CRT-510

LEVEL: The L-1.4 program for BC-GAUHESEQ regressions – Box-Cox Generalized AUtoregressive HEteroskedastic Single EQuation models

by

Tran Liem
Marcel Dagenais
Marc Gaudry

Centre de recherche sur les transports, Université de Montréal, C.P. 6128, succursale A, Montréal, Canada H3C 3J7

This work was supported at various stages by Transport Canada, by the F.C.A.R. program and the M.E.S.S.T. of Quebec, and by the S.S.H.R.C. and the N.S.E.R.C. of Canada.

Cette étude a été publiée grâce à une subvention du fonds F.C.A.R. pour l'aide et le soutien à la recherche.

Centre de recherche sur les transports — Publication #510
Département des sciences économiques — Cahier #8714
June 1980
Revised – January 1987, June 1991, April 1993

ABSTRACT

L-1.4 is a program designed to deal with the specification of the functional form in the generalized single-equation regression model when the functional form of the heteroskedasticity of the residuals,which may also be autocorrelated, is fully analysed. Fixed or estimated parameters in the Box-Cox transformations can be associated with the dependent and groups of independent variables, which may be simultaneously modified by a chosen autocorrelation structure in which the autoregressive parameters are estimated, and by a selected functional form of the heteroskedasticity in which the different parameters can be either fixed or estimated. The asymptotic covariance matrix of all the parameter estimates of the system is computed. A number of useful outputs, including elasticities of the expected value and standard error of the dependent variable, as well as forecasts made by simulation and maximum likelihood techniques and their elaborate statistics, can also be obtained.

Keywords: Box-Cox Regression, Generalized Heteroskedasticity, Multiple order Autocorrelation, Maximum Likelihood Estimation, Time series, Cross-sections, Elasticities, Maximum Likelihood Forecast.

RÉSUMÉ

Le programme L-1.4 est destiné à l'étude de la forme fonctionnelle du modèle de régression multiple lorsque la forme de l'hétéroscédasticité des erreurs, qui peuvent aussi être autocorrelées, doit aussi être analysée en profondeur. Le programme permet de fixer ou d'estimer les paramètres dans les transformations Box-Cox appliquées à la variable dépendante et aux groupes de variables indépendantes, d'estimer conjointement les paramètres d'une structure d'autocorrélation choisie et finalement de fixer ou d'estimer les différents paramètres associés à la forme de l'hétéroscédasticité des erreurs. La matrice de covariance asymptotique des estimations de tous ces paramètres est calculée. La procédure permet aussi de faire un certain nombre de calculs utiles, dont celui des élasticités de l'espérance mathématique et de l'écart-type de la variable dépendante, ainsi que celui de prévisions obtenues par des méthodes usuelles de simulation et par une nouvelle méthode du maximum de vraisemblance.

Mots-clés: Régression avec Box-Cox, Hétéroscédasticité généralisée, Autocorrélation d'ordre multiple, Estimation du Maximum de Vraisemblance, Séries chronologiques, Coupes transversales, Elasticités, Prévision par Maximisation de la vraisemblance.

OTHER TRIO DOCUMENTATION AND ORDER FORM . . . . See last two pages

Contents

List of Tables

INTRODUCTION

In applied regression analysis, the most important aspect in model specification is the choice of the functional forms of the dependent and independent variables. As Zarembka (1968) has noted, economic theory rarely indicates the appropriate forms under which the variables should appear, except for the signs of the regression coefficients which are expected to be positive or negative according to the assumptions made on the economic behavior of the dependent variable with respect to the changes of the explanatory variables. The three classical forms often encountered in econometric studies are the linear, semilog and log-linear forms due to their computational ease with a standard regression computer package. One way of letting the data determine the most appropriate functional form is the use of a class of power transformations considered by Box and Cox (1964). The main advantage of this approach is that statistical tests can be performed on the Box-Cox parameters to discriminate the estimated functional forms against the classical forms which all appear as special cases of the Box-Cox transformation. Early applications of this transformation can be found in various fields of economic analysis: monetary economics [Zarembka (1968), White (1972), Spitzer (1976, 1977)], income analysis [Heckman and Polachek (1974), Welland (1976)], production theory [Appelbaum (1979), Berndt and Khaled (1979)], and transportation [Kau and Sirmans (1976), Hollyer et al. (1979)].

The assumption originally made by Box and Cox (1964) that the transformation, in addition to its main purpose of choosing the appropriate functional form, also renders the distribution of the transformed dependent variable nearly normal and homoskedastic may be unrealistic, since in the case where the residuals are heteroskedastic, Zarembka (1974) has shown that the estimated Box-Cox parameter on the dependent variable will be biased due to the effect needed to make the transformed dependent variable more nearly homoskedastic. To deal with this problem, models which estimate the flexible functional form by allowing for the simultaneous correction of heteroskedasticity have been proposed by Gaudry and Dagenais (1979) and Egy and Lahiri (1979). Another important case is the problem of autocorrelation of residuals which usually exists with time-series data: Savin and White (1978) have considered the simultaneous estimation of the functional form and the first order autocorrelation, and Gaudry and Wills (1978) have extended the approach to multiple order of autocorrelation.

Generalizing the procedure for a first autocorrelation order outlined in Gaudry and Dagenais (1979) to incorporate a higher autocorrelation order simultaneously in the estimation of the functional form and heteroskedasticity, this computer program performs for a single regression equation the maximum likelihood estimation of the functional forms of the dependent and independent variables with the Box-Cox Transformation (BCT), and the functional form of heteroskedasticity with the Inverse Box-Cox Transformation (IBCT) in which the variables used to explain the error variance are themselves subject to BCT. Further details of the general procedure for a higher autocorrelation order are also given in a working paper by Gaudry and Dagenais (1978).

1. STATISTICAL MODEL

1.1 Log-Likelihood Function

•Model

Following the approach considered above, for a sample of n observations the regression equation with a flexible functional form of the dependent and independent variables and a generalized structure of heteroskedasticity and autocorrelation of the residuals can be written as:

$$Y_t^{(\lambda_y)} = \sum_{k=1}^{K} \beta_k X_{kt}^{(\lambda_{xk})} + u_t, \quad (t = 1, \ldots, n) \tag{1}$$

$$u_t \quad = v_t f(Z_t)^{1/2}, \tag{2}$$

$$v_t \quad = \sum_{\ell=1}^{r} \rho_\ell v_{t-\ell} + w_t, \tag{3}$$

where

— the dependent variable Y_t and the independent variables X_{kt}'s are subject to the BCT which is defined as a power transformation with a parameter λ on any positive real variable V_t:

$$V_t^{(\lambda)} = \begin{cases} \left(V_t^\lambda - 1\right)/\lambda & \text{if } \lambda \neq 0, \\ \ln V_t & \text{if } \lambda = 0, \end{cases} \tag{4}$$

with the corresponding inverse function IBCT:

$$V_t^{(\lambda)^{-1}} = \begin{cases} \left[\lambda V_t^{(\lambda)} + 1\right]^{1/\lambda} & \text{if } \lambda \neq 0, \\ \exp\left(V_t\right) & \text{if } \lambda = 0, \end{cases} \tag{5}$$

where the expression in square brackets must be positive,

— the first-stage vector of residuals $u = \{u_t\}$ is assumed to be heteroskedastic with mean $E(u) = 0$ and covariance matrix $E\left(uu'\right) = \Omega =$ diag$(\omega_{11}, ..., \omega_{nn})$ where $\omega_{tt} = E\left(u_t^2\right) = E\left(v_t^2\right) f(Z_t)$ and$f(Z_t)$ is a function of a vector of M variables, $Z_t = (Z_{1t}, ..., Z_{Mt})$, used to explain the variance of u_t. Note that these variables can be chosen from the set of independent variables X_{kt}'s or totally exogenous,

— the second-stage vector of residuals $v = \{v_t\}$ is assumed to follow a stationary autoregressive process of order r, with mean $E(v) = 0$ and covariance matrix $E(vv') = \sigma_w^2 \Psi$, where $\sigma_w^2 = E\left(w_t^2\right)$ and Ψ is a symmetric matrix of order n, whose element can be expressed in terms of the ρ_ℓ's,

— and the third-stage vector of residuals $w = \{w_t\}$ is assumed to have a mean $E(w) = 0$ and a covariance matrix $E(ww') = \sigma_w^2 I_n$ where σ_w^2 is the variance of w_t and I_n is an identity matrix of order n.

In (1) it is clearly understood that some of the X_{kt}'s, such as the constant, the dummies and the ordinary variables not strictly positive, cannot be transformed by BCT.

In (2) the functional form of heteroskedasticity for $f(Z_t)$ is assumed to be an IBCT with a parameter λ_u applied to a linear combination of the variables Z_{mt}'s which are themselves subject to BCT:

$$f(Z_t) = \left\{ \lambda_u \left[\delta_0 + \sum_{m=1}^{M} \delta_m Z_{mt}^{(\lambda_{zm})} \right] + 1 \right\}^{1/\lambda_u} \tag{6}$$

Hence the variance of u_t associated with a typical diagonal element of Ω can be expressed as:

$$\begin{aligned} E\left(u_t^2\right) &= \omega_{tt} = \psi^2 f(Z_t) \\ &= \psi^2 \left\{ \lambda_u \left[\delta_0 + \sum_m \delta_m Z_{mt}^{(\lambda_{zm})} \right] + 1 \right\}^{1/\lambda_u} \end{aligned} \tag{7}$$

where the expression in curly brackets must be positive since it is associated with a variance. The constants δ_0 and ψ are necessary to preserve the invariance with respect to changes in measurement units of the Z_{mt}'s [Schlesselman (1971)].

The form (7) is quite general since it includes a great number of special cases encountered in empirical studies and fully discussed in Judge et al. (1985):

— Setting λ_u equal to zero yields a first form of heteroskedasticity:

$$\widetilde{\omega}_{tt} = \psi^2 exp\left[\delta_0 + \sum_m \delta_m Z_{mt}^{(\lambda_{zm})}\right] = \widetilde{\psi}^2 exp\left[\sum_m \delta_m Z_{mt}^{(\lambda_{zm})}\right] = \widetilde{\psi}^2 \tilde{f}(Z_t) \tag{8}$$

where $\widetilde{\psi}^2 = \psi^2 exp(\delta_0)$ and $\tilde{f}(Z_t) = exp\left[\sum_m \delta_m Z_{mt}^{(\lambda_{zm})}\right]$.

This form is commonly called “multiplicative" heteroskedasticity since $\tilde{f}(Z_t)$ can also be expressed as a product of M exponential functions, or equivalently the logarithm of the variance $\widetilde{\omega}_{tt}$ is a linear combination of the $Z_{mt}^{(\lambda_{zm})}$'s:

$$\widetilde{\omega}_{tt} = \widetilde{\psi}^2 \prod_m exp\left[\delta_m Z_{mt}^{(\lambda_{zm})}\right] \tag{9}$$

$$\Leftrightarrow \ln \widetilde{\omega}_{tt} = \ln \widetilde{\psi}^2 + \sum_m \delta_m Z_{mt}^{(\lambda_{zm})} = \tilde{\delta}_0 + \sum_m \delta_m Z_{mt}^{(\lambda_{zm})} \tag{10}$$

where $\tilde{\delta}_0 = \ln \widetilde{\psi}^2 = \ln \psi^2 + \delta_0$.

Harvey (1976) considered a special case of (9)-(10) where every λ_{zm} is set equal to one:

$$\widetilde{\omega}_{tt} = \widetilde{\psi}^2 \prod_m exp[\delta_m(Z_{mt} - 1)] = \widetilde{\widetilde{\psi}}^2 \prod_m exp(\delta_m Z_{mt}) \tag{11}$$

$$\Leftrightarrow \ln \widetilde{\omega}_{tt} = \tilde{\delta}_0 + \sum_m \delta_m (Z_{mt} - 1) = \tilde{\tilde{\delta}}_0 + \sum_m \delta_m Z_{mt} \tag{12}$$

where $\widetilde{\widetilde{\psi}}^2 = \widetilde{\psi}^2 \prod_m exp(-\delta_m)$ and $\tilde{\tilde{\delta}}_0 = \tilde{\delta}_0 - \sum_m \delta_m$.

Another special case of (9)-(10) can be obtained if every λ_{zm} is set equal to zero:

$$\widetilde{\omega}_{tt} = \widetilde{\psi}^2 \prod_m exp(\delta_m ln Z_{mt}) = \widetilde{\psi}^2 \prod_m Z_{mt}^{\delta m} \tag{13}$$

$$\Leftrightarrow \ln \widetilde{\omega}_{tt} = \ln \widetilde{\psi}^2 + \sum_m \delta_m \ln Z_{mt} = \tilde{\delta}_0 + \sum_m \delta_m \ln Z_{mt} \tag{14}$$

which was considered by Dagum and Dagum (1974). The univariate case (M=1) has been often used in practice, for example in Geary (1966), Park (1966), Kmenta (1971) or Glejser (1969) who has proposed in his test for heteroskedasticity specific values of δ_1 such as 2, 1, 1/2, –1/2 and –1 to give what he called “pure” heteroskedasticity.

— A second form of heteroskedasticity corresponds to the case where $\lambda_u \neq 0$. An important case found in the literature [Hildreth and Houck (1968), Theil (1971), Goldfeld and Quandt (1972), Froehlich (1973), Harvey (1974), Amemiya (1977)] corresponds to the special case where λ_u and every λ_{zm} are set equal to one, i.e. where the variance ω_{tt} is assumed to be a linear combination of the Z_{mt}'s:

$$\begin{aligned}
\bar{\omega}_{tt} &= \psi^2\left[\delta_0 + \sum_m \delta_m (Z_{mt} - 1) + 1\right] \\
&= \psi^2\left(\delta_0 - \sum_m \delta_m + 1\right) + \sum_m \psi^2 \delta_m Z_{mt} \\
&= \bar{\delta}_0 + \sum_m \bar{\delta}_m Z_{mt}
\end{aligned} \tag{15}$$

where $\bar{\delta}_0 = \psi^2\left(\delta_0 - \sum_m \delta_m + 1\right)$ and $\bar{\delta}_m = \psi^2 \delta_m$.

For the univariate case (M=1), Glejser (1969) considered the cases where $\lambda_u = 1$ and 2 with $\lambda_{z1} = 1$, and also the subcases which he called “mixed" heteroskedasticity corresponding to $\lambda_u = 1$ and 1/2 with $\lambda_{z1} = \delta_1 = 1$.

Due to the positivity constraint on the diagonal elements of Ω in the special form (15) as well as in the more general form (7) with $\lambda_u \neq 0$, it is hard to estimate the δ-coefficients and λ-parameters without violating the constraint, as experienced in preliminary tests with an earlier version of our computer program. Therefore, in this program version, we will only consider the first form of heteroskedasticity (8) with $\lambda_u = 0$ which still includes a great number of interesting cases.

•Likelihood Function

Before considering the likelihood function for the observed Y_t's, it is convenient to rewrite the model (1)-(2)-(3) into a more compact form where various expressions which will be frequently used throughout the manual will be defined.

The equation (3) for the residuals v_t's can be expressed as a function of the residuals u_t's given in (2):

$$\frac{u_t}{f(Z_t)^{1/2}} = \sum_\ell \rho_\ell \frac{u_{t-\ell}}{f(Z_{t-\ell})^{1/2}} + w_t \tag{16}$$

where $f(Z_t) \equiv \bar{f}(Z_t)$ following (8).

Replacing u_t which is derived from (1) as $Y_t^{(\lambda_y)} - \sum_k \beta_k X_{kt}^{(\lambda_{xk})}$ and $u_{t-\ell}$ by an analogous expression in $t - \ell$ yields:

$$\begin{aligned}
\frac{Y_t^{(\lambda_y)}}{f(Z_t)^{1/2}} - \sum_k \beta_k \frac{X_{kt}^{(\lambda_{xk})}}{f(Z_t)^{1/2}} &= \sum_\ell \rho_\ell \frac{Y_{t-\ell}^{(\lambda_y)}}{f(Z_{t-\ell})^{1/2}} - \sum_k \beta_k \sum_\ell \rho_\ell \frac{X_{k,t-\ell}^{(\lambda_{xk})}}{f(Z_{t-\ell})^{1/2}} + w_t \\
Y_t^* - \sum_k \beta_k X_{kt}^* \quad &= \sum_\ell \rho_\ell Y_{t-\ell}^* - \sum_k \beta_k \sum_\ell \rho_\ell X_{k,t-\ell}^* + w_t
\end{aligned} \tag{17}$$

where $Y_t^* = Y_t^{(\lambda_y)}/f(Z_t)^{1/2}$ and $X_{kt}^* = X_{kt}^{(\lambda_{xk})}/f(Z_t)^{1/2}$. The corresponding expressions for $t - \ell$ are obtained by replacing t by $t - \ell$.

The resulting form (17) can be more compactly rewritten as:

$$\begin{aligned}
Y_t^* - \sum_\ell \rho_\ell Y_{t-\ell}^* &= \sum_k \beta_k \left(X_{kt}^* - \sum_\ell \rho_\ell X_{k,t-\ell}^*\right) + w_t \\
Y_t^{**} &= \sum_k \beta_k X_{kt}^{**} + w_t
\end{aligned} \tag{18}$$

where $Y_t^{**} = Y_t^* - \sum_\ell \rho_\ell Y_{t-\ell}^*$ and $X_{kt}^{**} = X_{kt}^* - \sum_\ell \rho_\ell X_{k,t-\ell}^*$.

Assuming that the residuals w_t's are independently and normally distributed $N\left(0, \sigma_w^2\right)$ and dropping the first r observations to simplify the procedure for higher order autocorrelation, i.e. assuming that the first observations on Y_t are given, the likelihood function associated with the last $n-r$ observations on Y_t can be written as follows:

$$\mathcal{L} = \prod_{t=1+r}^{n} \frac{1}{\sqrt{2\pi\sigma_w^2}} exp\left(-\frac{w_t^2}{2\sigma_w^2}\right) \left|\frac{\partial w_t}{\partial Y_t}\right| \tag{19}$$

where the residual w_t is given by (18) as $Y_t^{**} - \sum_k \beta_k X_{kt}^{**}$, and the jacobian of the transformation from w_t to the observed Y_t is: $|\partial w_t / \partial Y_t| = Y_t^{\lambda_y - 1} / f(Z_t)^{1/2}$.

