\]

Finally, the Shrinkage estimator we consider is

\[
\hat{\theta}^* = I_{(a,\infty)}(l)\left(1 - \frac{a}{\mu}\right)(\hat{\theta} - \bar{\theta}) + \bar{\theta}, \quad \ldots (4.9.10)
\]

Where \( a \) is constant that controls the amount of shrinkage of the unrestricted estimator toward the restricted estimator.
Maasoumi (1978) proposed a modified Stein–like estimator for the reduced form of a simultaneous equations model (MSRF) similar to $\theta^+$. In the MSRF $\hat{\theta}$ is the OLS estimator of the reduced form, $\tilde{\theta}$ is the Three SLS reduced form estimator and the test statistic $u$ has an asymptotic $\chi^2$ distribution with degrees of freedom equal to the degree of over identification. Maasoumi restricted the value of the constant $a \leq C_a$ and did not consider the risk properties of the modified Stein–like reduced form estimator.

As the specification error in the restriction increases, the weight on the MLE increases. Setting $W = 1 - I_{(a,a)}(u)\left(1 - \frac{a}{\mu}\right)$, the Shrinkage estimator $\theta^+$ can be written as a linear combination of the MLE and the constrained MLE

$$\theta^+ = W \tilde{\theta} + (1 - W) \hat{\theta}. \quad \ldots \ (4.9.11)$$

When the test statistic $u$ is less them the value of $a$, the Shrinkage estimator $\theta^+$ is reduced to the constrained MLE $\tilde{\theta}$, while the linear combination of the MLE and the constrained MLE will be obtained if the statistic $u$ exceeds the Value of $a$. The Shrinkage estimator has the property of shrinking every element of the MLE $\hat{\theta}$ toward the corresponding element of the constrained MLE $\tilde{\theta}$.

Therefore, the Stein–like estimator may have significant risk gain if the constrained MLE $\tilde{\theta}$ is close to the true parameter vector $\hat{\theta}$. 
4.10 **ALTERNATIVE ESTIMATORS FOR BOX-COX TRANSFORMATION**

Let \( Y_1, Y_2, \ldots, Y_n \) be a random sample of size \( n \) from a population whose functional form is unknown. Box and Cox (1964) suggested that if the transformation

\[
X = \begin{cases} 
Y^{\lambda} - 1, & \lambda \neq 0 \\
\ln(Y), & \lambda = 0
\end{cases} \quad \ldots \quad (4.10.1)
\]

is performed on the data then \( X \) will have approximately normal distribution with mean \( \mu \) and Variance \( \sigma^2 \). In Equation (4.10.1), \( \lambda \) is unknown considered as the Box – Cox power transformation parameter and ‘\( \ln \)’ represents the natural logarithm.

I. **Normal Likelihood Estimator**

After applying the transformation mentioned in equation (4.10.1) the density function of the data can be written as

\[
f(Y, \lambda, \mu, \sigma) = \sqrt{2\pi\sigma} \exp\left(-\frac{1}{2\sigma^2}\left(\frac{Y^{\lambda} - 1}{\lambda^{\lambda-1}}\right)^2\right) Y^{\lambda-1}, \lambda \neq 0 \\
1 \quad \left\{ \begin{array}{ll}
\frac{1}{2\sigma^2} \left(\ln(Y) - \mu\right)^2 & \lambda = 0 \\
\frac{1}{Y} & \lambda = 0
\end{array} \right. \quad \ldots \quad (4.10.2)
\]
The log-likelihood function \( l_L = l(\lambda, \mu, \sigma; Y_1, Y_2, \ldots, Y_n) \), where the subscript 'L' stands for the normal-likelihood, can be written as

\[
l_L = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \left( \frac{Y_i^2 - 1}{\lambda} - \mu \right)^2 + (\lambda - 1) \sum_{i=1}^{n} \ln(Y_i), \quad \lambda \neq 0
\]
\[
= \frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (\ln(Y_i) - \mu)^2 - \sum_{i=1}^{n} \ln(Y_i), \quad \lambda = 0
\]

... (4.10.3)

The equation (4.10.3) I maximized when the partial derivatives of (4.10.3) w.r.t to \( \mu, \sigma, \) and \( \lambda \) are equated to zero and solved simultaneously, which leads to solving.

\[
\left. \frac{\partial}{\partial \lambda} \left( \frac{Y_i^2 - 1}{\lambda} - \mu_L \right) \right|_{\lambda = \hat{\lambda}_L} = \frac{\hat{\sigma}^2}{\lambda} \sum_{i=1}^{n} \ln(Y_i).
\]

\[
\hat{\mu}_L = \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i^2 \ln(Y_i)}{\hat{\lambda}_L} - 1
\]  

