

The Natural Sciences and Engineering Council of Canada (NSERC) and the *Fonds Québécois de Recherche sur la Nature et les Technologies* (FQRNT) have supported this profoundly modified version of the L-1.5 algorithm usually referenced as: Liem, T.C., Gaudry, M., Dagenais, M. and U. Blum, “LEVEL: The L-1.5 program for BC-GAUHESEQ (Box-Cox Generalized Autoregressive Heteroskedastic Single Equation) regression and multimoment analysis”, in Gaudry, M. and S. Lassarre, eds, *Structural Road Accident Models: The International DRAG Family*, Elsevier Science Publishers, Oxford, Ch. 12, 263–324, 2000. This document is complemented by a user guide for the program: Tran, C-L., and Gaudry, M., “L-1.6 Program user guide with databases, examples and Tablex outputs”, Publication AJD-106, Agora Jules Dupuit, Université de Montréal, April 2008.

## **PUBLICATION**

**AJD-105**

***LEVEL : The L-1.6 estimation procedures for BC-GAUHESEQ  
(Box-Cox Generalized Autoregressive Heteroskedastic Single Equation)  
regression and multimoment analysis***

by

**Cong-Liem Tran<sup>1</sup>**

**Marc Gaudry<sup>1,2</sup>**

**Marcel Dagenais<sup>2</sup>**

<sup>1</sup> Agora Jules Dupuit, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Québec, CANADA H3C 3J7 <marc.gaudry@umontreal.ca>

<sup>2</sup> Département de sciences économiques, Université de Montréal

## ABSTRACT

L-1.6 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 first three moments of the dependent variable (expected value, standard error and skewness), marginal rates of substitution (and their elasticities) among the moments, and among the independent variables, 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, Skewness, Marginal Rates of Substitution, Maximum Likelihood Forecast.

## RÉSUMÉ

Le programme L-1.6 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 autocorrélées, 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 des trois premiers moments de la variable dépendante (espérance, écart-type et coefficient d'asymétrie), celui des taux marginaux de substitution (et de leurs élasticités) entre les moments, et entre les variables indépendantes, ainsi que celui des 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, Coefficient d'asymétrie, Taux marginaux de substitution, Prévision par Maximisation de la vraisemblance.

## ZUSAMMENFASSUNG

L-1.6 ist ein Programm, das es erlaubt, die Spezifikation der funktionalen Form im verallgemeinerten Eingleichungsregressionsmodell gleichzeitig mit der funktionalen Form der Heteroskedastizität der Residuen, die auch autokorreliert sein können, vollständig zu untersuchen. Fixe oder geschätzte Parameter der Box-Cox Transformation können der Abhängigen und Gruppen unabhängiger Variablen zugeordnet werden. Diese lassen sich zugleich innerhalb eine vorbestimmten Autokorrelationsstruktur verändern, wobei die autoregressiven Parameter und die Parameter einer ausgewählten funktionale Form der heteroskedastizität fixiert oder geschätzt werden. Die asymptotische Kovarianzmatrix aller Parameterschätzer des Systems wird berechnet. Es ist möglich, eine Anzahl zweckmäßiger Ergebnisse, einschließlich Elastizitäten der ersten drei Momente der abhängigen Variablen (Erwartungswert, Standardfehler und zwischen den unabhängigen Variablen auszuwerten. Ebenso können Prognosen auf Grundlage von Simulation und Maximum-Likelihood-Techniken und der zugehörigen Statistiken durchgeführt werden.

**Schlüsselworte:** Box-Cox-Regression, verallgemeinerte Heteroskedastizität, Autokorrelation multipler Ordnung, Maximum-Likelihood-Schätzung, Zeitreihen, Querschnittsreihen, Elastizität, Schiefe, Grenzrate der Substitution, Maximum-Likelihood-Prognose.

# Contents

|     |                                                                                                    |    |
|-----|----------------------------------------------------------------------------------------------------|----|
| 1   | INTRODUCTION AND STATISTICAL MODEL . . . . .                                                       | 1  |
| 1.1 | Introduction . . . . .                                                                             | 1  |
| 1.2 | Log-likelihood function . . . . .                                                                  | 2  |
| 1.3 | Computational aspects . . . . .                                                                    | 7  |
| 1.4 | Model types . . . . .                                                                              | 12 |
| 1.5 | Model estimation . . . . .                                                                         | 13 |
| 2   | ESTIMATION RESULTS . . . . .                                                                       | 14 |
| 2.1 | Definitions of moments of the dependent variable . . . . .                                         | 14 |
| 2.2 | Derivatives and elasticities of the sample and expected values of the dependent variable . . . . . | 19 |
| 2.3 | Derivatives and elasticities of the standard error of the dependent variable . . . . .             | 24 |
| 2.4 | Derivatives and elasticities of the skewness of the dependent variable . . . . .                   | 29 |
| 2.5 | Ratios of derivatives of the moments of the dependent variable . . . . .                           | 32 |
| 2.6 | Evaluation of moments, their derivatives, rates of substitution and elasticities . . . . .         | 37 |
| 2.7 | Student's t-statistics . . . . .                                                                   | 41 |
| 2.8 | Goodness-of-fit measures . . . . .                                                                 | 42 |
| 3   | SPECIAL OPTIONS . . . . .                                                                          | 44 |
| 3.1 | Correlation matrix and table of variance-decomposition proportions . . . . .                       | 44 |
| 3.2 | Analysis of heteroskedasticity of the residuals . . . . .                                          | 46 |
| 3.3 | Analysis of autocorrelation of the residuals . . . . .                                             | 46 |
| 3.4 | Forecasting: Maximum likelihood and simulation forecasts . . . . .                                 | 49 |
| 4   | REFERENCES . . . . .                                                                               | 56 |

# List of Tables

|         |                                                                                                                                                            |    |
|---------|------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| TABLE 1 | Relationships between the original $(\beta, \delta, L)$ and scaled $(\tilde{\beta}, \tilde{\delta}, \tilde{L})$ for Box-Cox transformed variables. . . . . | 11 |
| TABLE 2 | Explicit forms of the sample value of $Y_t$ and its derivatives and elasticities for the linear and logarithmic cases of $Y_t$ . . . . .                   | 22 |
| TABLE 3 | Explicit forms of the expected value of $Y_t$ and its derivative and elasticity for the linear and logarithmic cases of $Y_t$ . . . . .                    | 23 |
| TABLE 4 | Explicit forms of the standard error of $Y_t$ and its derivatives and elasticities for the linear and logarithmic cases of $Y_t$ . . . . .                 | 28 |
| TABLE 5 | Explicit forms of the skewness of $Y_t$ and its derivatives and elasticities for the linear and logarithmic cases of $Y_t$ . . . . .                       | 31 |