The corresponding log-likelihood function is:

$$L = -\frac{N}{2} \ln\left(2\pi\sigma_w^2\right) - \frac{1}{2\sigma_w^2} \sum_t w_t^2 - \frac{1}{2} \sum_t \ln f(Z_t) + (\lambda_y - 1) \sum_t \ln Y_t \tag{20}$$

where $N = n - r$ and the index t for the summation runs from $1 + r$ to n. Note that this function depends on all the parameters of the model: $\Pi = \left(\beta', \sigma_w^2, \lambda_y, \lambda_x', \lambda_z', \delta', \rho'\right)'$ where σ_w^2 and λ_y are scalars, and $\beta, \lambda_x, \lambda_z, \delta$ and ρ are the column vectors associated with the β_k's, λ_{xk}'s, λ_{zm}'s, δ_m's and ρ_ℓ's respectively.

•Concentrated Log-Likelihood Function

Since the model (18) rewritten in terms of the transformed variables Y_t^{**} and X_{kt}^{**}'s is just linear in the β-coefficients, we can concentrate the log-likelihood function on the β_k's and σ_w^2 by setting the first derivatives of the function with respect to these parameters to zero and solving for their values which will be replaced in (20).

In matrix notation, the compact form (18) can be expressed as:

$$Y^{**} = X^{**}\beta + w \tag{21}$$

where in view of (19) and (20), Y^{**} is a column vector containing the last N observations, X^{**} is an $(N \times K)$ matrix, β is a $(K \times 1)$ vector of coefficients and w is an $(N \times 1)$ vector of residuals.

Using (21) to replace $\sum_t w_t^2$ in the log-likelihood function (20) by $w'w$, the first derivatives of the function with respect to β and σ_w^2 are given by:

$$\begin{aligned} \frac{\partial L}{\partial \beta} &= -\frac{1}{2\sigma_w^2} \frac{\partial w'w}{\partial \beta} = -\frac{1}{\sigma_w^2} \frac{\partial (Y^{**} - X^{**}\beta)'}{\partial \beta} (Y^{**} - X^{**}\beta) \\ &= \frac{1}{\sigma_w^2} X^{**'} (Y^{**} - X^{**}\beta) = \frac{1}{\sigma_w^2} \left(X^{**'} Y^{**} - X^{**'} X^{**} \beta\right) \end{aligned} \tag{22}$$

$$\begin{aligned} \frac{\partial L}{\partial \sigma_w^2} &= -\frac{N}{2} \frac{1}{\sigma_w^2} + \frac{1}{2\sigma_w^4} w'w \\ &= \frac{1}{2\sigma_w^2} \left(-N + \frac{1}{\sigma_w^2} w'w\right) . \end{aligned} \tag{23}$$

By equating these two derivatives to zero and solving for β and σ_w^2, we obtain:

$$\hat{\beta} = \left(X^{**'} X^{**}\right)^{-1} X^{**'} Y^{**} \tag{24}$$

$$\hat{\sigma}_w^2 = \frac{1}{N} w' w = \frac{1}{N} (Y^{**} - X^{**}\beta)'(Y^{**} - X^{**}\beta) \tag{25}$$

Replacing β by $\hat{\beta}$ in (25) gives a value of σ_w^2 in function of $\hat{\beta}$:

$$\bar{\sigma}_w^2 = \frac{1}{N}\left(Y^{**} - X^{**}\hat{\beta}\right)'\left(Y^{**} - X^{**}\hat{\beta}\right). \tag{26}$$

Substitution of $\hat{\beta}$ and $\bar{\sigma}_w^2$ in L yields the concentrated log-likelihood function:

$$\bar{L} = -\frac{N}{2}[1 + \ln(2\pi)] - \frac{N}{2}\ln \bar{\sigma}_w^2 - \frac{1}{2}\sum_t \ln f(Z_t) + (\lambda_y - 1)\sum_t \ln Y_t \tag{27}$$

which now depends only on $\bar{\Pi} = \left(\lambda_y, \lambda_x', \lambda_z', \delta', \rho'\right)'$.

1.2 Computational Aspects

•Maximization Procedure

The Davidon-Fletcher-Powell (DFP) algorithm [Fletcher and Powell (1963)] is used to maximize the concentrated log-likelihood function $\bar{L}$ with respect to $\bar{\Pi} = \left(\lambda_y, \lambda_x', \lambda_z', \delta', \rho'\right)'$.

The gradient of $\bar{L}$ which is used in DFP can be written as:

$$\frac{\partial \bar{L}}{\partial \bar{\Pi}} = -\frac{1}{\bar{\sigma}_w^2}\sum_t w_t \frac{\partial w_t}{\partial \bar{\Pi}} - \frac{1}{2}\sum_t \frac{\partial \ln f(Z_t)}{\partial \bar{\Pi}} + \frac{\partial(\lambda_y - 1)}{\partial \bar{\Pi}}\sum_t \ln Y_t \tag{28}$$

where

$$\frac{\partial w_t}{\partial \lambda_y} = \frac{1}{f(Z_t)^{1/2}}\frac{\partial Y_t^{(\lambda_y)}}{\partial \lambda_y} - \sum_\ell \frac{\rho_\ell}{f(Z_{t-\ell})^{1/2}}\frac{\partial Y_{t-\ell}^{(\lambda_y)}}{\partial \lambda_y} \tag{29}$$

$$\frac{\partial w_t}{\partial \lambda_{xk}} = -\beta_k\left[\frac{1}{f(Z_t)^{1/2}}\frac{\partial X_{kt}^{(\lambda_{xk})}}{\partial \lambda_{xk}} - \sum_\ell \frac{\rho_\ell}{f(Z_{t-\ell})^{1/2}}\frac{\partial X_{k,t-\ell}^{(\lambda_{xk})}}{\partial \lambda_{xk}}\right] \tag{30}$$

$$\frac{\partial w_t}{\partial \lambda_{zm}} = -\frac{1}{2}\delta_m\left[\frac{u_t}{f(Z_t)^{1/2}}\frac{\partial Z_{mt}^{(\lambda_{zm})}}{\partial \lambda_{zm}} - \sum_\ell \frac{\rho_\ell u_{t-\ell}}{f(Z_{t-\ell})^{1/2}}\frac{\partial Z_{m,t-\ell}^{(\lambda_{zm})}}{\partial \lambda_{zm}}\right] \tag{31}$$

$$\frac{\partial w_t}{\partial \delta_m} = -\frac{1}{2}\left[\frac{u_t}{f(Z_t)^{1/2}}Z_{mt}^{(\lambda_{zm})} - \sum_\ell \frac{\rho_\ell u_{t-\ell}}{f(Z_{t-\ell})^{1/2}}Z_{m,t-\ell}^{(\lambda_{zm})}\right] \tag{32}$$

$$\frac{\partial w_t}{\partial \rho_\ell} = -\frac{u_{t-\ell}}{f(Z_{t-\ell})^{1/2}} \tag{33}$$

$$\frac{\partial \ln f(Z_t)}{\partial \bar{\Pi}_i} = \begin{cases} 0 & \text{if } \bar{\Pi}_i \neq \lambda_{zm} \text{ and } \delta_m, \\ \delta_m \dfrac{\partial Z_{mt}^{(\lambda_{zm})}}{\partial \lambda_{zm}} & \text{if } \bar{\Pi}_i = \lambda_{zm}, \\ Z_{mt}^{(\lambda_{zm})} & \text{if } \bar{\Pi}_i = \delta_m, \end{cases} \tag{34}$$

$$\frac{\partial (\lambda_y - 1)}{\partial \bar{\Pi}_i} = \begin{cases} 0 & \text{if } \bar{\Pi}_i \neq \lambda_y, \\ 1 & \text{if } \bar{\Pi}_i = \lambda_y. \end{cases} \tag{35}$$

The derivatives $\partial Y_t^{(\lambda_y)}/\partial \lambda_y$ in (29), $\partial X_{kt}^{(\lambda_{xk})}/\partial \lambda_{xk}$ in (30) and $\partial Z_{mt}^{(\lambda_{zm})}/\partial \lambda_{zm}$ in (31) and (34) as well as their lagged expressions are computed by the generic formula:

$$\frac{\partial V_t^{(\lambda)}}{\partial \lambda} = \begin{cases} \frac{1}{\lambda}\left[V_t^{\lambda} \ln V_t - V_t^{(\lambda)}\right] & \text{if } \lambda \neq 0, \\ \frac{1}{2} \ln^2 V_t & \text{if } \lambda = 0. \end{cases} \tag{36}$$

•Asymptotic Covariance Matrix of the Parameter Estimates

At the maximum point of $\bar{L}$, hence L, the asymptotic covariance matrix of all the parameter estimates $\widehat{\Pi}$ is evaluated by the method of Berndt et al. (1974). The log-likelihood function L in (20) can be rewritten as a sum of N individual log-likelihood functions associated with each observation t:

$$L = \sum_t L_t = \sum_t \left[-\frac{1}{2} \ln\left(2\pi\sigma_w^2\right) - \frac{1}{2\sigma_w^2} w_t^2 - \frac{1}{2} \ln f(Z_t) + (\lambda_y - 1) \ln Y_t \right] \tag{37}$$

where L_t is the log-likelihood function corresponding to the observation t and is equal to the expression in square brackets.

The asymptotic covariance matrix of $\widehat{\Pi}$ is given by:

$$\text{Var}\left(\widehat{\Pi}\right) = \left[\sum_t \frac{\partial L_t}{\partial \Pi} \frac{\partial L_t}{\partial \Pi'} \right]^{-1}_{\Pi = \widehat{\Pi}} \tag{38}$$

where the column vector $\partial L_t / \partial \Pi$ has the following form:

$$\frac{\partial L_t}{\partial \Pi} = -\frac{1}{2\sigma_w^2} \frac{\partial \sigma_w^2}{\partial \Pi} - \frac{1}{2} \frac{\partial \left(w_t^2/\sigma_w^2\right)}{\partial \Pi} - \frac{1}{2} \frac{\partial \ln f(Z_t)}{\partial \Pi} + \frac{\partial (\lambda_y - 1)}{\partial \Pi} \ln Y_t \tag{39}$$

with

$$\frac{\partial \sigma_w^2}{\partial \Pi_i} = \begin{cases} 0 & \text{if } \Pi_i \neq \sigma_w^2, \\ 1 & \text{if } \Pi_i = \sigma_w^2, \end{cases} \tag{40}$$

$$\frac{\partial \left(w_t^2/\sigma_w^2\right)}{\partial \Pi_i} = \begin{cases} \dfrac{2w_t}{\sigma_w^2} \dfrac{\partial w_t}{\partial \Pi_i} & \text{if } \Pi_i \neq \sigma_w^2, \\ -\dfrac{w_t^2}{\sigma_w^4} & \text{if } \Pi_i = \sigma_w^2, \end{cases} \tag{41}$$

$$\frac{\partial w_t}{\partial \Pi_i} = \begin{cases} \dfrac{\partial w_t}{\partial \bar{\Pi}_i} & \text{if } \Pi_i = \bar{\Pi}_i, \\ -X_{kt}^{**} & \text{if } \Pi_i = \beta_k \end{cases} \tag{42}$$

Note that the derivatives $\partial w_t/\partial\bar{\Pi}_i$ in (41) for the typical elements of $\bar{\Pi}$ are already given in (29)-(30)-(31)-(32)-(33). Since $f(Z_t)$ is a function of the λ_{zm}'s and δ_m's only, the derivative $\partial \ln f(Z_t)/\partial\Pi_i$ is also given in (34) for $\Pi_i = \bar{\Pi}_i = \lambda_{zm}$ and δ_m. Finally the derivative $\partial(\lambda_y - 1)/\partial\Pi_i$ is equal to zero if $\Pi_i \neq \lambda_y$ and one if $\Pi_i = \lambda_y$ as in (35).

•Scaling of the Variables

Although at the initial and final steps of the maximization of $\bar{L}$, all the outputs are given in terms of the original units of the variables Y_t, X_{kt} and Z_{mt}, an automatic scaling of these variables is performed during the iterations to avoid numerical problems which can slow down or inhibit the convergence process.

Each variable V_t is scaled as follows:

$$\widetilde{V}_t = g_v V_t \tag{43}$$

where V_t represents Y_t, X_{kt} or Z_{mt} in their original units, and g_v is the scaling factor of the form $10^{-\left[\log\left(\max\limits_t |V_t|\right)\right]}$ in which the square brackets indicate that the greatest integer is being taken for the expression inside.

In general, the β and δ-coefficients and their unconditional t-values computed from (38) are not invariant with respect to the scaling of Y_t, X_{kt} and Z_{mt}, but the $\lambda_y, \lambda_x, \lambda_z$ and ρ-parameters as well as their unconditional t-values remain invariant. Note that the concentrated log-likelihood function $\bar{L}$, hence L, and the error variance σ_w^2 are affected by the scaling of Y_t only. Therefore the concentrated log-likelihood values which are listed in terms of the scaled units of Y_t during the iterations would not be the same as if they were in the original units of Y_t, unless the scaling factor of Y_t happens to be equal to one for a particular data set. When a test for stability of the β-coefficients over various data sets is performed in terms of the log-likelihood values or the error variance estimates, these quantities must be in the original units of Y_t, or more generally in a system of units which is common to all data sets used.

•Constraints on the ρ-parameters

Strictly speaking, the autocorrelation parameters ρ_ℓ's in (3) must satisfy a very large number of conditions to ensure stationarity in an autoregressive process of order r. For simplicity, a constraint $-1 < \rho_\ell < 1$ on every ρ_ℓ is implicitly introduced by using the Fisher's z-transformation:

$$z_\ell = \frac{1}{2}\ln\frac{1+\rho_\ell}{1-\rho_\ell} \quad (-\infty < z_\ell < \infty) \tag{44}$$

with the corresponding inverse transformation $\rho_\ell = \tanh z_\ell$.

This parameter change from ρ_ℓ to z_ℓ implies that the maximization of $\bar{L}$ is performed with respect to the z-parameters. Therefore, during the iterations the z-values are listed instead of the ρ-values which are only given at the convergence of $\bar{L}$ when all the estimated parameters of the model $\widehat{\Pi}$ are printed with their standard-errors and t-statistics.

The derivative of $\bar{L}$ with respect to each z-parameter is computed as follows:

$$\frac{\partial\bar{L}}{\partial z_\ell} = \frac{\partial\bar{L}}{\partial\rho_\ell}\frac{\partial\rho_\ell}{\partial z_\ell} \tag{45}$$

where $\partial\bar{L}/\partial\rho_\ell$ is already given in (28) and (33), and $\partial\rho_\ell/\partial z_\ell = sech^2 z_\ell = 1 - \tanh^2 z_\ell = 1 - \rho_\ell^2$.

1.3 Model Types

Four types of Box-Cox regression models can be specified using the different options available in the program:

1. **BC**
   This model type only includes the Box-Cox (**BC**) transformation on the dependent and independent variables in the regression equation (1) without considering the problems of heteroskedasticity and autocorrelation for the residuals u_t's and v_t's respectively. The parameters involved are $\beta, \sigma_w^2, \lambda_y$ and λ_x where the variances of u_t, v_t and w_t are all identical;
2. **BC–HE**
   This model type extends the previous one to the case where the functional form of HEteroskedasticity (**HE**) of the first-stage residuals u_t's, $f(Z_t) = \exp\left[\sum_m \delta_m Z_{mt}^{(\lambda_{zm})}\right]$, is specified simultaneously with the BC transformation on the dependent and independent variables. The additional parameters to be considered are the λ_z and δ-vectors which are involved in $f(Z_t)$. Since the problem of autocorrelation of the residuals v_t's is not considered, the variance of w_t remains identical to the variance of v_t, but the latter is different from the variance of u_t;
3. **BC–GAU**
   This model type allows for a Generalized AUtoregressive (**GAU**) structure of the second-stage residuals v_t's, to be estimated jointly with the BC transformation on the dependent and independent variables, while the first-stage residuals u_t's are assumed to be homoskedastic. Only the ρ-parameters are added to the set of parameters already included in the BC model. Hence the variance of w_t is different from the variance of v_t, but the latter is identical to the variance of u_t;
4. **BC–GAUHE**
   This model type corresponds to the general case where the first-stage residuals u_t's are assumed to be heteroskedastic with a functional form $f(Z_t)$ and the second-stage residuals v_t's follow a stationary autoregressive process of order r. All the parameters of the model, $\Pi = \left(\beta', \sigma_w^2, \lambda_y, \lambda_x', \lambda_z', \delta', \rho'\right)'$, are simultaneously estimated. Therefore the variances of u_t, v_t and w_t are all different.

   The **BC–GAUHE** model type includes all the first three as special cases provided that:

   i) The same set of dependent and independent variables, Y and X, is used for the estimation of the parameters, with the same set of λ_y and λ_x-parameters being estimated or fixed across the four model types. For example, if λ_y or some of the λ_x-parameters are fixed at zero, then the same constraints should be retained across all the model types;
   ii) The same set of variables Z is specified in the functional form of heteroskedasticity for both BC-HE and BC–GAUHE, with the same set of λ_z and δ-parameters being estimated or fixed in both model types. For instance, if every λ_{zm} is set to one, then the same constraints must be used in both;
   iii) The same structure of the autoregressive parameters is estimated for both BC–GAU and BC–GAUHE so that the simpler model type is nested in the general model type. For example, if BC–GAU includes three estimated ρ-parameters of orders 1, 4 and 12, then the same structure should be preserved in BC–GAUHE.

Note that the BC–HE and BC–GAU model types are completely different since they are based on different assumptions on the structure of the residuals, hence they cannot be compared between each other, although the simpler BC model type is nested in both provided that the condition (i) is satisfied.