(4.10.4)

and

\[
\hat{\sigma}_L = \left| \frac{1}{n} \sum_{i=1}^{n} \left( \frac{Y_i^2 - 1}{\hat{\lambda}_L} - \hat{\mu}_L \right) \right|
\]  

... (4.10.5)
iteratively for $\hat{\lambda}_2 \neq 0$. For $\lambda = 0$, the system of equation (4.10.4) and (4.10.5) become

$$\hat{\mu}_L = \frac{1}{n} \sum_{i=1}^{n} \ln(Y_i)$$

and

$$\hat{\sigma}_L = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\ln Y_i - \hat{\mu}_L)^2}$$

Where $\hat{\mu}_L, \hat{\sigma}_L,$ and $\hat{\lambda}_L$ are the corresponding maximum likelihood estimates of $\mu, \sigma$ and $\lambda$ respectively. In the process of maximizing the log-likelihood $l_L$, to incorporate $\lambda = 0$ in the parameter space, in each iteration, the expression (4.10.3) needs to be evaluated. Note that $\hat{\mu}_L, \hat{\sigma}_L,$ and $\hat{\lambda}_L$ are asymptotically normal, consistent and asymptotically unbiased estimates.

II Artificial Regression Model Estimator

Linearizing the trend between the observations and the normal percentiles in a normal probability plot justifies the normality assumption. On the other hand, the problem of observing large variances of the point estimates, of the model parameters, is reduced through augmenting the model by adding some covariates that are related to the response variable. Putting these two concepts together, Halawa (1996) proposed an artificial augmenting regression model as
\[ X_{(i)} = \mu + \sigma Z_{(i)} + \epsilon_{(i)}, \quad i = 1, 2, \ldots, n, \]

Where \( X_{(i)} \) is the \( i \)th smallest transformed observation, using the transformation (4.10.1) for a fixed \( \lambda \), \( Z_{(i)} = \phi^{-1} \left( \frac{i - 0.375}{n + 0.25} \right) \), \( \phi \) being the cumulative standard normal distribution function, \( \epsilon_{(i)} \) are assumed to be normal variates with zero means, unit variances and zero covariances. It follows by definition that \( \sum_{i=1}^{n} Z_{(i)} = 0 \) and

\[ L_{T} \sum_{i=1}^{n} Z_{(i)}^2 = 1. \]

Then the log-likelihood

\[ l_{R} = l(\lambda, \mu, \sigma; Y_{1}, Y_{2}, \ldots, Y_{n}), \]

Where the subscript \( R \) denotes the artificial regression model procedure, can be written as

\[ l_{R} = \frac{n}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^{n} \left( \frac{Y_{(i)} - 1}{\lambda} - \mu - \sigma Z_{(i)} \right)^2 + (\lambda - 1) \sum_{i=1}^{n} \ln(Y_{i}), \lambda \neq 0 \]

\[ -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^{n} (\ln(Y_{i}) - \mu - \sigma Z_{(i)})^2 - \sum_{i=1}^{n} \ln(Y_{i}), \quad \lambda = 0 \]

\[ \ldots (4.10.6) \]

The equation (4.10.3) is maximized when the partial derivatives of (4.10.6) with respect to \( \mu, \sigma \) and \( \lambda \) are equated to zero and solved simultaneously in which \( \sum_{i=1}^{n} Z_{(i)} = 0 \) and \( \sum_{i=1}^{n} Z_{(i)}^2 = 1 \) are substituted, leads to solving,
\sum_{i=1}^{n} \ln(Y_i) = \sum_{i=1}^{n} \frac{Y_i^{\hat{\lambda}_R} - 1}{\hat{\lambda}_R} - \hat{\mu}_R - \hat{\sigma}_R Z(\hat{\lambda}_R) \frac{Y_i^{\hat{\lambda}_R} - 1}{\hat{\lambda}_R} \ln(Y(\hat{\lambda}_R)) - \frac{Y_i^{\hat{\lambda}_R} - 1}{\hat{\lambda}_R}

\hat{\mu}_R = \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i^{\hat{\lambda}_R} - 1}{\hat{\lambda}_R} \quad (4.10.7)

and

\hat{\sigma}_R = \frac{1}{n} \sum_{i=1}^{n} Z(\hat{\lambda}_R) \frac{Y_i^{\hat{\lambda}_R} - 1}{\hat{\lambda}_R} \quad (4.10.8)

iteratively for \( \hat{\lambda}_R \neq 0 \). In the process of maximizing the log-likelihood \( l_R \), to incorporate \( \lambda = 0 \) in the parameter space, in each iteration, the expression (4.10.6) needs to be evaluated. For \( \lambda = 0 \), the system of equations (4.10.8) becomes

\[ \hat{\mu}_R = \frac{1}{n} \sum_{i=1}^{n} \ln(Y_i) \]

and

\[ \hat{\sigma}_R = \frac{1}{n} \sum_{i=1}^{n} Z(\hat{\lambda}_R) \ln(Y(\hat{\lambda}_R)) \]

The asymptotic properties of \( \hat{\mu}_R, \hat{\sigma}_R \) and \( \hat{\lambda}_R \) are given in Halawa (1996).