|                                                                                                                                                                                                                                  |    |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| TABLE 6 Feasible combinations of signs of the MRS's among moments. . . . .                                                                                                                                                       | 36 |
| TABLE 7 Feasible combinations of signs of the elasticities of substitution among<br>moments. . . . .                                                                                                                             | 37 |
| TABLE 8 Correction of the elasticities $\eta_{X_{jt}}^s$ , $\eta_{X_{jt}}^e$ , $\eta_{X_{jt}}^\sigma$ and $\eta_{X_{jt}}^\gamma$ at the sample means,<br>for the positive observations of a quasi-dummy or a real dummy. . . . . | 40 |
| TABLE 9 Evaluations provided for the derivatives and elasticities of the moments, and<br>for the ratios of the derivatives of the moments and their elasticities. . . . .                                                        | 41 |
| TABLE 10 Moments of $Y_t$ observed and estimated, and of $E(Y_t)$ estimated. . . . .                                                                                                                                             | 44 |
| TABLE 11 Eigenvalues of $X'X$ , Condition Indexes of $X$ and Proportions of $\text{Var}(b_k)$ . . . . .                                                                                                                          | 45 |
| TABLE 12 Components of the second derivatives of<br>$L_{n+i}$ with respect to $\Pi_{(i)}$ and $Y_{n+i}$ . . . . .                                                                                                                | 53 |

# 1. INTRODUCTION AND STATISTICAL MODEL

## 1.1 Introduction<sup>1</sup>

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

<sup>1</sup> The Natural Sciences and Engineering Council of Canada (NSERC) and the *Fonds Québécois de Recherche sur la Nature et les Technologies* (FQRNT) have supported this profoundly modified version of the L-1.5 algorithm usually referenced as: Liem, T.C., Gaudry, M., Dagenais, M. and U. Blum, “LEVEL: The L-1.5 program for BC-GAUHESEQ (Box-Cox Generalized Autoregressive Heteroskedastic Single Equation) regression and mutimoment analysis”, in Gaudry, M. and S. Lassarre, eds, Structural Road Accident Models: The International DRAG Family, Elsevier Science Publishers, Oxford, Ch. 12, 263–324, 2000. This document is complemented by a user guide for the program: Tran, C.L., and Gaudry, M., “L-1.6 Program user guide with databases, examples and Tablex outputs”, Publication AJD-106, Agora Jules Dupuit, Université de Montréal, April 2008.

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).

The program gives a number of useful outputs such as the elasticities of the first three moments of the dependent variable (expected value, standard error and skewness), the marginal rates of substitution and elasticities of substitution among the moments, and among the independent variables, as well as forecasts made by simulation and maximum likelihood techniques. An application of the analysis of marginal rates of substitution and elasticities of substitution among moments of the dependent variable is found in Blum and Gaudry (1999).

## 1.2 Log-likelihood function

### A. 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:

$$\begin{aligned} (1) \quad Y_t^{(\lambda_y)} &= \sum_{k=1}^K \beta_k X_{kt}^{(\lambda_{xk})} + u_t, \quad (t = 1, \dots, n) \\ (2) \quad u_t &= v_t f(Z_t)^{1/2}, \\ (3) \quad v_t &= \sum_{\ell=1}^r \rho_\ell v_{t-\ell} + w_t, \end{aligned}$$

where

- (i) 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  $\lambda$  on any positive real variable  $V_t$ :

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

with the corresponding "Box-Cox-Gaudry" BCG or inverse function IBCT:

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

where the expression in square brackets must be positive. Note that in (4), if  $\lambda = 0$ , then  $V_t^{(\lambda)} = (V_t^\lambda - 1)/\lambda|_{\lambda=0} = 0/0$ , which is an indeterminate form. Using L'Hospital's rule gives  $\lim_{\lambda \rightarrow 0} V_t^{(\lambda)} = \lim_{\lambda \rightarrow 0} \frac{\partial(V_t^\lambda - 1)/\partial \lambda}{\partial \lambda / \partial \lambda} = \lim_{\lambda \rightarrow 0} V_t^\lambda \ln V_t = \ln V_t$ ;

- (ii) the first-stage vector of residuals  $u = \{u_t\}$  is assumed to be heteroskedastic with mean  $E(u) = 0$  and covariance matrix  $E(uu') = \Omega = \text{diag}(\omega_{11}, \dots, \omega_{nn})$  where  $\omega_{tt} = E(u_t^2) = E(v_t^2)f(Z_t)$  and  $f(Z_t)$  is a function of a vector of  $M$  variables,  $Z_t = (Z_{1t}, \dots, Z_{Mt})$ , 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;
- (iii) 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  $\sigma_w^2 = E(w_t^2)$  and  $\Psi$  is a symmetric matrix of order  $n$ , whose element can be expressed in terms of the  $\rho_\ell$ 's;
- (iv) 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  $\sigma_w^2$  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 regression 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  $\lambda_u$  applied to a linear combination of the variables  $Z_{mt}$ 's which are themselves subject to BCT:

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

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

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

where the expression in curly brackets must be positive since it is associated with a variance. The constants  $\delta_0$  and  $\psi$  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):

- (i) Setting  $\lambda_u$  equal to zero yields a first form of heteroskedasticity:

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

where  $\tilde{\psi}^2 = \psi^2 \exp(\delta_0)$  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  $\tilde{\omega}_{tt}$  is a linear combination of the  $Z_{mt}^{(\lambda_{zm})}$ 's:

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

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

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

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

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

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

where  $\tilde{\psi}^2 = \tilde{\psi}^2 \prod_m \exp (-\delta_m)$  and  $\tilde{\delta}_0 = \tilde{\delta}_0 - \sum_m \delta_m$ . Another special case of (9) - (10) can be obtained if every  $\lambda_{zm}$  is set equal to zero:

$$(13) \quad \tilde{\omega}_{tt} = \tilde{\psi}^2 \prod_m \exp (\delta_m \ln Z_{mt}) = \tilde{\psi}^2 \prod_m Z_{mt}^{\delta_m}$$

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

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  $\delta_1$  such as 2, 1, 1/2, -1/2 and -1 to give what he called "pure" heteroskedasticity.

- (ii) 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  $\lambda_u$  and every  $\lambda_{zm}$  are set equal to one, i.e. where the variance  $\omega_{tt}$  is assumed to be a linear combination of the  $Z_{mt}$ 's:

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

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  $\Omega$  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  $\delta$ -coefficients and  $\lambda$ -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.

## B. 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):

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

where  $f(Z_t) \equiv \tilde{f}(Z_t) = \exp \left[ \sum_m \delta_m Z_{mt}^{(\lambda_{zm})} \right]$  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:

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

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:

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

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  $N(0, \sigma_w^2)$  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:

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

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:

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

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:  $\Pi = (\beta', \sigma_w^2, \lambda_y, \lambda_x', \lambda_z', \delta', \rho')$  where  $\sigma_w^2$  and  $\lambda_y$  are scalars, and  $\beta, \lambda_x, \lambda_z, \delta$  and  $\rho$  are the column vectors associated with the  $\beta_k$ 's,  $\lambda_{xk}$ 's,  $\lambda_{zm}$ 's,  $\delta_m$ 's and  $\rho_\ell$ 's respectively.