1.4 Model Estimation

Since the four model types presented above are by increasing degrees of nonlinearity in the parameters as the specification of a model under study is allowed to be more and more general by relaxing the constraints on the parameters of heteroskedasticity or autocorrelation or both with respect to the BC model type, a normal procedure to succeed in estimating all the four model types can be suggested as follows:

Step 1

Start the estimation of the model with the BC type. If the number of independent variables is too large to allow for a distinct BC transformation on each variable, regrouping all the variables which can be transformed by Box-Cox into a few main categories and specifying one λ_x-parameter for each category of variables will reduce the number of estimated λ_x-parameters.

The program includes two types of constraints on the λ_y and λ_x-parameters:

— λ_y and any λ_x-parameter can be fixed at a constant value instead of being estimated. This type of constraint permits specially the estimation of the classical functional forms such as the linear, semilog and loglinear forms: for example, setting λ_y equal to zero and every λ_x-parameter to one yields the semilog form;
— λ_y can be set equal to one of the estimated λ_x-parameter. This type of constraint is useful in particular for the case where the model is specified with only one estimated λ-parameter common to all the dependent and independent variables. This form can then be compared to the linear and loglinear forms considered above.

To improve the specification of the model, three special options are available:

— Detection of multicollinearity among the independent variables X with the correlation matrix and specially with the regression coefficient variance decomposition proportions based on the spectral decomposition of $X'X$ in terms of the original and Box-Cox transformed variables (Section 3.1);
— Graphical analysis of the estimated residuals u_t which are plotted by increasing values against a variable Z_m that is thought to explain the variance of u_t in the case of heteroskedasticity which can arise not only with the cross-section data but can also be induced by the BC transformation on the dependent variable even with the time-series data (Section 3.2);
— Statistical tests of the estimates of the autocorrelation function and the partial autocorrelation function at different lags of the residuals u_t's for large time-series samples in order to detect the most significant orders of autocorrelation (Section 3.3).

Step 2

Proceed with the estimation of the model according to the results of the analyses previously made:

— If the problem of severe multicollinearity among a group of independent variables exists, then it should be first treated before considering the problems of heteroskedasticity and autocorrelation: some of the variables must be dropped from the regression equation (1) or replaced by their proxies, i.e. the variables which are used to explain essentially the same phenomena as the original variables, but which are less collinear among themselves. The model with a new specification of the variables is reestimated following Step 1. The whole process is repeated until a satisfying set of independent variables is obtained.
— If only the problem of heteroskedasticity is the most serious, then different types of the "multiplicative" form of heteroskedasticity in (8) can be tried using the BC–HE model type since the program allows the λ_z-parameters and the δ-coefficients in $f(Z_t)$ to be estimated or fixed at a constant value. At the end of each estimation, the first special option can be used again to detect multicollinearity among the Box-Cox transformed independent variables corrected for heteroskedasticity X^*_{kt}, and for time-series data, the third special option can also be used to analyze the structure of autocorrelation of the estimated residuals v_t's. This will allow us to proceed further with the estimation of the model using the BC–GAUHE type if autocorrelation is present.
— If only the problem of autocorrelation is important, then using the BC–GAU model type is sufficient to improve the specification of the model.

Note that at the end of each estimation with the BC–GAU or BC–GAUHE model types where the correction for autocorrelation has been applied with a chosen set of estimated ρ-parameters, a further check on the estimates of the autocorrelation and partial autocorrelation functions will indicate whether the estimated residuals w_t behave approximately as a white noise or other autocorrelation orders remain to be corrected for.

2. ESTIMATION RESULTS

2.1 Expected Value of the Dependent Variable

Following our approach in which the dependent variable Y_t is assumed to be censored both downwards and upwards, $\varepsilon \leq Y_t \leq \nu$, where ε and ν are respectively the lower and upper censoring points common to all observations in the sample, the generalization of Tobin's model (1958) to a doubly censored dependent variable yields the following expression for the expected value of Y_t:

$$E(Y_t) = \varepsilon \int_{-\infty}^{w_t(\varepsilon)} \varphi(w)dw + \int_{w_t(\varepsilon)}^{w_t(\nu)} Y_t(w_t)\varphi(w)dw + \nu \int_{w_t(\nu)}^{\infty} \varphi(w)dw \tag{46}$$

where

— $\varphi(w)$ is the normal density function of w with zero mean and variance σ_w^2:

$$\varphi(w) = \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp\left(-\frac{w^2}{2\sigma_w^2}\right), \tag{47}$$

— $Y_t(w)$ is a function of w:

$$Y_t(w) = \begin{cases} \left[1 + \lambda_y f(Z_t)^{1/2}\left(\sum_\ell \rho_\ell Y^*_{t-\ell} + \sum_k \beta_k X^{**}_{kt} + w\right)\right]^{1/\lambda_y} & \text{if } \lambda_y \neq 0, \\ \exp\left[f(Z_t)^{1/2}\left(\sum_\ell \rho_\ell \frac{\ln Y_{t-\ell}}{f(Z_{t-\ell})^{1/2}} + \sum_k \beta_k X^{**}_{kt} + w\right)\right] & \text{if } \lambda_y = 0, \end{cases} \tag{48}$$

— and the generic formula for $w_t(\varepsilon)$ and $w_t(\nu)$ is:

$$w_t(\xi) = \begin{cases} \frac{\xi^{\lambda_y} - 1}{\lambda_y f(Z_t)^{1/2}} - \sum_\ell \rho_\ell Y^*_{t-\ell} - \sum_k \beta_k X^{**}_{kt} & \text{if } \lambda_y \neq 0, \\ \frac{\ln \xi}{f(Z_t)^{1/2}} - \sum_\ell \rho_\ell \frac{\ln Y_{t-\ell}}{f(Z_{t-\ell})^{1/2}} - \sum_k \beta_k X^{**}_{kt} & \text{if } \lambda_y = 0\,. \end{cases} \tag{49}$$

Depending on the sign of the BC parameter λ_y, three distinct cases can be derived for $E(Y_t)$:

• $\boxed{\boldsymbol{\lambda_y > 0}}$

If ε and ν tend towards 0 and ∞ respectively, following (49) the limits of $w_t(\varepsilon)$ and $w_t(\nu)$ can be obtained as:

$$w_t^* = \lim_{\varepsilon \to 0} w_t(\varepsilon) = -\frac{1}{\lambda_y f(Z_t)^{1/2}} - \sum_\ell \rho_\ell Y^*_{t-\ell} - \sum_k \beta_k X^{**}_{kt} \tag{50}$$

and $\lim_{\nu \to \infty} w_t(\nu) = \infty$.

Hence the first and last terms in (46) disappear and the expected value of Y_t reduces to the second term which can be written as:

$$E(Y_t) = \int_{w_t^*}^{\infty} Y_t(w)\varphi(w)dw. \tag{51}$$

•$\boxed{\boldsymbol{\lambda_y = 0}}$

If ε and ν tend towards 0 and ∞ respectively, then $\lim\limits_{\varepsilon\to 0} w_t(\varepsilon) = -\infty$ and $\lim\limits_{\nu\to 0} w_t(\nu) = \infty$. Hence the first and last terms in (46) disappear and the expected value of Y_t reduces to the second term which can be written as:

$$E(Y_t) = \int\limits_{-\infty}^{\infty} Y_t(w)\varphi(w)dw. \tag{52}$$

•$\boxed{\boldsymbol{\lambda_y < 0}}$

If ε tends towards 0, then $\lim\limits_{\varepsilon\to 0} w_t(\varepsilon) = -\infty$, and the first term disappears. Before taking the limit of $w_t(\nu)$ as ν tends towards ∞ in the last two terms of (46), the expected value of Y_t can be expressed as:

$$E(Y_t) = \int\limits_{-\infty}^{w_t(\nu)} Y_t(w)\varphi(w)dw + \nu \int\limits_{w_t(\nu)}^{\infty} \varphi(w)dw. \tag{53}$$

If ν tends towards ∞, the limit of $w_t(\nu)$ has a finite value:

$$\lim_{\nu\to\infty} w_t(\nu) = -\frac{1}{\lambda_y f(Z_t)^{1/2}} - \sum_\ell \rho_\ell Y^*_{t-\ell} - \sum_k \beta_k X^{**}_{kt}\ . \tag{54}$$

Hence $\lim\limits_{\nu\to\infty} \nu \int\limits_{w_t(\nu)}^{\infty} \varphi(w)dw$ does not exist and $E(Y_t)$ does not have a finite value. In this case, it is still possible to obtain an approximation of $E(Y_t)$ for every observation t provided that the user selects a large value of ν, say $\widetilde{\nu}$, such that the probability that Y_t exceeds this upper limit $\widetilde{\nu}$ will be negligible, i.e. given the values of Y^*_t, X^{**}_{kt} and Z_t in $w_t(\widetilde{\nu})$, the value of $E(Y_t)$ will be much smaller than $\widetilde{\nu}$. The value of ν can be chosen very approximately without affecting the numerical precision of $E(Y_t)$. By default, the program will generate a value of ν determined as follows:

$$\widetilde{\nu} = \min\left[\max_t Y_t + 10\sigma_y, (-0.8\sigma_w\lambda_y)^{1/\lambda_y}\right] \tag{55}$$

where σ_y is the sample standard error of Y_t and σ_w and λ_y are evaluated at their estimated values. Note that if λ_y is fixed at a certain value in an estimation run, then this value will be used.

For the three cases of λ_y, the integrals corresponding to the second term in the general expression of $E(Y_t)$ are numerically computed using the Gauss-Legendre 32-point quadrature. As a sample measure of the degree of numerical approximation of $E(Y_t)$ for the two cases $\lambda_y > 0$ and $\lambda_y < 0$, the mean probability of Y_t in the sample to be at the lower limit ε if $\lambda_y > 0$ or at the upper limit ν if $\lambda_y < 0$ is also computed:

$$\overline{Pr}(Y_t \le \varepsilon) = \frac{1}{N}\sum_t \int\limits_{-\infty}^{w^*_t} \varphi(w)dw \quad (\lambda_y > 0) \tag{56}$$

and

$$\overline{Pr}(Y_t \ge \nu) = \frac{1}{N}\sum_t \int\limits_{w_t(\nu)}^{\infty} \varphi(w)dw \ \ (\lambda_y < 0) \tag{57}$$

where the integrals involving just the normal density function $\varphi(w)$ are computed with the formula 26.2.17 in Abramowitz and Stegun (1965). A value of $\overline{Pr}(.)$ lower than 1% indicates that the sample contains practically no limit observations.

2.2 Goodness-of-fit Measures

Two goodness-of-fit measures which indicate the accuracy with which a specified model adjusts to the observed data are given in the program: the R_E^2 measure is computed with the expected value of Y_t derived in the previous section and the R_L^2 measure is based on the likelihood ratio test statistic Λ. Since both measures are just nonlinear extensions of the standard linear regression case, they are called *pseudo*$-R^2$ measures. Two types of values — unadjusted and adjusted for degrees of freedom — are computed for each R^2 measure. Only the adjusted type is valid when comparing non-nested models.

Pseudo-(E)$-R^2$

— unadjusted:

$$R_E^2 = 1 - \frac{\sum_t [Y_t - E(Y_t)]^2}{\sum_t \left(Y_t - \bar{Y}\right)^2} \tag{58}$$

— adjusted:

$$\bar{R}_E^2 = 1 - \frac{N-1}{N-K^*}\left(1 - R_E^2\right) \tag{59}$$

where K^* is the total number of estimated parameters in Π^* which includes all the β-coefficients and a set of more or less restricted $(\lambda_y, \lambda_x, \lambda_z, \delta, \rho)$ depending on the specification of the model considered. Note that both types can give a negative R^2 value since in the nonlinear case $E(Y_t)$ is no longer equal to the fitted value $\widehat{Y}_t$ as in the standard linear regression case, so that the sum of squares of residuals can be greater than the total sum of squares of Y_t in deviation around its sample mean $\bar{Y}$.

Pseudo-(L)$-R^2$

— unadjusted:

$$R_L^2 = 1 - \Lambda^{2/N} \tag{60}$$

— adjusted:

$$\bar{R}_L^2 = 1 - \frac{N-1}{N-K^*}\left(1 - R_L^2\right) \tag{61}$$

where Λ is the ratio of the likelihood function $\mathcal{L}$ when maximized with respect to the regression constant β_0 only, to the one with respect to Π^* as defined above:

$$\Lambda = \max_{\beta_0} \mathcal{L} / \max_{\Pi^*} \mathcal{L}\ . \tag{62}$$

Unlike R_E^2 which can be negative, the R_L^2 measure always remains inside the interval 0–1 since the maximum of $\mathcal{L}$ associated with β_0 — which corresponds to the most restrictive model where no independent variables except for the constant term are specified, the dependent variable is in linear form, and the problems of heteroskedasticity and autocorrelation are not considered — is necessarily smaller than the maximum associated with Π^* which includes less restricted parameters.

Note that for the standard linear regression case $(\lambda_y = \lambda_{xk} = 1, \forall k)$ without heteroskedasticity $(\delta = 0, \forall \lambda_z)$ and autocorrelation $(\rho = 0)$, the two measures, R_L^2 and R_E^2, hence their adjusted forms, coincide since $\Lambda^{2/N}$ reduces to the ratio of the unexplained sum of squares to the total sum of squares of Y_t in deviation form:

$$\Lambda^{2/N} = \frac{\sum_t \left(Y_t^{**} - \widehat{Y}_t^{**}\right)^2}{\sum_t \left(Y_t - \bar{Y}\right)^2} \left[\sum_t \frac{Y_t^{\lambda_y - 1}}{f(Z_t)^{1/2}}\right]^{-2/N} = \frac{\sum_t \left(Y_t - \widehat{Y}_t\right)^2}{\sum_t \left(Y_t - \bar{Y}\right)^2} \tag{63}$$

where $\widehat{Y}_t^{**}$ is the fitted value of Y_t^{**}.

2.3 Elasticities and Derivatives of the Dependent Variable

Two types of elasticity are computed in the program:

1. The first type of elasticity, denoted as $E^s_{X_{jt}}$, is defined in terms of the sample value of Y_t: it measures the percent variation of the dependent variable Y_t due to a percent variation of an independent variable X_{jt}, given all the other independent variables X_{kt}'s as well as all the variables Z_{mt}'s $(Z_{mt} \neq X_{jt}, \forall m)$ in $f(Z_t)$ fixed at their observed values.
2. The second type of elasticity, denoted as $E^e_{X_{jt}}$, which is more relevant since the model is nonlinear in Y_t, is defined in terms of the expected value of Y_t derived in Section 2.1.

When heteroskedasticity is present, two types of elasticity, namely $E^s_{Z_{mt}}$ and $E^e_{Z_{mt}}$, can also be computed with respect to a heteroskedasticity variable Z_{mt} specified in the function $f(Z_t)$. Two specifications of the variable Z_{mt} should be considered:

1. The variable Z_{mt} specified in $f(Z_t)$ is also used as an independent variable X_{jt}.
2. It is only specified in $f(Z_t)$, not elsewhere.

By definition, the elasticity can be evaluated for every observation t in the sample. To obtain an overall value of elasticity for the whole sample, it is usual to compute the measure at the sample means of Y_t, X_t and Z_t. In the program, two sets of observations are used to compute the sample means:

— the total set used for the estimation of the parameters, namely the observations $t = 1 + r, ..., n$, which will give the elasticity for the whole sample;

— the subset including the last observations from the total set, which can be specified by the user: for example, 12 last observations can be selected for a monthly time-series to compare the elasticity of this period to the overall value for the whole sample.

•Elasticities in terms of the Sample Value of Y_t

The sample value of Y_t has the same form as $Y_t(w)$ in (48), but with w replaced by ω_t.

The elasticity of the sample value of Y_t with respect to an independent variable X_{jt} can be written as:

$$E^s_{X_{jt}} = \frac{\partial Y_t}{\partial X_{jt}} \frac{X_{jt}}{Y_t} = \frac{1}{Y_t^{\lambda_y}} \left(\beta_j X_{jt}^{\lambda_{xj}} + G_{X_{jt}} \right) \tag{64}$$

where

$$G_{X_{jt}} = \begin{cases} 0 & \text{if } f(Z_t) \text{ is not a function of } X_{jt}, \\ \frac{1}{2}\delta_m Z_{mt}^{\lambda_{zm}} \left[f(Z_t)^{1/2} A_t - \sum_k \beta_k X_{kt}^{(\lambda_{xk})} \right] & \text{if } f(Z_t) \text{ contains one } Z_{mt} = X_{jt}. \end{cases} \tag{65}$$

and $A_t = \Sigma_\ell \rho_\ell Y^*_{t-\ell} + \Sigma_k \beta_k X^{**}_{kt}$.

Similarly, the elasticity of the sample value of Y_t with respect to a heteroskedasticity variable Z_{mt} which appears only in $f(Z_t)$ is as follows:

$$E^s_{Z_{mt}} = \frac{\partial Y_t}{\partial Z_{mt}} \frac{Z_{mt}}{Y_t} = \frac{G_{Z_{mt}}}{Y_t^{\lambda_y}} \tag{66}$$

where $G_{Z_{mt}} = \frac{1}{2}\delta_m Z_{mt}^{\lambda_{zm}} \left[f(Z_t)^{1/2} A_t - \sum_k \beta_k X_{kt}^{(\lambda_{xk})} \right]$.