III Shapiro–Wilk W Statistic Estimator

The Shapiro–Wilk W test statistic (1965) is obtained by dividing the square of appropriate linear combination of the sample order statistics by the usual symmetric estimate of the variance.
Let \((x_1, x_2, \ldots, x_n)\) be a random sample to be tested for normality ordered \(x_{(1)} < x_{(2)} < \ldots < x_{(n)}\), and Let \((m_1, m_2, \ldots, m_n)\) denote the vector of expected values of standard normal order statistics.

Define

\[
W = \frac{\sum_{i=1}^{n} m_i x_{(i)}}{\sum_{i=1}^{n} m_i^2 \sum_{i=1}^{n} (x_i - \bar{x})^2} \quad \ldots (4.10.9)
\]

Note that \(W\) equals the square of the standard product moment correlation coefficient between the \(x_{(i)}\) and \(m_i\), and therefore measures the straightens of the -normal probability plot of the \(x_{(i)}\). Small values of \(W\) indicate non-normality. The ratio \(W\) is both scale and origin in variant and have the statistic is appropriate for a test of the composite hypothesis of normality. One useful feature of the Shapiro–Wilk test is that several independent goodness of fit tests may be combined in to one over all test of normality. \(W\) can be written in another form, that is

\[
W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \quad \ldots (4.10.10)
\]

Where

\[
a_i = \frac{m_i}{\sum_{i=1}^{n} m_i^2} \quad \ldots (4.10.11)
\]

The value of \(W\) is closer to 1 means the data is closer to normality and the maximum value of \(W\) is 1. By utilizing this fact, an
estimation procedure for $\lambda$ in the Box–Cox transformation (4.10.1) is proposed, which maximizes $W$.

That is, maximize

$$W = \frac{\left( \sum_{i=1}^{n} a_i x_{(i)} \right)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \quad \text{... (4.10.12)}$$

Where

$$x_{(i)} = \begin{cases} 
\frac{y_i^\lambda - 1}{\lambda}, & \lambda \neq 0 \\
\ln(y_i), & \lambda = 0
\end{cases} \quad \text{... (4.10.13)}$$

Note that $x_i$'s are ordered for a fixed $\lambda$. This maximization can be done either by using the Newton–Raphson root finding procedure after taking the first iterative of $W$ with respect to $\lambda$ and equating to zero or by the Numerical grid search procedure. The explicit solutions are not presented because the expressions become tedious when it is taken in to consideration that the relation between $x$ and $y$ are not always monotonous for all $\lambda$ values. And the statistic $W$ depends on the ordered $x$'s.

Let us denote such estimate of $\lambda$ as $\hat{\lambda}_W$.

Shapiro and Wilk's (1965) $W$ statistic has been shown to be a powerful test of normality for a variety of non-normal distributions. The values of $a_i$'s are tabulated in Shapiro and Wilk (1965) for
\( n = 2(1)50 \). But for other sample sizes, \( a_i \)'s can be estimated using the following approximate results.

By definition,

\[
a = \frac{m|V^{-1}|}{(m|V^{-1}V^{-1}m|^{1/2})} \quad m^* r^{-1} C \quad \ldots (4.10.14)
\]

is such that \( a^t a = 1 \), \( m \) is the vector of the expected values of standard normal order statistics and \( V \) is the covariance matrix of the standard normal order statistics. Let \( a^d = m^* r^{-1} \) then \( C^2 = a^d a^* \).

Suggested approximations are \( \hat{a}_{i}^2 = 2m, i = 1, 2, \ldots, n \).

and

\[
\hat{a}_{i}^2 = \hat{a}_{*}^2 = \begin{cases} 
\frac{1}{2} \left( \frac{1}{2} (n+1) \right)^{1/n}, & n \leq 20 \\
\sqrt{2\Gamma \left( \frac{1}{2} (n+1) \right) / \Gamma \left( \frac{1}{2} (n+1) \right)}, & n > 20 
\end{cases} \quad \ldots (4.10.15)
\]

The values of \( m_i \)'s are tabulated in Harter (1961) for \( n = (1)100 \), using simulation, it has been shown that the maximization of Shapiro–Wilk \( W \) statistic procedure leads to the estimate of \( \lambda \) which forces the data to be closer to normal in the sense that of higher Shapiro–Wilk \( W \) statistic value than that of the normal likelihood procedure and the artificial regression model procedure, \( \hat{\lambda}_w \) has smaller bias than that of and \( \hat{\lambda}_r \) and \( \hat{\lambda}_L \).