### C. 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  $\beta$ -coefficients, we can concentrate the log-likelihood function on the  $\beta_k$ 's and  $\sigma_w^2$  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:

$$(21) \quad Y^{**} = X^{**}\beta + w$$

where in view of (19) and (20),  $Y^{**}$  is a column vector containing the last  $N$  observations,  $X^{**}$  is an  $(N \times K)$  matrix,  $\beta$  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  $\beta$  and  $\sigma_w^2$  are given by:

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

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

By equating these two derivatives to zero and solving for  $\beta$  and  $\sigma_w^2$ , we obtain:

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

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

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

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

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

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

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

### 1.3 Computational aspects

#### A. 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:

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

where

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

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

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

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

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

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

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

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:

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

## B. 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  $\hat{\Pi}$  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  $t$ :

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

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  $\hat{\Pi}$  is given by:

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

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

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

with

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

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

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

Note that the derivatives  $\partial w_t / \partial \bar{\Pi}_i$  in (42) 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  $\lambda_{zm}$ 's and  $\delta_m$ '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  $\delta_m$ . 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).

### C. 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:

$$(43) \quad \tilde{V}_t = s_v V_t$$

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

In general, the  $\beta$  and  $\delta$ -coefficients and their unconditional t-values computed from (38) are not invariant with respect to the scaling of  $Y_t$ ,  $X_{kt}$  and  $Z_{mt}$ , due to the fact that the  $\beta$ -coefficients depend on the measurement units of  $Y_t$  and  $X_{kt}$ 's, and the  $\delta$ -coefficients depend on those of  $Z_{mt}$ 's. In contrast, the  $\lambda_y$ ,  $\lambda_x$ ,  $\lambda_z$  and  $\rho$ -parameters as well as their unconditional t-values remain invariant, since these parameters are pure numbers, i.e. do not depend on the measurement units of the variables. Note that the concentrated log-likelihood function  $\bar{L}$ , hence  $L$ , and the error variance  $\sigma_w^2$  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  $\beta$ -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.

Table 1 summarizes the relationships between the original  $\beta$  and  $\delta$  and the scaled  $\tilde{\beta}$  and  $\tilde{\delta}$ , as well as the original and scaled log-likelihood functions  $L$  and  $\bar{L}$  due to the scaling of the variables  $Y_t$ ,  $X_{kt}$ 's and  $Z_{mt}$ 's, when each variable is transformed by Box-Cox. For the dependent variable, due to the Box-Cox transformation, all the  $\beta$ -coefficients including the regression constant, as well as the log-likelihood function, will change. For the independent variables, only a strictly positive variable  $X_{kt}$  or a quasi-dummy  $Q_t$ , which is defined as a nonnegative variable in Section 2.6, can be transformed by Box-Cox. An associated real dummy  $D_{Qt}$  is always created for each quasi-dummy  $Q_t$ , since the null observations of  $Q_t$  cannot be transformed by Box-Cox. The main difference between  $X_{kt}$  and  $Q_t$ , when they are scaled, is that due to the Box-Cox transformation, there will be a shift in the regression constant  $\beta_o$  for  $X_{kt}$  and in the  $\beta$ -coefficient of the associated real dummy,  $\beta_{D_Q}$ , for  $Q_t$ , in addition to the usual changes in the  $\beta$ -coefficients of  $X_{kt}$  and  $Q_t$  themselves. The log-likelihood function will not be affected by the scaling of  $X_{kt}$  or  $Q_t$ . For the heteroskedasticity variables, only the  $\delta$ -coefficient associated with each  $Z_{mt}$  will change. There will be no effect on the regression constant and the log-likelihood function.

TABLE 1 Relationships between the original  $(\beta, \delta, L)$  and scaled  $(\tilde{\beta}, \tilde{\delta}, \tilde{L})$  for Box-Cox transformed variables.

| Variable                                                               | Regression constant                                                                   | $\beta$ or $\delta$ -coefficient                                                                                                             | Log-likelihood                 |
|------------------------------------------------------------------------|---------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------|
| Dependent<br>$\tilde{Y}_t = s_y Y_t$                                   | $\tilde{\beta}_0 = s_y^{\lambda_y} \beta_0 + s_y^{(\lambda_y)}$                       | $\tilde{\beta}_k = s_y^{\lambda_y} \beta_k$ for all k's                                                                                      | $\tilde{L} = L - N \ln s_y$    |
| Independent<br>• Strictly positive<br>$\tilde{X}_{kt} = s_{xk} X_{kt}$ | $\tilde{\beta}_0 = \beta_0 - s_{xk}^{(\lambda_{xk})} \beta_k / s_{xk}^{\lambda_{xk}}$ | $\tilde{\beta}_k = \beta_k / s_{xk}^{\lambda_{xk}}$                                                                                          | $\tilde{L} = L$<br>(invariant) |
| • Quasi-dummy<br>$\tilde{Q}_t = s_Q Q_t$                               | $\tilde{\beta}_0 = \beta_0$ (invariant)                                               | $\tilde{\beta}_{\tilde{Q}} = \beta_Q / s_Q^{\lambda_Q}$<br>$\tilde{\beta}_{D_Q} = \beta_{D_Q} - s_Q^{(\lambda_Q)} \beta_Q / s_Q^{\lambda_Q}$ | $\tilde{L} = L$<br>(invariant) |
| Heteroskedasticity<br>$\tilde{Z}_{mt} = s_{zm} Z_{mt}$                 | $\tilde{\beta}_0 = \beta_0$ (invariant)                                               | $\tilde{\delta}_m = \delta_m / s_{zm}^{\lambda_{zm}}$                                                                                        | $\tilde{L} = L$<br>(invariant) |

#### D. Constraints on the $\rho$ -parameters

Strictly speaking, the autocorrelation parameters  $\rho_\ell$ '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  $\rho_\ell$  is implicitly introduced by using the Fisher's  $z$ -transformation:

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

with the corresponding inverse transformation  $\rho_\ell = \tanh z_\ell$ . This parameter change from  $\rho_\ell$  to  $z_\ell$  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  $\rho$ -values which are only given at the convergence of  $\bar{L}$  when all the estimated parameters of the model  $\hat{\Pi}$  are printed with their standard-errors and t-statistics. The derivative of  $\bar{L}$  with respect to each  $z$ -parameter is computed as follows:

$$(45) \quad \frac{\partial \bar{L}}{\partial z_\ell} = \frac{\partial \bar{L}}{\partial \rho_\ell} \frac{\partial \rho_\ell}{\partial z_\ell}$$

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

## 1.4 Model types

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

**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  $\lambda_x$  where the variances of  $u_t$ ,  $v_t$  and  $w_t$  are all identical;

**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  $\lambda_z$  and  $\delta$ -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  $u_t$ ;

**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  $\rho$ -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  $u_t$ ;

**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  $\lambda_y$  and  $\lambda_x$ -parameters being estimated or fixed across the four model types. For example, if  $\lambda_y$  or some of the  $\lambda_x$ -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  $\lambda_z$  and  $\delta$ -parameters being estimated or fixed in both model types. For instance, if every  $\lambda_{zm}$  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  $\rho$ -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.