•Elasticities in terms of the Expected Value of Y_t

Using $E(Y_t)$ in (51)-(52)-(53) depending on the value of λ_y, the elasticity of the expected value of Y_t with respect to an independent variable X_{jt} can be written as:

$$E^e_{X_{jt}} = \frac{\partial E(Y_t)}{\partial X_{jt}} \frac{X_{jt}}{E(Y_t)} = \frac{1}{E(Y_t)} \int\limits_{R_w} [Y_t(w)]^{1-\lambda_y} \Big[\beta_j X_{jt}^{\lambda_{xj}} + H_{X_{jt}}(w)\Big] \varphi(w) dw \tag{67}$$

where R_w is the integration domain of $w : [w_t^*, \infty], [-\infty, \infty]$ and $[-\infty, w_t(\nu)]$ if $\lambda_y > 0, \lambda_y = 0$ and $\lambda_y < 0$ respectively, and

$$H_{X_{jt}}(w) = \begin{cases} 0 & \text{if } f(Z_t) \text{ is not a function of } X_{jt}, \\ \frac{1}{2}\delta_m Z_{mt}^{(\lambda_{zm})} \Big[f(Z_t)^{1/2}(A_t + w) - \sum\limits_k \beta_k X_{kt}^{(\lambda_{xk})} \Big] & \text{if } f(Z_t) \text{ contains } Z_{mt} = X_{jt}. \end{cases} \tag{68}$$

Likewise, the elasticity of the expected value of Y_t with respect to a heteroskedasticity variable Z_{mt} which appears only in $f(Z_t)$ is as follows:

$$E^e_{Z_{mt}} = \frac{\partial E(Y_t)}{\partial Z_{mt}} \frac{Z_{mt}}{E(Y_t)} = \frac{1}{E(Y_t)} \int\limits_{R_w} [Y_t(w)]^{1-\lambda_y} H_{Z_{mt}}(w) \varphi(w) dw \tag{69}$$

where $H_{Z_{mt}}(w) = \frac{1}{2}\delta_m Z_{mt}^{\lambda_{zm}} \Big[f(Z_t)^{1/2}(A_t + w) - \sum\limits_k \beta_k X_{kt}^{(\lambda_{xk})} \Big]$.

•Derivatives of the Dependent Variable

The derivatives $D^s_{X_{jt}} = \partial Y_t / \partial X_{jt}, D^s_{Z_{mt}} = \partial Y_t / \partial Z_{mt}$, $D^e_{X_{jt}} = \partial E(Y_t) / \partial X_{jt}$ and $D^e_{Z_{mt}} = \partial E(Y_t) / \partial Z_{mt}$ which are implicitly given in (64), (66), (67) and (69) for each observation t are also computed at the sample means of the variables corresponding to the total set and the subset of observations. For a nonlinear model, these derivatives are more relevant for comparing the different specifications of the model than the β-coefficients themselves.

•Corrected Elasticities for a Quasi-Dummy and a Real Dummy

The concept of a derivative, hence an elasticity, strictly implies that an independent variable is of continuous type. Since all the independent variables specified in a model may not be necessarily continuous, a classification of the variables into four categories is needed to allow for the correction of the elasticities for the positive observations of two categories of variables, namely the “quasi-dummies” and the real dummies:

Category 0: Continuous variable which is strictly positive, hence can be transformed by Box-Cox.

Category 1: Any continuous variable which is not strictly positive, hence cannot be transformed by Box-Cox.

Category 2: Quasi-dummy, i.e. a continuous variable containing only positive and null values, for example the level of snowfall per month, which is positive in winter, but null in summer. Strictly speaking, this type of variable cannot be transformed by Box-Cox due to the null values, but if a Box-Cox transformation is absolutely necessary, then only the positive observations of the quasi-dummy are transformed. Since the null observations cannot be transformed, an associated real dummy, which has a value of 1 for the positive observations of the quasi-dummy and a value of 0 otherwise, must also be created and specified in the regression equation.

Category 3: Real dummy which has only two values: 0 and a positive constant, for example a binary (0 – 1) variable such as sex (0 for male and 1 for female), or a dummy associated with a quasi-dummy defined above. Unlike a quasi-dummy, a real dummy cannot be transformed by Box-Cox since all the values of 1 will be reduced to zero, if they are transformed with any value of the Box-Cox λ-parameter, so that the transformed real dummy will contain only null observations.

The correction of the elasticities for the positive observations of a quasi-dummy or a real dummy is made in a separate program called **TABLEX**, which allows the user to present the estimation results from various specifications of a model in a table for ease of comparison among the coefficients, the parameters, the elasticities, the log-likelihood values, etc., specially when the model includes a great number of independent variables and that different forms of heteroskedasticity and/or structures of autocorrelation are estimated along with the different types of constraints on the Box-Cox parameters λ_y and λ_x.

The correction is based on the fact that one is interested in the effect on the dependent variable only when an activity or a phenomenon represented by the quasi-dummy or the real dummy really occurs, i.e. when the observations of the dummy are only positive. The following correction formula considered in Dagenais et al. (1987) is used:

$$E_j^+ = E_j \frac{\bar{X}_j^+}{\bar{X}_j} = E_j \frac{\sum\limits_t X_{jt}^+ / N_j^+}{\sum\limits_t X_{jt}/N} = E_j \frac{N}{N_j^+} \tag{70}$$

where E_j represents $E^s_{X_{jt}}$ or $E^e_{X_{jt}}$ evaluated at the sample means, $\bar{X}_j$ is the sample mean of a quasi-dummy or real dummy X_j computed from the total set of observations used in estimation, and $\bar{X}_j^+$ is the sample mean computed only from the positive observations of X_j whose number is N_j^+. Note that for the subset of the last observations from the total set, the correction formula is analogous.

In the special case where a quasi-dummy, say Q_t, is transformed by Box-Cox, the correction will change slightly for $E^e_{Q_t}$ since the expected value of Y_t evaluated at the sample means of X and Z, say $E(Y_t)|_{\bar{X},\bar{Z}}$, is computed in fact from the sample mean of a strictly positive variable $\widetilde{Q}_t$ instead of Q_t so that the well-known properties, namely $E(Y_t)|_{\bar{X},\bar{Z}} = \bar{Y}$ and $E^e_{Q_t}|_{\bar{X},\bar{Z}} = E^s_{Q_t}|_{\bar{X},\bar{Z}}$ in the standard linear regression case, may be preserved. Transforming only the positive observations of Q_t by Box-Cox while leaving the null observations of the variable untouched is exactly equivalent to applying the Box-Cox transformation on a strictly positive variable $\widetilde{Q}_t$ such that $\widetilde{Q}_t$ is identical to Q_t for the portion of positive observations Q_t^+ and $\widetilde{Q}_t$ is equal to 1 for the portion of null observations of Q_t, since $\widetilde{Q}_t^{(\lambda_Q)} = 0, \forall \lambda_Q$. Thus, the correction formula in this case has the following form:

$$E^+_{Q_t} = E^e_Q \frac{\bar{Q}^+}{\bar{\widetilde{Q}}} = E^e_Q \frac{\Sigma_t Q_t^+ / N_Q^+}{\Sigma_t \widetilde{Q}_t / N} = E^e_Q \frac{N}{N_Q^+} \left(\frac{\Sigma_t Q_t^+}{N_Q^0 + \Sigma_t Q_t^+} \right) \tag{71}$$

where E^e_Q is evaluated at the sample means and N_Q^+ and N_Q^0 are respectively the numbers of positive and null observations of Q_t.

•Derivatives and Elasticities for the Linear and Logarithmic Forms of the Dependent Variable

The most usual forms of the dependent variable Y_t encountered in pratice are the linear $(\lambda_y = 1)$ and logarithmic $(\lambda_y = 0)$ cases:

1. For these two cases, the explicit forms of the sample value of Y_t and its derivative and elasticity are given in TABLE 1, depending on the presence or absence of heteroskedasticity. To be completely general, autocorrelation is considered in these forms which can be further reduced in the absence of autocorrelation by setting all the autoregressive coefficients ρ_ℓ's included in A_t equal to zero.
2. Similarly, for both cases, explicit forms of the expected value of Y_t and its derivative and elasticity — whenever possible due to the integrals which cannot be reduced to closed forms — are summarized in TABLE 2.

TABLE 1 Explicit forms of the sample value of Y_t and its derivative and elasticity for the linear and logarithmic cases of Y_t.

STATISTIC	CASE	HOMOSKEDASTICITY (No $f(Z_t)$ involved)	HETEROSKEDASTICITY $X_{jt} \neq Z_{mt}$	$X_{jt} = Z_{mt}$
SAMPLE VALUE Y_t	$\lambda_y = 1$	$1 + A_t$	$1 + f(Z_t)^{1/2} A_t$	
	$\lambda_y = 0$	$\exp(A_t)$	$\exp\left[f(Z_t)^{1/2} A_t\right]$	
DERIVATIVE $D^s_{X_{jt}}$	$\lambda_y = 1$	$\beta_j X_{jt}^{\lambda_{xj}-1}$		$\beta_j X_{jt}^{\lambda_{xj}-1} + Z_{mt}^{-1} G_{Z_{mt}}$
	$\lambda_y = 0$	$\beta_j X_{jt}^{\lambda_{xj}-1} Y_t$		$\left(\beta_j X_{jt}^{\lambda_{xj}-1} + Z_{mt}^{-1} G_{Z_{mt}}\right) Y_t$
$D^s_{Z_{mt}}$	$\lambda_y = 1$	Not applicable	$Z_{mt}^{-1} G_{Z_{mt}}$	Same as above
	$\lambda_y = 0$	Not applicable	$Z_{mt}^{-1} G_{Z_{mt}} Y_t$	Same as above
ELASTICITY $E^s_{X_{jt}}$	$\lambda_y = 1$	$\beta_j X_{jt}^{\lambda_{xj}} / Y_t$		$\left(\beta_j X_{jt}^{\lambda_{xj}} + G_{Z_{mt}}\right) / Y_t$
	$\lambda_y = 0$	$E^s_{X_{jt}} = E^e_{X_{jt}} = \beta_j X_{jt}^{\lambda_{xj}}$		$\beta_j X_{jt}^{\lambda_{xj}} + G_{Z_{mt}}$
$E^s_{Z_{mt}}$	$\lambda_y = 1$	Not applicable	$G_{Z_{mt}} / Y_t$	Same as above
	$\lambda_y = 0$	Not applicable	$G_{Z_{mt}}$	Same as above

where $A_t = \sum_\ell \rho_\ell Y^*_{t-\ell} + \sum_k \beta_k X^{**}_{kt}$ and $G_{Z_{mt}} = \frac{1}{2} \delta_m Z_{mt}^{\lambda_{zm}} \left[f(Z_t)^{1/2} A_t - \sum_k \beta_k X_{kt}^{(\lambda_{xk})} \right]$.

TABLE 2 Explicit forms of the expected value of Y_t and its derivative and elasticity for the linear and logarithmic cases of Y_t.

STATISTIC	CASE	HOMOSKEDASTICITY (No $f(Z_t)$ involved)	HETEROSKEDASTICITY $X_{jt} \neq Z_{mt}$	$X_{jt} = Z_{mt}$
EXPECTED VALUE $E(Y_t)$	$\lambda_y = 1$	$E(Y_t) = \int_{w_t^*}^{\infty} Y_t(w)\varphi(w)dw$ where $Y_t(w) = 1 + A_t + w$; $\boxed{\text{LIMIT CASE}}$ $\lim_{w_t^* \to \infty} E(Y_t) = 1 + A_t$	$E(Y_t) = \int_{w_t^*}^{\infty} Y_t(w)\varphi(w)dw$ where $Y_t(w) = 1 + f(Z_t)^{1/2}(A_t + w)$; $\boxed{\text{LIMIT CASE}}$ $\lim_{w_t^* \to -\infty} E(Y_t) = 1 + f(Z_t)^{1/2}A_t$	
	$\lambda_y = 0$	$E(Y_t) = \int_{-\infty}^{\infty} Y_t(w)\varphi(w)dw$ where $Y_t(w) = \exp(A_t + w)$	$E(Y_t) = \int_{-\infty}^{\infty} Y_t(w)\varphi(w)dw$ where $Y_t(w) = \exp\left[f(Z_t)^{1/2}(A_t + w)\right]$	
DERIVATIVE $D^e_{X_{jt}}$	$\lambda_y = 1$	$D^e_{X_{jt}} = \beta_j X_{jt}^{\lambda_{xj}-1} \int_{w_t^*}^{\infty} \varphi(w)dw$; $\boxed{\text{LIMIT CASE}}$ $\lim_{w_t^* \to -\infty} D^e_{X_{jt}} = \beta_j X_{jt}^{\lambda_{xj}-1}$		Same as left column $+D^e_{Z_{mt}}(\lambda_y = 1)$; $\boxed{\text{LIMIT CASE}}$ $\lim_{w_t^* \to -\infty} D^e_{X_{jt}} = \beta_j X_{jt}^{\lambda_{xj}-1} + Z_{mt}^{-1} G_{Z_{mt}}$
	$\lambda_y = 0$	$D^e_{X_{jt}} = \beta_j X_{jt}^{\lambda_{xj}-1} E(Y_t)$		Same as left column $+D^e_{Z_{mt}}(\lambda_y = 0)$
$D^e_{Z_{mt}}$	$\lambda_y = 1$	Not applicable	$D^e_{Z_{mt}}(\lambda_y = 1)$	Same as above
	$\lambda_y = 0$	Not applicable	$D^e_{Z_{mt}}(\lambda_y = 0)$	Same as above
ELASTICITY $E^e_{X_{jt}}$	$\lambda_y = 1$	$E^e_{X_{jt}} = \frac{\beta_j X_{jt}^{\lambda_{xj}}}{E(Y_t)} \int_{w_t^*}^{\infty} \varphi(w)dw$; $\boxed{\text{LIMIT CASE}}$ $\lim_{w_t^* \to -\infty} E^e_{X_{jt}} = \frac{\beta_j X_{jt}^{\lambda_{xj}}}{1+A_t}$		same as left column $+E^e_{Z_{mt}}(\lambda_y = 1)$; $\boxed{\text{LIMIT CASE}}$ $\lim_{w_t^* \to -\infty} E^e_{X_{jt}} = \frac{\beta_j X_{jt}^{\lambda_{xj}} + G_{Z_{mt}}}{1+f(Z_t)^{1/2}A_t}$
	$\lambda_y = 0$	$E^e_{X_{jt}} = E^s_{X_{jt}} = \beta_j X_{jt}^{\lambda_{xj}}$		Same as left column $+E^e_{Z_{mt}}(\lambda_y = 0)$
$E^e_{Z_{mt}}$	$\lambda_y = 1$	Not applicable	$E^e_{Z_{mt}}(\lambda_y = 1)$	Same as above
	$\lambda_y = 0$	Not applicable	$E^e_{Z_{mt}}(\lambda_y = 0)$	Same as above

where $A_t = \sum_l \rho_l Y^*_{t-l} + \sum_k \beta_k X^{**}_{kt}$, $G_{Z_{mt}} = \frac{1}{2}\delta_m Z_{mt}^{\lambda_{zm}}\left[f(Z_t)^{1/2}A_t - \sum_k \beta_k X_{kt}^{(\lambda_{xk})}\right]$,

$D^e_{Z_{mt}}(\lambda_y = 1) = Z_{mt}^{-1}\int_{w_t^*}^{\infty} H_{Z_{mt}}(w)\varphi(w)dw$, $D^e_{Z_{mt}}(\lambda_y = 0) = Z_{mt}^{-1}\int_{-\infty}^{\infty} Y_t(w)H_{Z_{mt}}(w)\varphi(w)dw$,

$E^e_{Z_{mt}}(\lambda_y = 1) = \frac{1}{E(Y_t)}\int_{w_t^*}^{\infty} H_{Z_{mt}}(w)\varphi(w)dw$, $E^e_{Z_{mt}}(\lambda_y = 0) = \frac{1}{E(Y_t)}\int_{-\infty}^{\infty} Y_t(w)H_{Z_{mt}}(w)\varphi(w)dw$,

and $H_{Z_{mt}}(w) = \frac{1}{2}\delta_m Z_{mt}^{\lambda_{zm}}\left[f(Z_t)^{1/2}(A_t + w) - \sum_k \beta_k X_{kt}^{(\lambda_{xk})}\right]$.

2.4 Elasticities and Derivatives of the Standard Error of the Dependent Variable

The concept of elasticity can be extended to the standard error of the dependent variable Y_t, $\sigma(Y_t)$, to give a measure of the percent variation of the standard error of Y_t due to a percent variation of an independent variable X_{jt} or a heteroskedasticity variable Z_{mt} specified in the function $f(Z_t)$.

Like for the elasticities and derivatives of the sample and expected values of Y_t, two cases of heteroskedasticity can be distinguished for the elasticities and derivatives of $\sigma(Y_t)$:

1. The heteroskedasticity variable Z_{mt} is also used as an independent variable X_{jt}.
2. It is specified only in the function $f(Z_t)$, not elsewhere.

•Elasticity and Derivative of $\sigma(Y_t)$ with respect to an Independent Variable

The elasticity of the standard error of Y_t with respect to an independent variable X_{jt} is defined as:

$$E^{\sigma}_{jt} = \frac{\partial \sigma(Y_t)}{\partial X_{jt}} \frac{X_{jt}}{\sigma(Y_t)} \tag{72}$$

where

— the standard error of Y_t is:

$$\sigma(Y_t) = \sqrt{\text{Var}(Y_t)} = \sqrt{E\left(Y_t^2\right) - [E(Y_t)]^2}\ , \tag{73}$$

— the derivative of the standard error of Y_t with respect to X_{jt} is:

$$D^{\sigma}_{X_{jt}} = \frac{\partial \sigma(Y_t)}{\partial X_{jt}} = \frac{1}{2\sigma(Y_t)} \left[\frac{\partial E\left(Y_t^2\right)}{\partial X_{jt}} - 2E(Y_t) \frac{\partial E(Y_t)}{\partial X_{jt}} \right]\ , \tag{74}$$

— the second moment of Y_t is:

$$E\left(Y_t^2\right) = \begin{cases} \int\limits_{w_t^*}^{\infty} Y_t^2(w)\varphi(w)dw & \text{if } \lambda_y > 0\ , \\ \int\limits_{-\infty}^{\infty} Y_t^2(w)\varphi(w)dw & \text{if } \lambda_y = 0\ , \\ \int\limits_{-\infty}^{w_t(\nu)} Y_t^2(w)\varphi(w)dw + \nu^2 \int\limits_{w_t(\nu)}^{\infty} \varphi(w)dw & \text{if } \lambda_y < 0\ , \end{cases} \tag{75}$$

— the first moment of Y_t which is already defined is repeated here for convenience:

$$E(Y_t) = \begin{cases} \int\limits_{w_t^*}^{\infty} Y_t(w)\varphi(w)dw & if\ \lambda_y > 0\ , \\ \int\limits_{-\infty}^{\infty} Y_t(w)\varphi(w)dw & if\ \lambda_y = 0\ , \\ \int\limits_{-\infty}^{w_t(\nu)} Y_t(w)\varphi(w)dw + \nu \int\limits_{w_t(\nu)}^{\infty} \varphi(w)dw & if\ \lambda_y < 0\ , \end{cases} \tag{76}$$

— the derivative of $E\left(Y_t^2\right)$ with respect to X_{jt} is:

$$\frac{\partial E\left(Y_t^2\right)}{\partial X_{jt}} = 2 \int\limits_{R_w} [Y_t(w)]^{2-\lambda_y} \left[\beta_j X_{jt}^{\lambda_{xj}-1} + X_{jt}^{-1} H_{jt}(w) \right] \varphi(w)dw \tag{77}$$

— and the derivative of $E(Y_t)$ with respect to X_{jt} is:

$$\frac{\partial E(Y_t)}{\partial X_{jt}} = \int\limits_{R_w} [Y_t(w)]^{1-\lambda_y} \left[\beta_j X_{jt}^{\lambda_{xj}-1} + X_{jt}^{-1} H_{X_{jt}}(w) \right] \varphi(w)dw\ . \tag{78}$$

Note that $Y_t(w), R_w$ and $H_{jt}(w)$ are already defined in (48), (67) and (68) respectively, and that $H_{X_{jt}}(w)$ includes the first case of heteroskedasticity where X_{jt} appears also as a heteroskedasticity variable Z_{mt} used in $f(Z_t)$.