## 1.5 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  $\lambda_x$ -parameter for each category of variables will reduce the number of estimated  $\lambda_x$ -parameters. The program includes two types of constraints on the  $\lambda_y$  and  $\lambda_x$ -parameters:

- (i)  $\lambda_y$  and any  $\lambda_x$ -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  $\lambda_y$  equal to zero and every  $\lambda_x$ -parameter to one yields the semilog form;
- (ii)  $\lambda_y$  can be set equal to one of the estimated  $\lambda_x$ -parameter. This type of constraint is useful in particular for the case where the model is specified with only one estimated  $\lambda$ -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:

- (iii) 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);
- (iv) 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);
- (v) 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 estimation according to the results of the previous analyses:

- (i) 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.
- (ii) 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  $\lambda_z$ -parameters and the  $\delta$ -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.

- (iii) 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  $\rho$ -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 Definitions of moments of the dependent variable

In a standard linear regression model where the dependent variable  $Y_t$  is not subject to a Box-Cox transformation, the calculated value of  $Y_t$  is equal to the expected value of  $Y_t$ . In contrast, in a Box-Cox regression model where  $Y_t$  is transformed by Box-Cox, this property no longer holds. In this case, the expected value of  $Y_t$  is more relevant than the calculated value of  $Y_t$ . Moreover, we know that the use of the Box-Cox transformation on the dependent variable will affect the standard error and the skewness of the distribution of  $Y_t$ . In this section, we will define these four elements, namely the calculated and expected values of  $Y_t$ , the standard error and the skewness of  $Y_t$ .

#### A. Calculated value of the dependent variable $Y_t$

Using the equation (18), the calculated value of the transformed dependent variable  $Y_t^{**}$  is obtained by replacing all the parameters of the model by their maximum likelihood estimates:

$$(46) \quad \hat{Y}_t^{**} = \sum_k \hat{\beta}_k X_{kt}^{**}$$

where

$$\hat{Y}_t^{**} = \hat{Y}_t^* - \sum_{\ell} \hat{\rho}_{\ell} \hat{Y}_{t-\ell}^*$$

$$\hat{Y}_t^* = \hat{Y}_t^{(\hat{\lambda}_y)} / \widehat{f(\hat{Z}_t)}^{1/2}$$

$$\hat{Y}_{t-\ell}^* = \hat{Y}_{t-\ell}^{(\hat{\lambda}_y)} / \widehat{f(\hat{Z}_{t-\ell})}^{1/2}$$

$$\begin{aligned}
\hat{X}_{kt}^{**} &= \hat{X}_{kt}^* - \sum_{\ell} \hat{\rho}_{\ell} \hat{X}_{k,t-\ell}^* \\
\hat{X}_{kt}^* &= X_{kt}^{(\hat{\lambda}_{xk})} / \widehat{f(Z_t)}^{1/2} \\
\hat{X}_{k,t-\ell}^* &= X_{k,t-\ell}^{(\hat{\lambda}_{xk})} / \widehat{f(Z_{t-\ell})}^{1/2} \\
(47) \quad \widehat{f(Z_t)} &= \exp \left[ \sum_m \hat{\delta}_m Z_{mt}^{(\hat{\lambda}_{zm})} \right] \\
\widehat{f(Z_{t-\ell})} &= \exp \left[ \sum_m \hat{\delta}_m Z_{m,t-\ell}^{(\hat{\lambda}_{zm})} \right].
\end{aligned}$$

Replacing  $\hat{Y}_t^{**}$  in the left-hand side of (46) by its expression given in (47), we solve the resulting equation for the calculated value of the original  $Y_t$ :

$$(48) \quad \hat{Y}_t = \begin{cases} \left[ 1 + \hat{\lambda}_y \widehat{f(Z_t)}^{1/2} \left( \sum_{\ell} \hat{\rho}_{\ell} \hat{Y}_{t-\ell}^* + \sum_k \hat{\beta}_k \hat{X}_{kt}^{**} \right) \right]^{1/\hat{\lambda}_y} & \text{if } \hat{\lambda}_y \neq 0 \\ \exp \left[ \widehat{f(Z_t)}^{1/2} \left( \sum_{\ell} \hat{\rho}_{\ell} \frac{\ln \hat{Y}_{t-\ell}}{\widehat{f(Z_{t-\ell})}^{1/2}} + \sum_k \hat{\beta}_k \hat{X}_{kt}^{**} \right) \right] & \text{if } \hat{\lambda}_y = 0 \end{cases}.$$

## B. Expected value of the dependent variable $Y_t$

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  $\varepsilon$  and  $\nu$  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  $Y_t$ :

$$(49) \quad E(Y_t) = \varepsilon \int_{-\infty}^{w_t(\varepsilon)} \varphi(w) dw + \int_{w_t(\varepsilon)}^{w_t(\nu)} Y_t(w_t) \varphi(w) dw + \nu \int_{w_t(\nu)}^{\infty} \varphi(w) dw$$

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

$$(50) \quad \varphi(w) = \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp \left( -\frac{w^2}{2\sigma_w^2} \right),$$

$Y_t(w)$  is a function of  $w$ :

$$(51) \quad Y_t(w) = \begin{cases} \left[ 1 + \lambda_y \widehat{f(Z_t)}^{1/2} \left( \sum_{\ell} \rho_{\ell} Y_{t-\ell}^* + \sum_k \beta_k X_{kt}^{**} + w \right) \right]^{1/\lambda_y} & \text{if } \lambda_y \neq 0, \\ \exp \left[ \widehat{f(Z_t)}^{1/2} \left( \sum_{\ell} \rho_{\ell} \frac{\ln Y_{t-\ell}}{\widehat{f(Z_{t-\ell})}^{1/2}} + \sum_k \beta_k X_{kt}^{**} + w \right) \right] & \text{if } \lambda_y = 0, \end{cases}$$

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

$$(52) \quad w_t(\xi) = \begin{cases} \frac{\xi^{\lambda_y} - 1}{\lambda_y f(Z_t)^{1/2}} - \sum_{\ell} \rho_{\ell} Y_{t-\ell}^* - \sum_k \beta_k X_{kt}^{**} & \text{if } \lambda_y \neq 0, \\ \frac{\ln \xi}{f(Z_t)^{1/2}} - \sum_{\ell} \rho_{\ell} \frac{\ln Y_{t-\ell}}{f(Z_{t-\ell})^{1/2}} - \sum_k \beta_k X_{kt}^{**} & \text{if } \lambda_y = 0. \end{cases}$$

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

1.  **$\lambda_y > 0$ .** If  $\varepsilon$  and  $\nu$  tend towards 0 and  $\infty$  respectively, following (52) the limits of  $w_t(\varepsilon)$  and  $w_t(\nu)$  can be obtained as:

$$(53) \quad w_t^* = \lim_{\varepsilon \rightarrow 0} w_t(\varepsilon) = -\frac{1}{\lambda_y f(Z_t)^{1/2}} - \sum_{\ell} \rho_{\ell} Y_{t-\ell}^* - \sum_k \beta_k X_{kt}^{**}$$

and  $\lim_{\nu \rightarrow \infty} w_t(\nu) = \infty$ . This expression is used in the program to compute  $w_t^*$  automatically and cannot be changed by the user, in contrast to the upper limit in case 3 below, in order to avoid negative values for some observations in the square brackets found on the first line of (51). Whence the first and last terms in (49) disappear and the expected value of  $Y_t$  reduces to the second term which can be written as:

$$(54) \quad E(Y_t) = \int_{w_t^*}^{\infty} Y_t(w) \varphi(w) dw.$$

2.  **$\lambda_y = 0$ .** If  $\varepsilon$  and  $\nu$  tend towards 0 and  $\infty$  respectively, then  $\lim_{\varepsilon \rightarrow 0} w_t(\varepsilon) = -\infty$  and  $\lim_{\nu \rightarrow 0} w_t(\nu) = \infty$ . Hence the first and last terms in (49) disappear and the expected value of  $Y_t$  reduces to the second term which can be written as:

$$(55) \quad E(Y_t) = \int_{-\infty}^{\infty} Y_t(w) \varphi(w) dw.$$

3.  **$\lambda_y < 0$ .** If  $\varepsilon$  tends towards 0, then  $\lim_{\varepsilon \rightarrow 0} w_t(\varepsilon) = -\infty$ , and the first term disappears. Before taking the limit of  $w_t(\nu)$  as  $\nu$  tends towards  $\infty$  in the last two terms of (49), the expected value of  $Y_t$  can be expressed as:

$$(56) \quad E(Y_t) = \int_{-\infty}^{w_t(\nu)} Y_t(w) \varphi(w) dw + \nu \int_{w_t(\nu)}^{\infty} \varphi(w) dw.$$

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

$$(57) \quad \lim_{\nu \rightarrow \infty} w_t(\nu) = -\frac{1}{\lambda_y f(Z_t)^{1/2}} - \sum_{\ell} \rho_{\ell} Y_{t-\ell}^* - \sum_k \beta_k X_{kt}^{**}.$$

Hence  $\lim_{\nu \rightarrow \infty} \nu \int_{w_t(\nu)}^{\infty} \varphi(w) dw$  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  $\nu$ , say  $\tilde{\nu}$ , such that the probability that  $Y_t$  exceeds this upper limit  $\tilde{\nu}$  will be negligible, i.e. given the values of  $Y_t^*$ ,  $X_{kt}^{**}$  and  $Z_t$  in  $w_t(\tilde{\nu})$ , the value of  $E(Y_t)$  will be much smaller than  $\tilde{\nu}$ . The value of  $\nu$  can be chosen very approximately without affecting the numerical precision of  $E(Y_t)$ . By default, the program will generate a value of  $\nu$  determined as follows:

$$(58) \quad \tilde{\nu} = \min \left[ \max_t Y_t + 10\sigma_y, (-0.8\sigma_w \lambda_y)^{1/\lambda_y} \right]$$

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

For the three cases of  $\lambda_y$ , 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  $\varepsilon$  if  $\lambda_y > 0$  or at the upper limit  $\nu$  if  $\lambda_y < 0$  is also computed:

$$(59) \quad \overline{Pr}(Y_t \leq \varepsilon) = \frac{1}{N} \sum_t \int_{-\infty}^{w_t^*} \varphi(w) dw$$

and

$$(60) \quad \overline{Pr}(Y_t \geq \nu) = \frac{1}{N} \sum_t \int_{w_t(\nu)}^{\infty} \varphi(w) dw$$

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  $\overline{Pr}(\cdot)$  lower than 1% indicates that the sample contains practically no limit observations.

### C. Standard error of the dependent variable $Y_t$

The standard error of  $Y_t$  is defined as the square root of the variance of  $Y_t$  which is also known as the second moment of  $Y_t$  centered about the mean  $E(Y_t)$ :

$$(61) \quad \sigma(Y_t) = \sqrt{\text{Var}(Y_t)} = \sqrt{E[Y_t - E(Y_t)]^2} = \sqrt{E(Y_t^2) - [E(Y_t)]^2}$$

where the first and second noncentral moments  $E(Y_t^r)$ ,  $r = 1, 2$ , are given by:

$$(62) \quad E(Y_t^r) = \begin{cases} \int_{w_t^*}^{\infty} Y_t^r(w) \varphi(w) dw & \text{if } \lambda_y > 0 \\ \int_{-\infty}^{\infty} Y_t^r(w) \varphi(w) dw & \text{if } \lambda_y = 0 \\ \int_{-\infty}^{w_t(\nu)} Y_t^r(w) \varphi(w) dw + \nu^r \int_{w_t(\nu)}^{\infty} \varphi(w) dw & \text{if } \lambda_y < 0 \end{cases} .$$

### D. Skewness of the dependent variable $Y_t$

Skewness is a measure of asymmetry of a distribution. It indicates how the data are distributed in a particular distribution relatively to a perfectly symmetric one. Several types of skewness can be defined, but the most usual one is the Fisher Skewness defined as:

$$(63) \quad \gamma = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\mu_3}{\sigma^3}$$

where  $\mu_2$  and  $\mu_3$  are respectively the second and third moments centered about the mean of the distribution, and  $\mu_2^{1/2} \equiv \sigma$  is the standard error of the distribution. Note that the third moment  $\mu_3$  is expressed in cubic units. In order to compare the results from one distribution to another, the skewness is expressed in standard units, i.e. as the third moment divided by  $\sigma^3$ . It does not depend on the units of measurement of the variable considered in the distribution. For example, the skewness of a normal distribution is zero because it is symmetric about its mean, whereas a lognormal distribution has positive skewness, i.e. the right tail is longer than the left one. Conversely a negative skewness indicates that the left tail is longer than the right one. Usually a distribution is considered to be:

Slightly asymmetric, if  $|\gamma| \leq 0.5$  .

Moderately asymmetric, if  $0.5 < |\gamma| \leq 1$  .

Highly asymmetric, if  $|\gamma| > 1$  .

Using the definition of the skewness in (63) applied to the distribution of the dependent variable  $Y_t$ ,  $\gamma(Y_t)$  can be written as:

$$(64) \quad \gamma(Y_t) = \frac{\mu_3(Y_t)}{\sigma^3(Y_t)}$$

where  $\mu_3(Y_t) = E[Y_t - E(Y_t)]^3$  is the third moment of  $Y_t$  centered about the mean  $E(Y_t)$  and  $\sigma(Y_t)$  is the standard error of  $Y_t$  defined in (61).