After some algebraic manipulations, the derivative of the standard error of Y_t with respect to X_{jt} can be expressed as:

$$D^{\sigma}_{X_{jt}} = \frac{1}{\sigma(Y_t)} \int\limits_{R_w} [Y_t(w)]^{1-\lambda_y}[Y_t(w) - E(Y_t)]\Big[\beta_j X_{jt}^{\lambda_{xj}-1} + X_{jt}^{-1} H_{X_{jt}}(w)\Big]\varphi(w)dw \tag{79}$$

Hence, the elasticity of the standard error of Y_t can be written as:

$$E^{\sigma}_{jt} = \frac{1}{Var(Y_t)} \int\limits_{R_w} [Y_t(w)]^{1-\lambda_y}[Y_t(w) - E(Y_t)]\Big[\beta_j X_{jt}^{\lambda_{xj}} + H_{X_{jt}}(w)\Big]\varphi(w)dw\ . \tag{80}$$

•Elasticity and Derivative of $\sigma(Y_t)$ with respect to a Heteroskedasticity Variable

Since the first case of heteroskedasticity has been previously treated, only the second case where Z_{mt} appears only in $f(Z_t)$ is considered here. The formulas for $D^{\sigma}_{Z_{mt}}$ and $E^{\sigma}_{Z_{mt}}$ are analogous to $D^{\sigma}_{X_{jt}}$ and $E^{\sigma}_{X_{jt}}$ without the terms $\beta_j X_{jt}^{\lambda_{xj}-1}$ and $\beta_j X_{jt}^{\lambda_{xj}}$ respectively:

$$D^{\sigma}_{Z_{mt}} = \frac{Z_{mt}^{-1}}{\sigma(Y_t)} \int\limits_{\mathcal{R}_w} [Y_t(w)]^{1-\lambda_y}[Y_t(w) - E(Y_t)] H_{Z_{mt}}(w)\varphi(w)dw. \tag{81}$$

$$E^{\sigma}_{Z_{mt}} = \frac{1}{\text{Var}(Y_t)} \int\limits_{\mathcal{R}_w} [Y_t(w)]^{1-\lambda_y}[Y_t(w) - E(Y_t)] H_{Z_{mt}}(w)\varphi(w)dw. \tag{82}$$

where $H_{Z_{mt}}(w)$ is already defined in (69).

In the program, three statistics are computed:

1. Standard error of Y_t for each observation t,
2. Derivatives of $\sigma(Y_t)$ with respect to X_{jt} and Z_{mt} at the sample means,
3. Elasticities of $\sigma(Y_t)$ with respect to X_{jt} and Z_{mt} at the sample means.

•Standard Error of Y_t for each observation t

The standard error of Y_t is computed for each observation t used in the estimation. In general, this statistic will vary with each observation t. In TABLE 4, explicit forms of $\sigma(Y_t)$ are given for the linear and logarithmic cases of Y_t depending on the presence or absence of heteroskedasticity. As in TABLES 1 and 2 for the sample value Y_t and the expected value $E(Y_t)$ respectively, autocorrelation is considered in these forms which can be further reduced in the absence of autocorrelation by setting all the autoregressive coefficients ρ_ℓ's included in A_t equal to zero.

1. **Linear Case** $(\lambda_y = 1)$

   - **Homoskedasticity:** $\sigma(Y_t)$ varies slightly depending on the value of the lower bound of integration w_t^* except if w_t^* tends towards $-\infty$, $\sigma(Y_t)$ is constant and equal to the standard error of the residuals w_t's :

$$\lim_{w_t^*\to-\infty}\sigma(Y_t)=\lim_{w_t^*\to-\infty}\left\{\int_{w_t^*}^{\infty}(1+A_t+w)^2\varphi(w)dw-\left[\int_{w_t^*}^{\infty}(1+A_t+w)\varphi(w)dw\right]^2\right\}^{\frac{1}{2}}$$
$$=\Bigg\{(1+A_t)^2\int_{-\infty}^{\infty}\varphi(w)dw+\int_{-\infty}^{\infty}w^2\varphi(w)dw+2(1+A_t)\int_{-\infty}^{\infty}w\varphi(w)dw$$
$$-\left[(1+A_t)\int_{-\infty}^{\infty}\varphi(w)dw+\int_{-\infty}^{\infty}w\varphi(w)dw\right]^2\Bigg\}^{\frac{1}{2}} \qquad (83)$$
$$=\left[\int_{-\infty}^{\infty}w^2\varphi(w)dw\right]^{\frac{1}{2}}=\sigma_w$$

since $\int_{-\infty}^{\infty}\varphi(w)dw=1$, $\int_{-\infty}^{\infty}w\varphi(w)dw=0$ and $\int_{-\infty}^{\infty}w^2\varphi(w)dw=\sigma_w^2$.

- **Heteroskedasticity:** in contrast, $\sigma(Y_t)$ varies for each observation t even if w_t^* tends towards $-\infty$. In this case, $\sigma(Y_t)$ changes proportionally to $f(Z_t)^{\frac{1}{2}}$:

$$\lim_{w_t^*\to-\infty}\sigma(Y_t)=\lim_{w_t^*\to-\infty}\Bigg\{\left[\int_{w_t^*}^{\infty}1+f(Z_t)^{\frac{1}{2}}(A_t+w)\right]^2\varphi(w)dw$$
$$-\left[\int_{w_t^*}^{\infty}\left[1+f(Z_t)^{\frac{1}{2}}(A_t+w)\right]\varphi(w)dw\right]^2\Bigg\}^{\frac{1}{2}}$$
$$=\Bigg\{\left[1+f(Z_t)^{\frac{1}{2}}A_t\right]^2\int_{-\infty}^{\infty}\varphi(w)dw+f(Z_t)\int_{-\infty}^{\infty}w^2\varphi(w)dw \qquad (84)$$
$$+2\left[1+f(Z_t)^{\frac{1}{2}}A_t\right]f(Z_t)^{\frac{1}{2}}\int_{-\infty}^{\infty}w\varphi(w)dw$$
$$-\left[\left[1+f(Z_t)^{\frac{1}{2}}A_t\right]\int_{-\infty}^{\infty}\varphi(w)dw+f(Z_t)^{\frac{1}{2}}\int_{-\infty}^{\infty}w\varphi(w)dw\right]^2\Bigg\}^{\frac{1}{2}}$$
$$=\left[f(Z_t)\int_{-\infty}^{\infty}w^2\varphi(w)dw\right]^{\frac{1}{2}}=f(Z_t)^{\frac{1}{2}}\sigma_w.$$

2. **Logarithmic Case** $(\lambda_y=0)$

- **Homoskedasticity:** $\sigma(Y_t)$ varies proportionally to $\exp(A_t)$:

$$\sigma(Y_t)=\exp(A_t)\left\{\int_{-\infty}^{\infty}\exp(2w)\varphi(w)dw-\left[\int_{-\infty}^{\infty}\exp(w)\varphi(w)dw\right]^2\right\}^{\frac{1}{2}}. \qquad (85)$$

- **Heteroskedasticity:** $\sigma(Y_t)$ is a nonlinear function of A_t and $f(Z_t)^{\frac{1}{2}}$:

$$\sigma(Y_t) = \exp\left[f(Z_t)^{\frac{1}{2}} A_t\right] \left\{ \int_{-\infty}^{\infty} \exp\left[2 f(Z_t)^{\frac{1}{2}} w\right] \varphi(w) dw - \left[\int_{-\infty}^{\infty} \exp\left[f(Z_t)^{\frac{1}{2}} w\right] \varphi(w) dw \right]^2 \right\}^{\frac{1}{2}}. \tag{86}$$

•Elasticity and Derivative of the Standard Error of Y_t at the sample means

The elasticities and the derivatives of $\sigma(Y_t)$ with respect to an independent variable X_{jt} and a heteroskedasticity variable Z_{mt} are computed at the sample means of all the variables for the two sets of observations previously defined for the elasticities of the sample and expected values of Y_t. The correction of the elasticity $E^{\sigma}_{X_{jt}}$ for the positive observations of a quasi-dummy or a real dummy is the same as for the elasticities $E^{s}_{X_{jt}}$ and $E^{e}_{X_{jt}}$ (see TABLE 3). Like the case of the derivatives $D^{s}_{X_{jt}}$ and $D^{e}_{X_{jt}}$, no correction of the derivative $D^{\sigma}_{X_{jt}}$ is made.

TABLE 3 Correction of the elasticities $E^{s}_{X_{jt}}$, $E^{e}_{X_{jt}}$ and $E^{\sigma}_{X_{jt}}$ at the sample means, for the positive observations of a quasi-dumy or a real dummy.

CATEGORY	ELASTICITY $E^{s}_{X_{jt}}$	$E^{e}_{X_{jt}}$	$E^{\sigma}_{X_{jt}}$
(2) Quasi-dummy • No Box-Cox	$E_j \frac{N}{N_j^+}$	$E_j \frac{N}{N_j^+}$	$E_j \frac{N}{N_j^+}$
• With Box-Cox	$E_j \frac{N}{N_j^+}\left(\frac{\sum_t X_{jt}^+}{N_j^0 + \sum_t X_{jt}^+}\right)$	$E_j \frac{N}{N_j^+}\left(\frac{\sum_t X_{jt}^+}{N_j^0 + \sum_t X_{jt}^+}\right)$	$E_j \frac{N}{N_j^+}\left(\frac{\sum_t X_{jt}^+}{N_j^0 + \sum_t X_{jt}^+}\right)$
(3) Real dummy • No Box-Cox	$E_j \frac{N}{N_j^+}$	$E_j \frac{N}{N_j^+}$	$E_j \frac{N}{N_j^+}$

where E_j represents $E^{s}_{X_{jt}}$, $E^{e}_{X_{jt}}$ or $E^{\sigma}_{X_{jt}}$ evaluated at the sample means, N_j^+ is the number of positive observations X_{jt}^+'s, N_j^0 is the number of null observations of X_{jt} and N is the total number of observations $\left(N = N_j^0 + N_j^+\right)$ included in each set of observations (total set used for estimation, or subset of the last observations from the total set).

In TABLE 4, explicit forms of the derivatives and the elasticities of $\sigma(Y_t)$ with respect to X_{jt} and Z_{mt} for the linear and logarithmic cases of Y_t are given in terms of an observation t to show how both statistics vary with each observation.

A. Derivative and Elasticity with respect to X_{jt}

1. **Linear Case** $(\lambda_y = 1)$

- **Homoskedasticity:** the derivative $D^{\sigma}_{X_{jt}}$ and the elasticity $E^{\sigma}_{X_{jt}}$ vary slightly depending on the value of the lower bound of integration w_t^*, but are both equal to zero as w_t^* tends towards $-\infty$, since $\lim_{w_t^* \to -\infty} \int_{w_t^*}^{\infty} [Y_t(w) - E(Y_t)]\varphi(w)dw$ is equal to zero.
- **Heteroskedasticity** $(X_{jt} \neq Z_{mt})$**:** this case is identical to the previous one since the heteroskedasticity function $f(Z_t)$ does not include any Z_{mt} which is used also as an independent variable X_{jt}.
- **Heteroskedasticity** $(X_{jt} = Z_{mt})$**:** both $D^{\sigma}_{X_{jt}}$ and $E^{\sigma}_{X_{jt}}$ include two components: one resulting from the variation of X_{jt} alone and the other from the variation of Z_{mt}. In the limit case where w_t^* tends towards $-\infty$, only the second component remains.

2. **Logarithmic Case** $(\lambda_y = 0)$

- **Homoskedasticity:** the derivative $D^{\sigma}_{X_{jt}}$ is a function of X_{jt} and $\sigma(Y_t)$, but the elasticity $E^{\sigma}_{X_{jt}}$ is a function of X_{jt} only and it is identical to both elasticities $E^{e}_{X_{jt}}$ and $E^{s}_{X_{jt}}$.
- **Heteroskedasticity** $(X_{jt} \neq Z_{mt})$**:** this case is identical to the previous one since the heteroskedasticity function $f(Z_t)$ does not include any Z_{mt} which is used also as an independent variable X_{jt}.
- **Heteroskedasticity** $(X_{jt} = Z_{mt})$**:** the derivative $D^{\sigma}_{X_{jt}}$ and the elasticity $E^{\sigma}_{X_{jt}}$ depend on $X_{jt}, Z_{mt}, \sigma(Y_t), Y_t(w), E(Y_t)$ and $H_{Z_{mt}}$.

B. Derivative and Elasticity with respect to Z_{mt}

1. **Linear Case** $(\lambda_y = 1)$

- **Homoskedasticity:** since $f(Z_t)$ is not involved in this case, the derivative $D^{\sigma}_{Z_{mt}}$ and the elasticity $E^{\sigma}_{Z_{mt}}$ do not apply.
- **Heteroskedasticity** $(X_{jt} \neq Z_{mt})$**:** the derivative and the elasticity depend on $w_t^*, Z_{mt}, \sigma(Y_t), Y_t(w), E(Y_t)$ and $H_{Z_{mt}}$. If w_t^* tends towards $-\infty$, both statistics have the same limits as in the case A.1 with heteroskedasticity $(X_{jt} = Z_{mt})$ since only the heteroskedasticity component remains.
- **Heteroskedasticity** $(X_{jt} = Z_{mt})$**:** this case is identical to the case A.1 with heteroskedasticity $(X_{jt} = Z_{mt})$.

2. **Logarithmic case** $(\lambda_y = 0)$

- **Homoskedasticity:** since $f(Z_t)$ is not involved in this case, the derivative $D^{\sigma}_{Z_{mt}}$ and the elasticity $E^{\sigma}_{Z_{mt}}$ do not apply.
- **Heteroskedasticity** $(X_{jt} \neq Z_{mt})$**:** the derivative and the elasticity depend on $Z_{mt}, \sigma(Y_t), Y_t(w), E(Y_t)$ and $H_{Z_{mt}}$.
- **Heteroskedasticity** $(X_{jt} = Z_{mt})$**:** this case is identical to the case A.2 with heteroskedasticity $(X_{jt} = Z_{mt})$.

TABLE 4 Explicit forms of the standard error of Y_t and its derivative and elasticity for the linear and logarithmic cases of Y_t.