The third central moment  $\mu_3(Y_t)$  can be expressed as a function of the first three noncentral moments of  $Y_t$ :

$$(65) \quad \mu_3(Y_t) = E(Y_t^3) - E(Y_t) \left\{ 3E(Y_t^2) - 2[E(Y_t)]^2 \right\}$$

where the first three noncentral moments  $E(Y_t^r)$ ,  $r = 1, 3$ , have the following forms:

$$(66) \quad E(Y_t^r) = \begin{cases} \int_{w_t^*}^{\infty} Y_t^r(w) \varphi(w) dw & \text{if } \lambda_y > 0 \\ \int_{-\infty}^{\infty} Y_t^r(w) \varphi(w) dw & \text{if } \lambda_y = 0 \\ \int_{-\infty}^{w_t(\nu)} Y_t^r(w) \varphi(w) dw + \nu^r \int_{w_t(\nu)}^{\infty} \varphi(w) dw & \text{if } \lambda_y < 0 \end{cases}.$$

## 2.2 Derivatives and elasticities of the sample and expected values of the dependent variable

Two types of elasticity are computed in the program:

The first type of elasticity, denoted as  $\eta_{X_{jt}}^s$ , 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.

The second type of elasticity, denoted as  $\eta_{X_{jt}}^e$ , 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  $\eta_{Z_{mt}}^s$  and  $\eta_{Z_{mt}}^e$ , 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:

The variable  $Z_{mt}$  specified in  $f(Z_t)$  is also used as an independent variable  $X_{jt}$ .

It is only specified in  $f(Z_t)$ , not elsewhere.

### A. Derivatives and Elasticities of the Sample Value of $Y_t$

The sample value of  $Y_t$  has the same form as  $Y_t(w)$  in (51), but with  $w$  replaced by  $w_t$ . The derivative and the elasticity of the sample value of  $Y_t$  with respect to an independent variable  $X_{jt}$  can be written as:

$$D_{X_{jt}}^s = \frac{\partial Y_t}{\partial X_{jt}} = \frac{1}{Y_t^{\lambda_y - 1}} \left( \beta_j X_{jt}^{\lambda_{xj} - 1} + X_{jt}^{-1} G_{X_{jt}} \right) \quad (67)$$

$$\eta_{X_{jt}}^s = \frac{\partial Y_t}{\partial X_{jt}} \frac{X_{jt}}{Y_t} = \frac{1}{Y_t^{\lambda_y}} \left( \beta_j X_{jt}^{\lambda_{xj}} + G_{X_{jt}} \right)$$

where

$$G_{X_{jt}} = \begin{cases} 0 & \text{if } f(Z_t) \text{ is not a function of } X_{jt}, \\ \frac{1}{2} \delta_m Z_{mt}^{\lambda_{zm}} \left[ f(Z_t)^{1/2} A_t - \sum_k \beta_k X_{kt}^{(\lambda_{xk})} \right] & \text{if } f(Z_t) \text{ contains one } Z_{mt} = X_{jt}. \end{cases} \quad (68)$$

and  $A_t = \sum_\ell \rho_\ell Y_{t-\ell}^* + \sum_k \beta_k X_{kt}^{**}$ . Similarly, the derivative and the elasticity of the sample value of  $Y_t$  with respect to a heteroskedasticity variable  $Z_{mt}$  which appears only in  $f(Z_t)$  are:

$$D_{Z_{mt}}^s = \frac{\partial Y_t}{\partial Z_{mt}} = \frac{Z_{mt}^{-1} G_{Z_{mt}}}{Y_t^{\lambda_y - 1}} \quad (69)$$

$$\eta_{Z_{mt}}^s = \frac{\partial Y_t}{\partial Z_{mt}} \frac{Z_{mt}}{Y_t} = \frac{G_{Z_{mt}}}{Y_t^{\lambda_y}}$$

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

### B. Derivatives and Elasticities of the Expected Value of $Y_t$

Using  $E(Y_t)$  in (54)-(55)-(56) depending on the value of  $\lambda_y$ , the derivative and the elasticity of the expected value of  $Y_t$  with respect to an independent variable  $X_{jt}$  can be written as:

$$D_{X_{jt}}^e = \frac{\partial E(Y_t)}{\partial X_{jt}} = \int_{R_w} [Y_t(w)]^{1-\lambda_y} \left[ \beta_j X_{jt}^{\lambda_{xj} - 1} + X_{jt}^{-1} H_{X_{jt}}(w) \right] \varphi(w) dw \quad (70)$$

$$\eta_{X_{jt}}^e = \frac{\partial E(Y_t)}{\partial X_{jt}} \frac{X_{jt}}{E(Y_t)} = \frac{1}{E(Y_t)} \int_{R_w} [Y_t(w)]^{1-\lambda_y} \left[ \beta_j X_{jt}^{\lambda_{xj}} + H_{X_{jt}}(w) \right] \varphi(w) dw$$

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

$$(71) \quad H_{X_{jt}}(w) = \begin{cases} 0 & \text{if } f(Z_t) \text{ is not a function of } X_{jt}, \\ \frac{1}{2}\delta_m Z_{mt}^{\lambda_{zm}} \left[ f(Z_t)^{1/2}(A_t + w) - \sum_k \beta_k X_{kt}^{(\lambda_{xk})} \right] & \text{if } f(Z_t) \text{ contains } Z_{mt} = X_{jt}. \end{cases}$$

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

$$(72) \quad D_{Z_{mt}}^e = \frac{\partial E(Y_t)}{\partial Z_{mt}} = Z_{mt}^{-1} \int_{R_w} [Y_t(w)]^{1-\lambda_y} H_{Z_{mt}}(w) \varphi(w) dw$$

$$\eta_{Z_{mt}}^e = \frac{\partial E(Y_t)}{\partial Z_{mt}} \frac{Z_{mt}}{E(Y_t)} = \frac{1}{E(Y_t)} \int_{R_w} [Y_t(w)]^{1-\lambda_y} H_{Z_{mt}}(w) \varphi(w) dw$$

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

### C. Derivatives and Elasticities for the Linear and Logarithmic Cases of $Y_t$

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

For these two cases, the explicit forms of the sample value of  $Y_t$  and its derivative and elasticity are given in Table 2, 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  $\rho_l$ 's included in  $A_t$  equal to zero.

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 3.