STATISTIC	CASE	HOMOSKEDASTICITY (No $f(Z_t)$ involved)	HETEROSKEDASTICITY $X_{jt} \neq Z_{mt}$	HETEROSKEDASTICITY $X_{jt} = Z_{mt}$
STANDARD ERROR $\sigma(Y_t)$	$\lambda_y = 1$	$\sigma(Y_t) = \left\{ \int_{w_t^*}^{\infty} Y_t^2(w)\varphi(w)dw - \left[\int_{w_t^*}^{\infty} Y_t(w)\varphi(w)dw \right]^2 \right\}^{1/2}$ where $Y_t(w) = 1 + A_t + w$	where $Y_t(w) = 1 + f(Z_t)^{1/2}(A_t + w)$	
		$\boxed{\text{LIMIT CASE}}$ $\lim_{w_t^* \to -\infty} \sigma(Y_t) = \sigma_w$	$\boxed{\text{LIMIT CASE}}$ $\lim_{w_t^* \to -\infty} \sigma(Y_t) = f(Z_t)^{1/2}\sigma_w$	
	$\lambda_y = 0$	$\sigma(Y_t) = \left\{ \int_{-\infty}^{\infty} Y_t^2(w)\varphi(w)dw - \left[\int_{-\infty}^{\infty} Y_t(w)\varphi(w)dw \right]^2 \right\}^{1/2}$ where $Y_t(w) = \exp(A_t + w)$	where $Y_t(w) = \exp\left[f(Z_t)^{1/2}(A_t + w)\right]$	
DERIVATIVE $D^{\sigma}_{X_{jt}}$	$\lambda_y = 1$	$D^{\sigma}_{X_{jt}} = \frac{\beta_j X_{jt}^{\lambda_{xj}-1}}{\sigma(Y_t)} \int_{w_t^*}^{\infty} [Y_t(w) - E(Y_t)]\varphi(w)dw$ $\boxed{\text{LIMIT CASE}}$ $\lim_{w_t^* \to -\infty} D^{\sigma}_{X_{jt}} = 0$		Same as left column $+D^{\sigma}_{Z_{mt}}(\lambda_y = 1)$ $\boxed{\text{LIMIT CASE}}$ $\lim_{w_t^* \to -\infty} D^{\sigma}_{X_{jt}} = \frac{1}{2}\delta_m Z_{mt}^{\lambda_{zm}-1} f(Z_t)^{1/2}\sigma_w$
	$\lambda_y = 0$	$D^{\sigma}_{X_{jt}} = \beta_j X_{jt}^{\lambda_{xj}-1}\sigma(Y_t)$		Same as left column $+D^{\sigma}_{Z_{mt}}(\lambda_y = 0)$
$D^{\sigma}_{Z_{mt}}$	$\lambda_y = 1$	Not applicable	$D^{\sigma}_{Z_{mt}}(\lambda_y = 1)$	Same as above
	$\lambda_y = 0$	Not applicable	$D^{\sigma}_{Z_{mt}}(\lambda_y = 0)$	Same as above
ELASTICITY $E^{\sigma}_{X_{jt}}$	$\lambda_y = 1$	$E^{\sigma}_{X_{jt}} = \frac{\beta_j X_{jt}^{\lambda_{xj}}}{\mathrm{Var}(Y_t)} \int_{w_t^*}^{\infty} [Y_t(w) - E(Y_t)]\varphi(w)dw$ $\boxed{\text{LIMIT CASE}}$ $\lim_{w_t^* \to -\infty} E^{\sigma}_{X_{jt}} = 0$		same as left column $+E^{\sigma}_{Z_{mt}}(\lambda_y = 1)$ $\boxed{\text{LIMIT CASE}}$ $\lim_{w_t^* \to -\infty} E^{\sigma}_{X_{jt}} = \frac{1}{2}\delta_m Z_{mt}^{\lambda_{zm}}$
	$\lambda_y = 0$	$E^{\sigma}_{X_{jt}} = E^{e}_{X_{jt}} = E^{s}_{X_{jt}} = \beta_j X_{jt}^{\lambda_{xj}}$		Same as left column $+E^{\sigma}_{Z_{mt}}(\lambda_y = 0)$
$E^{\sigma}_{Z_{mt}}$	$\lambda_y = 1$	Not applicable	$E^{\sigma}_{Z_{mt}}(\lambda_y = 1)$	Same as above
	$\lambda_y = 0$	Not applicable	$E^{\sigma}_{Z_{mt}}(\lambda_y = 0)$	Same as above

where $A_t = \sum_{\ell} \rho_\ell Y^*_{t-\ell} + \sum_k \beta_k X^{**}_{kt}$,

$$D^{\sigma}_{Z_{mt}}(\lambda_y = 1) = \frac{Z_{mt}^{-1}}{\sigma(Y_t)} \int_{w_t^*}^{\infty} [Y_t(w) - E(Y_t)] H_{Z_{mt}}(w)\varphi(w)dw,$$

$$D^{\sigma}_{Z_{mt}}(\lambda_y = 0) = \frac{Z_{mt}^{-1}}{\sigma(Y_t)} \int_{-\infty}^{\infty} Y_t(w)[Y_t(w) - E(Y_t)] H_{Z_{mt}}(w)\varphi(w)dw,$$

$$E^{\sigma}_{Z_{mt}}(\lambda_y = 1) = \frac{1}{\mathrm{Var}(Y_t)} \int_{w_t^*}^{\infty} [Y_t(w) - E(Y_t)] H_{Z_{mt}}(w)\varphi(w)dw,$$

$$E^{\sigma}_{Z_{mt}}(\lambda_y = 0) = \frac{1}{\mathrm{Var}(Y_t)} \int_{-\infty}^{\infty} Y_t(w)[Y_t(w) - E(Y_t)] H_{Z_{mt}}(w)\varphi(w)dw$$

and $H_{Z_{mt}}(w) = \frac{1}{2}\delta_m Z_{mt}^{\lambda_{zm}} \left[f(Z_t)^{1/2}(A_t + w) - \sum_k \beta_k X_{kt}^{(\lambda_{xk})} \right]$.

2.5 Student-t Statistics

The t-statistics for the estimated parameters can be computed as follows:

1. At the initial step of the maximization procedure, only the t-statistics for the estimated β-coefficients **conditional** upon the initial or fixed values of $\lambda_y, \lambda_x, \lambda_z, \delta$ and ρ can be given since they are based on the results of a standard regression of the transformed dependent variable Y^{**} on the transformed independent variables X^{**}. Hence these tests are **invariant** to changes of measurement units in the original variables Y, X and Z.
2. At the convergence of the log-likelihood function, two types of t-statistics can be obtained:
   a. **Unconditional** t-tests for the estimates of $\beta, \lambda_y, \lambda_x, \lambda_z, \delta, \rho$ and σ_w^2 using the standard errors derived from the covariance matrix (38) are computed under the null hypothesis that the parameter is equal to **zero**. These tests are **invariant** to changes of measurement units in the original variables Y, X and Z, except for β and δ-coefficients due to the application of the Box-Cox transformations on these variables. For the λ-parameters, apart from the value 0 which corresponds to a logarithmic form of the variable, another interesting value to test against is the value 1 since it corresponds to a linear specification of the variable, hence unconditional t-tests for the estimates of λ_y, λ_x and λ_z are also computed under the null hypothesis that the parameter is equal to **one**.
   b. To obtain **invariant** t-tests for the β and δ-estimates, which are more reliable in hypothesis testing than the unconditional ones which are not invariant, hence can be boosted at will by a proper choice of measurement units of the variables, a simple procedure is to compute the t-tests for β, δ, ρ and σ_w^2 **conditionally** upon the estimates of λ_y, λ_x and λ_z by using the conditional covariance matrix of β, δ, ρ and σ_w^2 which is equal to the inverse of a submatrix derived from the matrix $\left[\sum_t \frac{\partial L_t}{\partial \Pi} \frac{\partial L_t}{\partial \Pi'}\right]_{\Pi=\widehat{\Pi}'}$ in (38) by deleting the rows and columns associated with λ_y, λ_x and λ_z (Dagenais et al., 1987).

When the program fails to converge to a maximum point, the covariance matrix of the estimated parameters cannot be obtained. In this case, the standard errors will be set equal to unity so that the final table can be printed to give information about the gradient.

3. SPECIAL OPTIONS

3.1 Correlation Matrix and Table of Variance-Decomposition Proportions

A correlation matrix for the independent variables (excluding the constant) and the dependent variable in terms of the original variables $(X$ and $Y)$ is always given before the maximization procedure begins. Another correlation matrix in terms of the transformed variables (X^{**} and Y^{**}) is also computed at the maximum of the log-likelihood function.

The matrices are stored and output in a lower triangular form where the last row represents the pairwise correlations between the dependent variable and each of the independent variables.

To detect the presence of multiple linear dependencies among the original independent variables X, the spectral decomposition of $X^{'}X$ is used (Judge et al., 1985). This method is similar to the singular value decomposition of the matrix X given in Belsley et al. (1980). The analysis is also performed for the transformed variables X^{**} at the maximum of the log-likelihood function.

The spectral decomposition of $X^{'}X$ is defined as:

$$X^{'}X \ = \ \sum_{i=1}^{k} \gamma_i p_i p_{i,}^{'} \tag{88}$$

where p_i is the $(K \times 1)$ eigenvector associated with the i-th eigenvalue γ_i of $X^{'}X$ and the columns of X are scaled to unit length but not centered around their sample means, because centering obscures any linear dependency that involves the constant term.

Belsley et al. use a set of condition indexes which is a generalization of the concept of the condition number of a matrix to detect the presence of near dependencies among the columns of X:

— **Condition number** of X: $\kappa(X) \ = \ (\gamma_{max}/\gamma_{min})^{1/2}$, where γ_{max} and γ_{min} are respectively the greatest and smallest eigenvalues of the γ_i's. This number measures the sensitivity of b to changes in $X^{'}X$ or $Y^{'}Y$ in linear systems represented by the normal equations $X^{'}Xb \ = \ X^{'}Y$;
— **Condition indexes**: $\eta_i = \ (\gamma_{max}/\gamma_i)^{1/2}, i \ = \ 1, ..., K$. Note that if γ_i is equal to γ_{min}, then η_i has a maximum value which corresponds to the condition number of X: $\gamma_{max} \ = \ \kappa(X)$.

To determine which variables are involved in each near dependency, a decomposition of the variance of $b_k(k \ = \ 1, ..., K)$ is performed:

$$\text{Var}(b_k) \ = \ \sigma^2 \sum_{j=1}^{K} \left(p_{kj}^2/\gamma_j\right) \tag{89}$$

The proportion of $Var(b_k)$ associated with any γ_i is then computed:

$$\Pi_{ik} \ = \ \left(p_{ki}^2/\gamma_i\right) / \sum_{j=1}^{K} \left(p_{kj}^2/\gamma_j\right) \tag{90}$$

These results can be summarized in a table of variance-decomposition proportions where the elements in each column are reordered according to the increasing values of the γ_i's.

TABLE 5 Eigenvalues of $X'X$, Condition Indexes of X and Proportions of Var(b_k).

EIGENVALUE	CONDITION INDEX	VARIANCE-DECOMPOSITION PROPORTIONS			
		Var(b_1)	Var(b_2)	. . .	Var(b_K)
γ_1 ($\gamma_{\min}$)	$\eta_1 = \kappa$(X)	Π_{11}	Π_{12}	. . .	Π_{1K}
γ_2	η_2	Π_{21}	Π_{22}	. . .	Π_{2K}
.	.	.	.	. . .	.
.	.	.	.	. . .	.
.	.	.	.	. . .	.
γ_K ($\gamma_{\max}$)	$\eta_K = 1$	Π_{K1}	Π_{K2}	. . .	Π_{KK}

The sum of the proportions Π_{ik}'s in each column associated with Var(b_k) is equal to one. The following rules of thumb can be used to detect the presence of near dependencies:

1. High values of the condition indexes $(\eta_i > 30)$ signal the existence of near dependencies, while high associated Π_{ik}'s in excess of 0.5 indicate which variable X_k is involved in the collinear relations.
2. When a given variable X_k is involved in several collinear relations, its proportions Π_{ik}'s can be individually small across the high η_i's. In this case, the sum of these proportions which is in excess of 0.5 also diagnose variable involvement.

3.2 Analysis of Heteroskedasticity of the Residuals

At the initial and final steps of the maximization of the log-likelihood function, the estimated residuals of u_t, v_t and w_t are plotted around their means, against the observations ordered by increasing values of one selected independent variable to provide a graphical analysis of the functional form of heteroskedasticity related to this variable.

The residuals in standard units are plotted to scale on the horizontal axis, whereas on the vertical axis, the independent variable is not, due to the large number of observations: one line represents an observation, irrespective of the real distance between two consecutive observations.

3.3 Analysis of Autocorrelation of the Residuals

At the initial and final steps of the maximization of $\bar{L}$, autocorrelation and partial autocorrelation functions are estimated for a time-series $\widetilde{w}_t$ that is produced by a differencing operation applied to the estimated residuals $\widehat{e}_t$'s which can be the $\widehat{u}_t$'s, $\widehat{v}_t$'s or $\widehat{w}_t$'s depending on which model type is specified:

$$\widetilde{w}_t = (1-B)^d(1-B^s)^D\widehat{e}_t \ , \ (t = 1+r, \ldots, n) \tag{91}$$

where B is the backward shift operator defined as $B\widehat{e}_t = \widehat{e}_{t-1}$, hence $B^m\widehat{e}_t = \widehat{e}_{t-m}$, d is the degree of consecutive differencing $(0 \le d \le 3)$, D is the degree of seasonal differencing $(0 \le D \le 3)$ and s is the period of seasonal differencing $(1 \le s \le 31)$. An additional constraint for d, D and s is $d + sD \le n - r$. For example, if only the residuals $\widehat{e}_t$'s are to be analyzed, but not the higher order differences produced by consecutive and/or seasonal differencing, then d and D should be set at 0 and s which is not relevant in this case can be set at any positive integer value.

The following estimates are computed for the autocorrelation and partial autocorrelation functions of $\widetilde{w}_t$. For a complete description and use of these functions, see Box and Jenkins (1976).

•Estimate of the Autocorrelation Function

The autocorrelation function is extremely useful in providing a partial description of a stochastic process, i.e. a measure of how much correlation, hence interdependency, exists between neighboring data points in a time-series.

For a normal stationary process $\widetilde{w}_t$, the autocorrelation function with lag k is defined as:

$$\widetilde{p}_k = \frac{E\left[\left(\widetilde{w}_t - \mu_{\widetilde{w}}\right)\left(\widetilde{w}_{t+k} - \mu_{\widetilde{w}}\right)\right]}{\sqrt{E\left[\left(\widetilde{w}_t - \mu_{\widetilde{w}}\right)^2\right]E\left[\left(\widetilde{w}_{t+k} - \mu_{\widetilde{w}}\right)^2\right]}} = \frac{Cov(\widetilde{w}_t, \widetilde{w}_{t+k})}{\sigma^2_{\widetilde{w}}} = \frac{\widetilde{\gamma}_k}{\widetilde{\gamma}_0}, \left(k = 0, 1, ..., \widetilde{K}\right) \tag{92}$$

where the mean and variance of $\widetilde{w}_t$ are respectively $\mu_{\widetilde{w}} = E(\widetilde{w}_t) = E(\widetilde{w}_{t+k})$ and $\sigma^2_{\widetilde{w}} = E\left[\left(\widetilde{w}_t - \mu_{\widetilde{w}}\right)^2\right] = E\left[\left(\widetilde{w}_{t+k} - \mu_{\widetilde{w}}\right)^2\right]$, and the autocovariance at lag k is $\widetilde{\gamma}_k = Cov(\widetilde{w}_t, \widetilde{w}_{t+k}) = Cov(\widetilde{w}_{t+m}, \widetilde{w}_{t+m+k})$ for any t, k and m. Note that $\widetilde{\rho}_0 = 1$ and $\widetilde{\gamma}_0 = \sigma^2_{\widetilde{w}}$.

The autocorrelation function $\{\widetilde{\rho}_k\}$ is dimensionless, i.e. independent of the scale of measurement of the time-series. Since $\widetilde{\gamma}_{\ } > |\widetilde{\gamma}_k|$ $(k = 1, 2, ...)$, all the $\widetilde{\rho}_k$'s lie between -1 and 1. It is symmetric about zero, that is $\widetilde{\rho}_k = \widetilde{\rho}_{-k}$, hence considering the positive half of the function $(k > 0)$ is sufficient to analyze the time-series. Since $\widetilde{\gamma}_k = \widetilde{\rho}_k\widetilde{\gamma}_0 = \widetilde{\rho}_k\sigma^2_{\widetilde{w}}$, a normal stationary process $\widetilde{w}_t$ is completely characterized by its mean $\mu_{\widetilde{w}}$ and autocovariance function $\{\widetilde{\gamma}_k\}$, or equivalently by its mean $\mu_{\widetilde{w}}$, variance $\sigma^2_{\widetilde{w}}$ and autocorrelation function $\{\widetilde{\rho}_k\}$.

An estimate of the autocorrelation function, called the **sample autocorrelation function**, can be computed as:

$$\widetilde{r}_k = \frac{c_k}{c_0}, \left(k = 0, 1, ..., \widetilde{K}\right) \tag{93}$$

where c_k is the estimate of the autocovariance function at lag k:

$$c_k = \sum_{t=1+r}^{n-k} \left(\widetilde{w}_t - \bar{\widetilde{w}}\right)\left(\widetilde{w}_{t+k} - \bar{\widetilde{w}}\right)/(n-r), \tag{94}$$

$\bar{\widetilde{w}}$ is the sample mean of $\widetilde{w}_t$ computed as $\sum\limits_{t=1+r}^{n} \widetilde{w}_t/(n-r)$ and c_0 is the sample variance of $\widetilde{w}_t$ computed as $\sum\limits_{t=1+r}^{n} \left(\widetilde{w}_t - \bar{\widetilde{w}}\right)^2/(n-r)$. Note that for long time-series, the sample autocorrelation function will closely approximate the true population autocorrelation function, but for small samples, e.g. less than 50 observations, it will be biased downward from the latter.

•Standard Error of the Autocorrelation Estimate

To check whether the theoretical autocorrelation $\widetilde{\rho}_k$ is effectively zero beyond a certain lag q, Bartlett's (1946) approximation for the variance of the estimated autocorrelation coefficient of a normal stationary process can be used:

$$\text{Var}(\widetilde{r}_k) \simeq \frac{1}{N}\sum_{\upsilon=-\infty}^{+\infty} \left(\widetilde{\rho}^2_\upsilon + \widetilde{\rho}_{\upsilon+k}\widetilde{\rho}_{\upsilon-k} - 4\widetilde{\rho}_k\widetilde{\rho}_\upsilon\widetilde{\rho}_{\upsilon-k} + 2\widetilde{\rho}^2_\upsilon\widetilde{\rho}^2_k\right). \tag{95}$$

If all the autocorrelations $\widetilde{\rho}_\upsilon$ are zero for $\upsilon > q$, all terms except the first between the parentheses are zero when $k > q$. Thus at lags k greater than q, the variance of the sample autocorrelations $\widetilde{r}_k$ can be expressed as:

$$Var(\widetilde{r}_k) \simeq \frac{1}{N}\left(1 + 2\sum_{\upsilon=1}^{q}\widetilde{\rho}_\upsilon^2\right)\ ,\ (k > q)\ . \tag{96}$$

To obtain an estimate of Var($\widetilde{r}_k$), say $\widehat{Var}(\widetilde{r}_k)$, the sample autocorrelations $\widetilde{r}_\upsilon$ $(\upsilon = 1, 2, ..., q)$ are substituted for $\widetilde{\rho}_\upsilon$. The square root of this estimate is referred as the **large-lag standard error**.

For example, if $q = 0$, i.e. the series $\widetilde{w}_t$ is assumed to be completely random, then for all lags, the large-lag standard error reduces to:

$$\sqrt{\widehat{Var}(\widetilde{r}_k)} \simeq \frac{1}{\sqrt{N}} \tag{97}$$

which is printed next to each row of the sample autocorrelations $\widetilde{r}_k$ and can be used in a first step to test the null hypothesis that $\widetilde{\rho}_k = 0,\ (k = 1, 2, ...)$.