TABLE 2 Explicit forms of the sample value of  $Y_t$  and its derivatives and elasticities for the linear and logarithmic cases of  $Y_t$ .

| STATISTIC                                                                                                                                                                                                                 | CASE              | HOMOSKEDASTICITY                                                    | HETEROSKEDASTICITY                  |                                                                             |               |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------|---------------------------------------------------------------------|-------------------------------------|-----------------------------------------------------------------------------|---------------|
|                                                                                                                                                                                                                           |                   | (No $f(Z_t)$ involved)                                              | $X_{jt} \neq Z_{mt}$                | $X_{jt} = Z_{mt}$                                                           |               |
| SAMPLE<br>VALUE<br><br>$Y_t$                                                                                                                                                                                              | $\lambda_y = 1$   | $1 + A_t$                                                           | $1 + f(Z_t)^{1/2} A_t$              |                                                                             |               |
|                                                                                                                                                                                                                           | $\lambda_y = 0$   | $\exp(A_t)$                                                         | $\exp\left[f(Z_t)^{1/2} A_t\right]$ |                                                                             |               |
| DERIVATIVE<br><br>$D_{X_{jt}}^s$                                                                                                                                                                                          | $\lambda_y = 1$   | $\beta_j X_{jt}^{\lambda_{xj}-1}$                                   |                                     | $\beta_j X_{jt}^{\lambda_{xj}-1} + Z_{mt}^{-1} G_{Z_{mt}}$                  |               |
|                                                                                                                                                                                                                           | $\lambda_y = 0$   | $\beta_j X_{jt}^{\lambda_{xj}-1} Y_t$                               |                                     | $\left(\beta_j X_{jt}^{\lambda_{xj}-1} + Z_{mt}^{-1} G_{Z_{mt}}\right) Y_t$ |               |
|                                                                                                                                                                                                                           | $D_{Z_{mt}}^s$    | $\lambda_y = 1$                                                     | 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<br><br>$\eta_{X_{jt}}^s$                                                                                                                                                                                       | $\lambda_y = 1$   | $\beta_j X_{jt}^{\lambda_{xj}}/Y_t$                                 |                                     | $\left(\beta_j X_{jt}^{\lambda_{xj}} + G_{Z_{mt}}\right)/Y_t$               |               |
|                                                                                                                                                                                                                           | $\lambda_y = 0$   | $\eta_{X_{jt}}^s = \eta_{X_{jt}}^e = \beta_j X_{jt}^{\lambda_{xj}}$ |                                     | $\beta_j X_{jt}^{\lambda_{xj}} + G_{Z_{mt}}$                                |               |
|                                                                                                                                                                                                                           | $\eta_{Z_{mt}}^s$ | $\lambda_y = 1$                                                     | 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 $G_{Z_{mt}} = \frac{1}{2} \delta_m Z_{mt}^{\lambda_{z,m}} \left[ f(Z_t)^{1/2} A_t - \sum_k \beta_k X_{kt}^{(\lambda_{xk})} \right]$ . |                   |                                                                     |                                     |                                                                             |               |

TABLE 3 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<br>(No $f(Z_t)$ involved)                                                                                                                                                                                                                     |                                                           | HETEROSKEDASTICITY<br>$X_{jt} \neq Z_{mt}$ <span style="float:right;"><math>X_{jt} = Z_{mt}</math></span>                                                                                                                                                                                       |                               |                                                                                                                                                                                                                                                                         |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| EXPECTED<br>VALUE                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | $\lambda_y = 1$                                                     | $E(Y_t) = \int\limits_{w_t^*}^{\infty} Y_t(w) \varphi(w) dw$<br>where $Y_t(w) = 1 + A_t + w$ <span style="float:right;">where <math>Y_t(w) = 1 + f(Z_t)^{1/2}(A_t + w)</math></span>                                                                           |                                                           |                                                                                                                                                                                                                                                                                                 |                               |                                                                                                                                                                                                                                                                         |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | $E(Y_t)$                                                            | <span style="border: 1px solid black; padding: 2px;">LIMIT CASE</span> $\lim\limits_{w_t^* \rightarrow -\infty} E(Y_t) = 1 + A_t$                                                                                                                              |                                                           | <span style="border: 1px solid black; padding: 2px;">LIMIT CASE</span> $\lim\limits_{w_t^* \rightarrow -\infty} E(Y_t) = 1 + f(Z_t)^{1/2} A_t$                                                                                                                                                  |                               |                                                                                                                                                                                                                                                                         |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | $\lambda_y = 0$                                                     | $E(Y_t) = \int\limits_{-\infty}^{\infty} Y_t(w) \varphi(w) dw$<br>where $Y_t(w) = \exp(A_t + w)$ <span style="float:right;">where <math>Y_t(w) = \exp\left[f(Z_t)^{1/2}(A_t + w)\right]</math></span>                                                          |                                                           |                                                                                                                                                                                                                                                                                                 |                               |                                                                                                                                                                                                                                                                         |
| DERIVATIVE                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | $\lambda_y = 1$                                                     | $D_{X_{jt}}^e = \beta_j X_{jt}^{\lambda_{xj}-1} \int\limits_{w_t^*}^{\infty} \varphi(w) dw$<br><span style="border: 1px solid black; padding: 2px;">LIMIT CASE</span> $\lim\limits_{w_t^* \rightarrow -\infty} D_{X_{jt}}^e = \beta_j X_{jt}^{\lambda_{xj}-1}$ |                                                           | Same as left column<br>$+ D_{Z_{mt}}^e(\lambda_y = 1)$<br><span style="border: 1px solid black; padding: 2px;">LIMIT CASE</span> $\lim\limits_{w_t^* \rightarrow -\infty} D_{X_{jt}}^e = \beta_j X_{jt}^{\lambda_{xj}-1} + Z_{mt}^{-1} G_{Z_{mt}}$                                              |                               |                                                                                                                                                                                                                                                                         |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | $\lambda_y = 0$                                                     | $D_{X_{jt}}^e = \beta_j X_{jt}^{\lambda_{xj}-1} E(Y_t)$                                                                                                                                                                                                        |                                                           | Same as left column<br>$+ D_{Z_{mt}}^e(\lambda_y = 0)$                                                                                                                                                                                                                                          |                               |                                                                                                                                                                                                                                                                         |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | $D_{Z_{mt}}^e$                                                      | $\lambda_y = 1$                                                                                                                                                                                                                                                | Not applicable                                            |                                                                                                                                                                                                                                                                                                 | $D_{Z_{mt}}^e(\lambda_y = 1)$ | Same as above                                                                                                                                                                                                                                                           |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                                                                     | $\lambda_y = 0$                                                                                                                                                                                                                                                | Not applicable                                            |                                                                                                                                                                                                                                                                                                 | $D_{Z_{mt}}^e(\lambda_y = 0)$ | Same as above                                                                                                                                                                                                                                                           |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                                                                     | ELASTICITY                                                                                                                                                                                                                                                     | $\lambda_y = 1$                                           | $\eta_{X_{jt}}^e = \frac{\beta_j X_{jt}^{\lambda_{xj}}}{E(Y_t)} \int\limits_{w_t^*}^{\infty} \varphi(w) dw$<br><span style="border: 1px solid black; padding: 2px;">LIMIT CASE</span> $\lim\limits_{w_t^* \rightarrow -\infty} \eta_{X_{jt}}^e = \frac{\beta_j X_{jt}^{\lambda_{xj}}}{1 + A_t}$ |                               | same as left column<br>$+ \eta_{Z_{mt}}^e(\lambda_y = 1)$<br><span style="border: 1px solid black; padding: 2px;">LIMIT CASE</span> $\lim\limits_{w_t^* \rightarrow -\infty} \eta_{X_{jt}}^e = \frac{\beta_j X_{jt}^{\lambda_{xj}} + G_{Z_{mt}}}{1 + f(Z_t)^{1/2} A_t}$ |
| $\lambda_y = 0$                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | $\eta_{X_{jt}}^e = \eta_{X_{jt}}^s = \beta_j X_{jt}^{\lambda_{xj}}$ |                                                                                                                                                                                                                                                                | Same as left column<br>$+ \eta_{Z_{mt}}^e(\lambda_y = 0)$ |                                                                                                                                                                                                                                                                                                 |                               |                                                                                                                                                                                                                                                                         |
| $\eta_{Z_{mt}}^e$                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | $\lambda_y = 1$                                                     | Not applicable                                                                                                                                                                                                                                                 |                                                           | $\eta_{Z_{mt}}^e(\lambda_y = 1)$                                                                                                                                                                                                                                                                | Same as above                 |                                                                                                                                                                                                                                                                         |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | $\lambda_y = 0$                                                     | Not applicable                                                                                                                                                                                                                                                 |                                                           | $\eta_{Z_{mt}}^e(\lambda_y = 0)$                                                                                                                                                                                                                                                                | Same as above                 |                                                                                                                                                                                                                                                                         |
| where $A_t = \sum_l \rho_l Y_{t-l}^* + \sum_k \beta_k X_{kt}^{**}$ , $G_{Z_{mt}} = \frac{1}{2} \delta_m Z_{mt}^{\lambda_{zm}} \left[ f(Z_t)^{1/2} A_t - \sum_k \beta_k X_{kt}^{(\lambda_{xk})} \right]$ ,<br>$D_{Z_{mt}}^e(\lambda_y = 1) = Z_{mt}^{-1} \int\limits_{w_t^*}^{\infty} H_{Z_{mt}}(w) \varphi(w) dw$ , $D_{Z_{mt}}^e(\lambda_y = 0) = Z_{mt}^{-1} \int\limits_{-\infty}^{\infty} Y_t(w) H_{Z_{mt}}(w) \varphi(w) dw$ ,<br>$\eta_{Z_{mt}}^e(\lambda_y = 1) = \frac{1}{E(Y_t)} \int\limits_{w_t^*}^{\infty} H_{Z_{mt}}(w) \varphi(w) dw$ , $\eta_{Z_{mt}}^e(\lambda_y = 0) = \frac{1}{E(Y_t)} \int\limits_{-\infty}^{\infty} Y_t(w) H_{Z_{mt}}(w) \varphi(w) dw$ ,<br>and $H_{Z_{mt}}(w) = \frac{1}{2} \delta_m Z_{mt}^{\lambda_{zm}} \left[ f(Z_t)^{1/2} (A_t + w) - \sum_k \beta_k X_{kt}^{(\lambda_{xk})} \right]$ . |                                                                     |                                                                                                                                                                                                                                                                |                                                           |                                                                                                                                                                                                                                                                                                 |                               |                                                                                                                                                                                                                                                                         |