A second step would be to select the first lag q at which an autocorrelation coefficient was significant and then use (95) to test for significant autocorrelations at lags k longer than q. For example, if $\widetilde{r}_1$ were significant, $\widetilde{r}_2, \widetilde{r}_3, ...$ could be tested using the standard error $\left[\left(1 + 2\widetilde{r}_1^2\right)/N\right]^{1/2}$.

For moderate N, the distribution of an estimated autocorrelation coefficient, whose theoretical value is zero, is approximately Normal (Anderson, 1942): on the hypothesis that $\widetilde{\rho}_k$ is zero, the estimate $\widetilde{r}_k$ divided by its standard error will follow approximately a unit Normal distribution.

•Estimate of the Partial Autocorrelation Function

Whereas the sample autocorrelation function is used as a first guess which is certainly not conclusive for the significance of each autocorrelation coefficient, the partial autocorrelation function $\phi_{\ell\ell}$ which is based on the fact that for an autoregressive process of order p which has an autocorrelation function infinite in extent, $\phi_{\ell\ell}$ is nonzero for $\ell \le p$ and zero for $\ell > p$, i.e. it has a cutoff after lag p, provides a means for choosing which order of autoregressive process has to be fit to an observed time-series.

Let $\phi_{\ell j}$ be the jth coefficient in an autoregressive process of order ℓ so that $\phi_{\ell\ell}$ is the last coefficient. The $\phi_{\ell j}$'s can be shown to satisfy the set of equations:

$$\widetilde{\rho}_j = \phi_{\ell 1}\widetilde{\rho}_{j-1} + \phi_{\ell 2}\widetilde{\rho}_{j-2} + ... + \phi_{\ell\ell}\widetilde{\rho}_{j-\ell}\ ,\ (j = 1, 2, ..., \ell) \tag{98}$$

leading to the Yule-Walker equations:

$$\begin{bmatrix} 1 & \widetilde{\rho}_1 & \widetilde{\rho}_2 & ... & \widetilde{\rho}_{\ell-1} \\ \widetilde{\rho}_1 & 1 & \widetilde{\rho}_1 & ... & \widetilde{\rho}_{\ell-2} \\ \vdots & \vdots & \vdots & ... & \vdots \\ \widetilde{\rho}_{\ell-1} & \widetilde{\rho}_{\ell-2} & \widetilde{\rho}_{\ell-3} & ... & 1 \end{bmatrix} \begin{bmatrix} \phi_{\ell 1} \\ \phi_{\ell 2} \\ \vdots \\ \phi_{\ell\ell} \end{bmatrix} = \begin{bmatrix} \widetilde{\rho}_1 \\ \widetilde{\rho}_2 \\ \vdots \\ \widetilde{\rho}_\ell \end{bmatrix}. \tag{99}$$

The coefficient $\phi_{\ell\ell}$, regarded as a function of the lag ℓ, is called the **partial autocorrelation function**. It measures the correlation between $\widetilde{w}_t$ and $\widetilde{w}_{t-\ell}$ given $\widetilde{w}_{t-1}, \ldots, \widetilde{w}_{t-(\ell-1)}$. The estimated partial autocorrelations $\widehat{\phi}_{\ell j}$'s can be obtained by substituting the sample autocorrelations $\widetilde{r}_j$'s for the $\widetilde{\rho}_j$'s in (97) to yield:

$$\widetilde{r}_j = \widehat{\phi}_{\ell 1}\widetilde{r}_{j-1} + \widehat{\phi}_{\ell 2}\widetilde{r}_{j-2} + \cdots + \widehat{\phi}_{\ell\ell}\widetilde{r}_{j-l}\ ,\ (j = 1, 2, \ldots, \ell) \tag{100}$$

and solving the resulting Yule-Walker equations.

A simple recursive method for estimating the $\widehat{\phi}_{\ell j}$'s, due to Durbin (1960), may be used if the values of the parameters are not too close to the nonstationary boundaries:

$$\widehat{\phi}_{\ell\ell} = \begin{cases} \widetilde{r}_1 & \text{if } \ell = 1, \\ \dfrac{\widetilde{r}_\ell - \sum\limits_{j=1}^{\ell-1} \widehat{\phi}_{\ell-1,j}\widetilde{r}_{\ell-j}}{1 - \sum\limits_{j=1}^{\ell-1} \widehat{\phi}_{\ell-1,j}\widetilde{r}_j} & \text{if } \ell = 2, 3, \ldots, \widetilde{L}. \end{cases} \tag{101}$$

where $\widehat{\phi}_{\ell j} = \widehat{\phi}_{\ell-1,j} - \widehat{\phi}_{\ell\ell}\widehat{\phi}_{\ell-1,\ell-j}$ $(j = 1, 2, \ldots, \ell - 1)$ and $\widetilde{L}$ is the maximum lag of the partial autocorrelation function. Note that $\widetilde{L}$ should be less than or equal to $\widetilde{K}$ which is the maximum lag of the autocorrelation function used in (92).

•Standard Error of the Partial Autocorrelation Estimate

If the process is autoregressive of order p, the estimated partial autocorrelations of order $p + 1$, and higher, are approximately independently distributed (Quenouille, 1949), and the standard error for these autocorrelations is:

$$\sqrt{\widehat{\text{Var}}\left(\widehat{\phi}_{\ell\ell}\right)} \simeq \frac{1}{\sqrt{N}} \;,\; (\ell > p) \tag{102}$$

which can be used to check whether the partial autocorrelations are effectively zero after some specific lag p. Note that on the hypothesis that $\phi_{\ell\ell}$ is zero, the estimate $\widehat{\phi}_{\ell\ell}$ divided by its standard error is approximately distributed as a unit Normal deviate.

3.4 Forecasting: Maximum Likelihood and Simulation Forecasts

•Maximum Likelihood Forecast

To obtain the predicted values of the dependent variable over the p periods for $t = n + 1, \ldots, n + p,$, the following approach is adopted: for the first period of forecast $n + 1$, the maximum likelihood predicted value of Y_{n+1} is given by the first order condition $\partial L_{n+1}/\partial Y_{n+1} = 0$, since maximizing the total log-likelihood function $\sum\limits_{t=1+r}^{n+1} L_t = \sum\limits_{t=1+r}^{n} L_t + L_{n+1}$ with respect to Y_{n+1} is equivalent to maximizing the log-likelihood function L_{n+1} associated with the $(n + 1)$-th period only.

The first derivative of the log-likelihood function L_{n+1} with respect to Y_{n+1} can be written as:

$$\frac{\partial L_{n+1}}{\partial Y_{n+1}} = \begin{cases} \dfrac{1}{Y_{n+1}}\left[\lambda_y - 1 - \dfrac{Y_{n+1}}{\sigma_w^2 f(Z_{n+1})^{1/2}}\left(\dfrac{u_{n+1}}{f(Z_{n+1})^{1/2}} - \sum\limits_\ell \rho_\ell \dfrac{u_{n+1-\ell}}{f(Z_{n+1-\ell})^{1/2}}\right)\right] & \text{if } \lambda_y \neq 0 \\ \dfrac{1}{Y_{n+1}}\left[-1 - \dfrac{1}{\sigma_w^2 f(Z_{n+1})^{1/2}}\left(\dfrac{u_{n+1}}{f(Z_{n+1})^{1/2}} - \sum\limits_\ell \rho_\ell \dfrac{u_{n+1-\ell}}{f(Z_{n+1-\ell})^{1/2}}\right)\right] & \text{if } \lambda_y = 0 \end{cases} \tag{103}$$

where $u_{n+1} = Y_{n+1}^{(\lambda_y)} - \sum\limits_k \beta_k X_{k,n+1}^{(\lambda_{xk})}$.

• $\boxed{\boldsymbol{\lambda_y \neq 0}}$

Two subcases must be distinguished:

— If $\lambda_y \neq 1$, setting the derivative equal to zero and solving for $Y_{n+1}^{\lambda_y}$, yield an equation of the second degree in $Y_{n+1}^{\lambda_y}$:

$$Y_{n+1}^{2\lambda_y} + BY_{n+1}^{\lambda_y} + C = 0 \tag{104}$$

where

$$B = -\left[1 + \lambda_y f(Z_{n+1})^{1/2}\left(\sum_k \beta_k \frac{X_{k,n+1}^{(\lambda_{xk})}}{f(Z_{n+1})^{1/2}} + \sum_\ell \rho_\ell \frac{u_{n+1-\ell}}{f(Z_{n+1-\ell})^{1/2}}\right)\right] \tag{105}$$

and

$$C = -\lambda_y(\lambda_y - 1)\sigma_w^2 f(Z_{n+1}). \tag{106}$$

To ensure that the predicted value of Y_{n+1}, say $\widehat{Y}_{n+1}$, is real and positive, both roots of (104), namely $Y_{n+1}^{\lambda_y} = \left(-B \pm \sqrt{B^2 - 4C}\right)/2$, must be real, i.e. if $B^2 - 4C \geq 0$, and the greater root will be chosen since it will be always positive as shown in the next section. Finally, the chosen root must also satisfy the second order condition $\partial^2 L_{n+1}/\partial Y_{n+1}^2 < 0$.

— If $\lambda_y = 1$, i.e. the dependent variable is specified in a linear form, the first derivative reduces to:

$$\frac{\partial L_{n+1}}{\partial Y_{n+1}} = -\frac{1}{\sigma_w^2 f(Z_{n+1})^{1/2}}\left(\frac{u_{n+1}}{f(Z_{n+1})^{1/2}} - \sum_\ell \rho_\ell \frac{u_{n+1-\ell}}{f(Z_{n+1-\ell})^{1/2}}\right) \tag{107}$$

which gives a unique maximum likelihood predicted value of Y_{n+1} by setting the derivative equal to zero and solving for Y_{n+1}:

$$\widehat{Y}_{n+1} = 1 + \sum_k \beta_k X_{k,n+1}^{(\lambda_{xk})} + f(Z_{n+1})^{1/2} \sum_\ell \rho_\ell \frac{u_{n+1-\ell}}{f(Z_{n+1-\ell})^{1/2}}\,. \tag{108}$$

• $\boxed{\boldsymbol{\lambda_y = 0}}$

Setting the first derivative equal to zero and solving for Y_{n+1} also give a unique maximum likelihood predicted value of Y_{n+1}:

$$\widehat{Y}_{n+1} = \exp\left[\sum_k \beta_k X_{k,n+1}^{(\lambda_{xk})} + f(Z_{n+1})^{1/2} \sum_\ell \rho_\ell \frac{u_{n+1-\ell}}{f(Z_{n+1-\ell})^{1/2}} - \sigma_w^2 f(Z_{n+1})\right], \tag{109}$$

Note that for all cases, the predicted value $\widehat{Y}_{n+1}$is computed at $\Pi = \widehat{\Pi}$.

Clearly, the approach can be easily extended to the next forecast periods, since in each single period, only the log-likelihood function associated with that period is needed for the maximization, given the sample observed values of the dependent variable in the estimation period and its predicted values in the preceding forecast periods. Thus for the i-th period of forecast, the predicted value of Y_{n+i} is also given by the first order condition $\partial L_{n+i}/\partial Y_{n+i} = 0$, subject to the second order condition $\partial^2 L_{n+i}/\partial Y_{n+i}^2 < 0$. The formulas are analogous to (103) – (109) with $n+1$ replaced by $n+i$.

•Variance of the Forecast Error

The variance of the forecast error for the first period of forecast, σ_{n+1}^2, is computed from the covariance matrix of the parameters Π and Y_{n+1}, which is equal to minus the inverse matrix of the second derivatives of the total log-likelihood function $L_{(1)} = \Sigma_{t=1+r}^{n+1} L_t$ with respect to Π and Y_{n+1}, evaluated at $\Pi = \widehat{\Pi}$ and $Y_{n+1} = \widehat{Y}_{n+1}$:

$$COV_{(1)} = \begin{bmatrix} COV_{(0)} & C_{(1)} \\ C'_{(1)} & \sigma^2_{n+1} \end{bmatrix} = -\begin{bmatrix} \partial^2 L_{(1)}/\partial\Pi\partial\Pi' & \partial^2 L_{(1)}/\partial\Pi\partial Y_{n+1} \\ \partial^2 L_{(1)}/\partial Y_{n+1}\partial\Pi' & \partial^2 L_{(1)}/\partial Y^2_{n+1} \end{bmatrix}^{-1} \tag{110}$$

where

1. $COV_{(0)}$ is the covariance matrix of $\widehat{\Pi}$ which is already estimated in (38) and is used instead of $-\left[\partial^2 L_{(1)}/\partial\Pi\partial\Pi'\right]^{-1}$ which is too complex to be evaluated,
2. $C_{(1)}$ is the column vector of covariances between the elements of $\widehat{\Pi}$ and $\widehat{Y}_{n+1}$:

$$C_{(1)} = -COV_{(0)}\left[\frac{\partial^2 L_{(1)}}{\partial\Pi\partial Y_{n+1}}\right]\left[\frac{\partial^2 L_{(1)}}{\partial Y^2_{n+1}}\right]^{-1} \tag{111}$$

3. and σ^2_{n+1} is the variance of the forecast error for Y_{n+1}:

$$\sigma^2_{n+1} = -\left[\frac{\partial^2 L_{(1)}}{\partial Y^2_{n+1}}\right]^{-1} + C'_{(1)} COV^{-1}_{(0)} C_{(1)}\ . \tag{112}$$

In general, the variance of the forecast error for the i-th period of forecast, σ^2_{n+i}, can be recursively computed from:

$$COV_{(i)} = \begin{bmatrix} COV_{(i-1)} & C_{(i)} \\ C'_{(i)} & \sigma^2_{n+1} \end{bmatrix} = -\begin{bmatrix} \partial^2 L_{(i)}/\partial\Pi_{(i-1)}\partial\Pi'_{(i-1)} & \partial^2 L_{(i)}/\partial\Pi_{(i-1)}\partial Y_{n+i} \\ \partial^2 L_{(i)}/\partial Y_{n+i}\partial\Pi'_{(i-1)} & \partial^2 L_{(i)}/\partial Y^2_{n+i} \end{bmatrix}^{-1} \tag{113}$$

where

1. COV_{i-1} is the covariance matrix of $\widehat{\Pi}_{(i-1)} = \left(\widehat{\Pi}', \widehat{Y}_{n+1}, \ldots, \widehat{Y}_{n+i-1}\right)'$ which is already estimated in the preceding period of forecast, and is used instead of $-\left[\partial^2 L_{(i)}/\partial\Pi_{(i-1)}\partial\Pi'_{(i-1)}\right]^{-1}$ which is too complex to be evaluated,
2. $C_{(i)}$ is the column vector of covariances between the elements of $\widehat{\Pi}_{(i-1)}$ and $\widehat{Y}_{n+i}$:

$$C_{(i)} = -COV_{(i-1)}\left[\frac{\partial^2 L_{(i)}}{\partial\Pi_{(i-1)}\partial Y_{n+i}}\right]\left[\frac{\partial^2 L_{(i)}}{\partial Y^2_{n+i}}\right]^{-1} \tag{114}$$

3. and σ^2_{n+i} is the variance of the forecast error for Y_{n+i}:

$$\sigma^2_{n+i} = -\left[\frac{\partial^2 L_{(i)}}{\partial Y^2_{n+i}}\right]^{-1} + C'_{(i)} COV^{-1}_{(i-1)} C_{(i)}\ . \tag{115}$$

Note that $\partial^2 L_{(i)}/\partial\Pi_{(i-1)}\partial Y_{n+i} = \partial L_{n+i}/\partial\Pi_{(i-1)}\partial Y_{n+i}$ and $\partial^2 L_{(i)}/\partial Y^2_{n+i} = \partial^2 L_{n+i}/\partial Y^2_{n+i}$, since Y_{n+i} appears only in L_{n+i}. The second derivatives $\partial^2 L_{n+i}/\partial\Pi_{(i-1)}\partial Y_{n+i}$ and $\partial^2 L_{n+i}/\partial Y^2_{n+i}$ are in the form $Y^{\lambda-1}_{n+i}/\left[\sigma^2_w f(Z_{n+i})^{1/2}\right]$ times a component which is given in TABLE 6 for each element of $\Pi_{(i)} = \left(\Pi'_{(i-1)}, Y_{n+i}\right)'$ crossed with Y_{n+i}.

TABLE 6 Components of the second derivatives of L_{n+i} with respect to $\Pi_{(i)}$ and Y_{n+i}

$\Pi_{(i)}$	Y_{n+i}
β_k	$\dfrac{X_{k,n+i}^{(\lambda_{xk})}}{f(Z_{n+i})^{1/2}} - \sum_\ell \rho_\ell \dfrac{X_{k,n+i-\ell}^{(\lambda_{xk})}}{f(Z_{n+i-\ell})^{1/2}}$
σ_w^2	w_{n+i}/σ_w^2
λ_y	$\dfrac{\sigma_w^2 f(Z_{n+i})^{1/2}}{Y_{n+i}^{\lambda_y}} - w_{n+i}\ln Y_{n+i} - \left[\dfrac{1}{f(Z_{n+i})^{1/2}}\dfrac{\partial Y_{n+i}^{(\lambda_y)}}{\partial\lambda_y} - \sum_\ell \dfrac{\rho_\ell}{f(Z_{n+i-\ell})^{1/2}}\dfrac{\partial Y_{n+i-\ell}^{(\lambda_y)}}{\partial\lambda_y}\right]$
λ_{xk}	$\beta_k\left[\dfrac{1}{f(Z_{n+i})^{1/2}}\dfrac{\partial X_{k,n+i}^{(\lambda_{xk})}}{\partial\lambda_{xk}} - \sum_\ell \dfrac{\rho_\ell}{f(Z_{n+i-\ell})^{1/2}}\dfrac{\partial X_{k,n+i-\ell}^{(\lambda_{xk})}}{\partial\lambda_{xk}}\right]$
λ_{zm}	$\dfrac{\delta_m}{2}\left[\left(\dfrac{2u_{n+i}}{f(Z_{n+i})^{1/2}} - \sum_\ell \rho_\ell \dfrac{u_{n+i-\ell}}{f(Z_{n+i-\ell})^{1/2}}\right)\dfrac{\partial Z_{m,n+i}^{(\lambda_{zm})}}{\partial\lambda_{zm}} - \sum_\ell \rho_\ell \dfrac{u_{n+i-\ell}}{f(Z_{n+i-\ell})^{1/2}}\dfrac{\partial Z_{m,n+i-\ell}^{(\lambda_{zm})}}{\partial\lambda_{zm}}\right]$
δ_m	$\dfrac{1}{2}\left[\left(\dfrac{2u_{n+i}}{f(Z_{n+i})^{1/2}} - \sum_\ell \rho_\ell \dfrac{u_{n+i-\ell}}{f(Z_{n+i-\ell})^{1/2}}\right) Z_{m,n+i}^{(\lambda_{zm})} - \sum_\ell \rho_\ell \dfrac{u_{n+i-\ell}}{f(Z_{n+i-\ell})^{1/2}} Z_{m,n+i-\ell}^{(\lambda_{zm})}\right]$
ρ_ℓ	$\dfrac{Y_{n+i-\ell}^{(\lambda_y)} - \Sigma_k \beta_k X_{k,n+i-\ell}^{(\lambda_{xk})}}{f(Z_{n+i-\ell})^{1/2}}$
Y_{n+i-s}	$\dfrac{\rho_s}{f(Z_{n+i-s})^{1/2}} Y_{n+i-s}^{(\lambda_y-1)}$ $\quad (s = 1, \ldots, \min(i-1, r)$ and $i \geq 2)$
Y_{n+i}	$-\left(2Y_{n+i}^{(\lambda_y)} + B\right)/\left[Y_{n+i} f(Z_{n+i})^{1/2}\right]$

The different derivatives $\partial Y_{n+i}^{(\lambda_y)}/\partial\lambda_y$, $\partial X_{k,n+i}^{(\lambda_{xk})}/\partial\lambda_{xk}$, and $\partial Z_{m,n+i}^{(\lambda_{zm})}/\partial\lambda_{zm}$ as well as their lagged expressions are computed by the generic formula given in (36).

In evaluating the second derivatives of L_{n+i} at the maximum point of the total log-likelihood function $L_{(i)} = \Sigma_{t=1+r}^{n+i} L_t$, the predicted values of $(Y_{n+1}, \ldots, Y_{n+i})$ obtained by the approach above are used just as when the predicted value of Y_{n+i} is computed, the predicted values of $(Y_{n+1}, \ldots, Y_{n+i-1})$ are also used.

The error of forecast is computed in percentage whenever the observed value of the dependent variable is available in the forecast periods $t = n+1, \ldots, n+p$:

$$\widehat{e}_t = \left(1 - \frac{\widehat{Y}}{Y_t^0}\right) 100 \tag{116}$$

where $\widehat{Y}_t$ and Y_t^0 are respectively the predicted and observed values of Y_t.

The 95% confidence interval for Y_t is also computed:

$$\widehat{Y}_t - 1.96\widehat{\sigma}_t \leq Y_t \leq \widehat{Y}_t + 1.96\widehat{\sigma}_t \tag{117}$$

where $\widehat{\sigma}_t$ is estimated from (115).

•Simulation Forecast

In addition to the maximum likelihood predicted value $\widehat{Y}_t \;\; (t = n+1, \ldots, n+p)$, the simulated value of the dependent variable $\widetilde{Y}_t$ is computed at $\Pi = \widehat{\Pi}$ as:

$$\widetilde{Y}_t = \begin{cases} \left[1 + \lambda_y f(Z_t)^{1/2} \left(\Sigma_\ell \rho_\ell Y_{t-\ell}^* + \Sigma_k \beta_k X_{kt}^{**}\right)\right]^{1/\lambda_y} & \text{if } \lambda_y \neq 0 \\ \exp\left[f(Z_t)^{1/2} \left(\Sigma_\ell \rho_\ell \frac{lnY_{t-\ell}}{f(Z_{t-\ell})^{1/2}} + \Sigma_k \beta_k X_{kt}^{**}\right)\right] & \text{if } \lambda_y = 0 \end{cases} \tag{118}$$

where $Y_{t-\ell}$ is replaced by $\widetilde{Y}_{t-\ell}$ if $Y_{t-\ell}$ is not observed, i.e. if $t - \ell > n$. Note that for the case $\lambda_y \neq 0$, if the expression between the squared brackets happens to be negative, then it cannot be raised to the power $1/\lambda_y$ and the program will be stopped at that point of the run.

•$\boxed{\lambda_y \neq 0}$

$\boxed{\textbf{Subcase } \lambda_y \neq 1}$: If both $\widetilde{Y}_t$ and $\widehat{Y}_t$ are positive, the bias $\widetilde{Y}_t - \widehat{Y}_t$, or equivalently $\widetilde{Y}_t^{\lambda_y} - \widehat{Y}_t^{\lambda_y}$, can be shown to be negative if $\lambda_y < 0$ or $\lambda_y > 1$ and positive if $0 < \lambda_y < 1$.

For the first period of forecast $t = n+1$, if the two roots of (104) are denoted by $\widehat{Y}_t^{\lambda_y} = \left(-B + \sqrt{B^2 - 4C}\right)/2$ and $\widehat{\widehat{Y}}_t^{\lambda_y} = \left(-B - \sqrt{B^2 - 4C}\right)/2$, then replacing $-B$ by its value which is equal to $\widetilde{Y}_t^{\lambda_y}$ in the greater root $\widehat{Y}_t^{\lambda_y}$ yields:

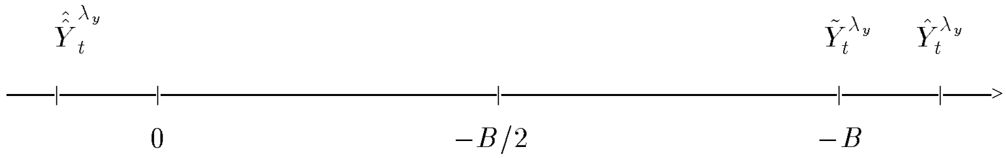
$$\widehat{Y}_t^{\lambda_y} = \widetilde{Y}_t^{\lambda_y} \left(1 + \sqrt{1 - 4C/\widetilde{Y}_t^{2\lambda_y}}\right)/2\ . \tag{119}$$

If$\lambda_y < 0$ or $\lambda_y > 1$, then C is negative, $\sqrt{1 - 4C/\widetilde{Y}_t^{2\lambda_y}} > 1$ and the bias $\widetilde{Y}_t^{\lambda_y} - \widehat{Y}_t^{\lambda_y}$ is negative.

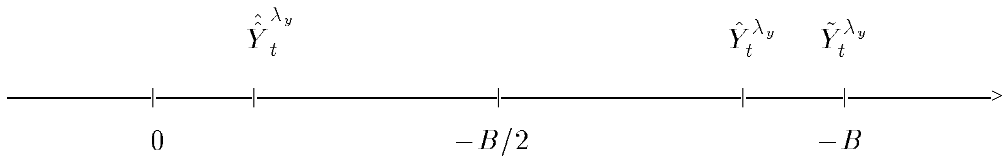
If $0 < \lambda_y < 1$, then C is positive, $\sqrt{1 - 4C/\widetilde{Y}_t^{2\lambda_y}} < 1$ and the bias $\widetilde{Y}_t^{\lambda_y} - \widehat{Y}_t^{\lambda_y}$ is positive.

Furthermore, since $\widetilde{Y}_t^{\lambda_y} = \widehat{Y}_t^{\lambda_y} + \widehat{\widehat{Y}}_t^{\lambda_y}$, the bias $\widetilde{Y}_t^{\lambda_y} - \widehat{Y}_t^{\lambda_y}$ will be given by the smaller root $\widehat{\widehat{Y}}_t^{\lambda_y}$. These results can be illustrated by the following diagrams:

$\boxed{\text{Negative bias}}$: $\tilde{Y}_t^{\lambda_y} < \hat{Y}_t^{\lambda_y}$ ($\lambda_y < 0$ or $\lambda_y > 1$)



$\boxed{\text{Positive bias}}$: $\tilde{Y}_t^{\lambda_y} > \hat{Y}_t^{\lambda_y}$ $(0 < \lambda_y < 1)$



Choosing the greater root $\widehat{Y}_t^{\lambda_y}$ will ensure that the maximum likelihood predicted value of Y_t is always positive since the smaller root $\widehat{\widehat{Y}}_t^{\lambda_y}$ can be negative. These results will hold for every period of forecast $t = n+1, \ldots, n+p$.

$\boxed{\textbf{Subcase } \boldsymbol{\lambda_y = 1}}$: In this case, the bias $\widetilde{Y}_t - \widehat{Y}_t$ is null, i.e. the simulated value of the dependent variable $\widetilde{Y}_t$ is identical to the maximum likelihood predicted value $\widehat{Y}_t$ given in (108):

$$\begin{aligned}
\widetilde{Y}_t &= 1 + f(Z_t)^{1/2}\left(\sum_\ell \rho_\ell Y^*_{t-\ell} + \sum_k \beta_k X^{**}_{kt}\right) \\
&= 1 + f(Z_t)^{1/2}\left[\sum_\ell \rho_\ell \left(Y^*_{t-\ell} - \sum_k \beta_k X^*_{k,t-\ell}\right) + \sum_k \beta_k X^*_{kt}\right] \\
&= 1 + \sum_k \beta_k X^{(\lambda x t)}_{kt} + f(Z_t)^{1/2} \sum_\ell \rho_\ell \frac{u_{t-\ell}}{f(Z_{t-\ell})^{1/2}} \\
&= \widehat{Y}_t
\end{aligned} \tag{120}$$

• $\boxed{\boldsymbol{\lambda_y = 0}}$

In this case, the bias $\widetilde{Y}_t - \widehat{Y}_t$ is positive:

$$\begin{aligned}
\widetilde{Y}_t &= \exp\left[f(Z_t)^{1/2}\left(\sum_\ell \rho_\ell \frac{\ln Y_{t-\ell}}{f(Z_{t-\ell})^{1/2}} + \sum_k \beta_k X^{**}_{kt}\right)\right] \\
&= \exp\left\{f(Z_t)^{1/2}\left[\sum_\ell \rho_\ell \left(\frac{\ln Y_{t-\ell}}{f(Z_{t-\ell})^{1/2}} - \sum_k \beta_k X^*_{k,t-\ell}\right) + \sum_k \beta_k X^*_{kt}\right]\right\} \\
&= \exp\left[\sum_k \beta_k X^{(\lambda_{xt})}_{kt} + f(Z_t)^{1/2} \sum_\ell \rho_\ell \frac{u_{t-\ell}}{f(Z_{t-\ell})^{1/2}}\right] \\
&> \widehat{Y}_t
\end{aligned} \tag{121}$$

since $\widehat{Y}_t$ contains in the argument of the exponential (109) an additional negative term equal to $-\sigma_w^2 f(Z_t)$.

4. BIBLIOGRAPHY FOR THE LEVEL CLASS

1. **Abramowitz, M. and I.A. Stegun** (1965),*Handbook of Mathematical Functions*, Dover Publications, New York.
2. **Amemiya, T.** (1977), “A Note on a Heteroscedastic Model", *Journal of Econometrics*, 6, 365–370; and “Corrigenda", *Journal of Econometrics*, 8, 275.
3. **Anderson, R.L.** (1942), “Distribution of the Serial Correlation Coefficient",*Annals of Mathematical Statistics*, 13, 1.
4. **Appelbaum, E.** (1979), “On the Choice of Functional Forms", *International Economic Review*, 20, 449–458.
5. **Bartlett, M.S.** (1946), “On the Theoretical Specification of Sampling Properties of Autocorrelated Time Series", *Journal of the Royal Statistical Society*, Series B, 8, 27.
6. **Belsley, D., E. Kuh and R.E. Welsh** (1980), *Regression Diagnostics*, John Wiley, New York.
7. **Berndt, E.K., B.H. Hall, R.E. Hall and J.A. Hausman** (1974), “Estimation and Inference in Nonlinear Structural Models", *Annals of Economic and Social Measurement*, 3, 653–665.
8. **Berndt, E.K. and M.S. Khaled** (1979), “Parametric Productivity Measurement and Choice among Flexible Functional Form", *Journal of Political Economy*, 87, 1220–1245.
9. **Box, G.P. and D.R. Cox** (1964), “An Analysis of Transformations", *Journal of the Royal Statistical Society*, Series B, 26, 211–243.
10. **Box, G.P. and G.M. Jenkins** (1976),*Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, Revised Edition.
11. **Dagenais, M.G., M.J.I. Gaudry and T.C. Liem** (1980), “Multiple Regression Analysis with Box-Cox Transformation and Nonspherical Residual Errors: A Transportation Application", Publication no. 166, Centre de Recherche sur les Transports, Université de Montréal.
12. **Dagenais, M.G., M.J.I. Gaudry and T.C. Liem** (1987), “Urban Travel Demand: The Impact of Box-Cox Transformations with Nonspherical Residual Errors",*Transportation Research*, 21 B, 6, 443–477.
13. **Dagum, C. and E.B. Dagum** (1974), “Construction de Modèles et Analyse Econométrique", Cahiers de l’Institut des Sciences Mathématiques et Economiques Appliquées, 8, no. 11–12.
14. **Egy, D. and K. Lahiri** (1979), “On Maximum Likelihood Estimation of Functional Form and Heteroscedasticity", *Economics Letters*, 2, 155–159.
15. **Fletcher, R. and M.J.D. Powell** (1963), “A Rapidly Convergent Descent Method for Minimization",*Computer Journal*, 6, 163–168.
16. **Froehlich, B.R.** (1973), “Some Estimators for a Random Coefficient Regression Model", *Journal of the American Statistical Association*, 68, 329–334.
17. **Gaudry, M.J.I. and M.G. Dagenais** (1978), “The Use of Box-Cox Transformations in Regression Models with Heteroskedastic Autoregressive Residuals", Cahier no. 7814, Département de Sciences Economiques, Université de Montréal.
18. **Gaudry, M.J.I. and M.G. Dagenais** (1979), “Heteroscedasticity and the Use of Box-Cox Transformations",*Economic Letters*, 2, 225–229.
19. **Gaudry, M.J.I. and J.J. Wills** (1978), “Estimating the Functional Form of Travel Demand Models",*Transportation Research*, 12, 253–289.
20. **Geary, R.C.** (1966), “A Note on Residual Heterovariance and Estimation Efficiency in Regression",*American Statistician*, 20, (4), 30–31.
21. **Glejser, H.** (1969), “A New Test for Heteroscedasticity", *Journal of the American Statistical Association*, 64, 316–323.
22. **Goldfeld, S.M. and R.E. Quandt** (1972),*Nonlinear Methods in Econometrics*, North-Holland, Amsterdam.
23. **Harvey, A.C.** (1974), “Estimation of Parameters in a Heteroscedastic Regression Model”, European Meeting of the Econometric Society, Grenoble, France.
24. **Harvey, A.C.** (1976), “Estimating Regression Models with Multiplicative Heteroscedasticity",*Econometrica*, 44, 461–465.

25. **Heckman, J. and S. Polachek** (1974), "Empirical Evidence on the Functional Form of the Earnings-Schooling Relationship", *Journal of the American Statistical Association*, 69, 350–354.
26. **Hildreth, C. and J.P. Houck** (1968), "Some Estimators for a Linear Model with Random Coefficients",*Journal of the American Statistical Association*, 63, 584–595.
27. **Hollyer, M., W. Maling and G. Wang** (1979), "Domestic and International Air Cargo Activity", Report no. FAA-AVP-79-10, U.S. Department of Transportation.
28. **Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lütkepohl and T.C. Lee** (1985), *The Theory and Practice of Econometrics*, Second Edition, John Wiley, New York.
29. **Kau, J.B. and C.F. Sirmans** (1976), "The Functional Form of the Gravity Model: A New Technique with Empirical Results", Working paper, Department of Finance, University of Illinois. Presented at the North American Meeting of the Regional Science Association in Toronto.
30. **Kmenta, J.** (1971), *Elements of Econometrics*, Macmillan, New York.
31. **Park, R.E.** (1966), "Estimation with Heteroskedastic Error Terms", *Econometrica*, 34, 888.
32. **Quenouille, M.H.** (1949), "Approximate Tests of Correlation in Time Series",*Journal of the Royal Statistical Society*, Series B, 11, 68.
33. **Savin, N.E. and K.J. White** (1978), "Estimation and Testing for Functional Form and Autocorrelation: A Simultaneous Approach", *Journal of Econometrics*, 8, 1–12.
34. **Schlesselman, J.** (1971), "Power Families: A Note on the Box and Cox Transformation", *Journal of the Royal Statistical Society*, Series B, 33, 307–311.
35. **Spitzer, J.J.** (1976), "The Demand for Money, the Liquidity Trap, and Functional Forms",*International Economic Review*, 17, 220–227.
36. **Spitzer, J.J.** (1977), "A Simultaneous Equations System of Money Demand and Supply Using Generalized Functional Forms", *Journal of Econometrics*, 5, 117–128.
37. **Theil, H.** (1971), *Principles of Econometrics*, John Wiley, New York.
38. **Tobin, J.** (1958), "Estimation of Relationships for Limited Dependent Variables",*Econometrica*, 25, 24–36.
39. **Welland, J.D.** (1976), "Cognitive Abilities, Schooling and Earnings: the Question of Functional Form", Working paper 76-14, Department of Economics, McMaster University, Hamilton.
40. **White, K.J.** (1972), "Estimation of the Liquidity Trap with a Generalized Functional Form",*Econometrica*, 40, 193–199.
41. **Zarembka, P.** (1968), "Functional Form in the Demand for Money", *Journal of the American Statistical Association*, 63, 502–511.
42. **Zarembka, P.** (1974), "Transformation of Variables in Econometrics", in P. Zarembka (ed.), *Frontiers in Econometrics*, Academic Press, New York.