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

The heteroskedasticity variable  $Z_{mt}$  is also used as an independent variable  $X_{jt}$ .

It is specified only in the function  $f(Z_t)$ , not elsewhere.

#### A. Derivative and elasticity of $\sigma(Y_t)$ with respect to an independent variable

The derivative and the elasticity of the standard error of  $Y_t$  with respect to an independent variable  $X_{jt}$  are given by:

$$(73) \quad \begin{aligned} D_{X_{jt}}^{\sigma} &= \frac{\partial \sigma(Y_t)}{\partial X_{jt}} = \frac{1}{2\sigma(Y_t)} \left[ \frac{\partial E(Y_t^2)}{\partial X_{jt}} - 2E(Y_t) \frac{\partial E(Y_t)}{\partial X_{jt}} \right] \\ \eta_{X_{jt}}^{\sigma} &= \frac{\partial \sigma(Y_t)}{\partial X_{jt}} \frac{X_{jt}}{\sigma(Y_t)} \end{aligned}$$

where the derivatives of the first and second noncentral moments  $E(Y_t^r)$ ,  $r = 1, 2$ , with respect to  $X_{jt}$  are:

$$(74) \quad \frac{\partial E(Y_t^r)}{\partial X_{jt}} = r \int_{R_w} [Y_t(w)]^{r-\lambda_y} \left[ \beta_j X_{jt}^{\lambda_{xj}-1} + X_{jt}^{-1} H_{X_{jt}}(w) \right] \varphi(w) dw \quad .$$

Note that  $Y_t(w)$ ,  $R_w$  and  $H_{X_{jt}}(w)$  are already defined in (51), (70) and (71) 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 and the elasticity of the standard error of  $Y_t$  with respect to  $X_{jt}$  can be expressed as:

$$(75) \quad \begin{aligned} D_{X_{jt}}^{\sigma} &= \frac{1}{\sigma(Y_t)} \int_{R_w} [Y_t(w)]^{1-\lambda_y} [Y_t(w) - E(Y_t)] \left[ \beta_j X_{jt}^{\lambda_{xj}-1} + X_{jt}^{-1} H_{X_{jt}}(w) \right] \varphi(w) dw, \\ \eta_{X_{jt}}^{\sigma} &= \frac{1}{Var(Y_t)} \int_{R_w} [Y_t(w)]^{1-\lambda_y} [Y_t(w) - E(Y_t)] \left[ \beta_j X_{jt}^{\lambda_{xj}} + H_{X_{jt}}(w) \right] \varphi(w) dw \quad . \end{aligned}$$

#### B. Derivative and elasticity 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  $D_{Z_{mt}}^{\sigma}$  and  $\eta_{Z_{mt}}^{\sigma}$  are analogous to  $D_{X_{jt}}^{\sigma}$  and  $\eta_{X_{jt}}^{\sigma}$  without the terms  $\beta_j X_{jt}^{\lambda_{xj}-1}$  and  $\beta_j X_{jt}^{\lambda_{xj}}$  respectively:

$$\begin{aligned}
(76) \quad D_{Z_{mt}}^\sigma &= \frac{Z_{mt}^{-1}}{\sigma(Y_t)} \int_{\mathcal{R}_w} [Y_t(w)]^{1-\lambda_y} [Y_t(w) - E(Y_t)] H_{Z_{mt}}(w) \varphi(w) dw, \\
\eta_{Z_{mt}}^\sigma &= \frac{1}{\text{Var}(Y_t)} \int_{\mathcal{R}_w} [Y_t(w)]^{1-\lambda_y} [Y_t(w) - E(Y_t)] H_{Z_{mt}}(w) \varphi(w) dw.
\end{aligned}$$

where  $H_{Z_{mt}}(w)$  is already defined in (72).

### C. Standard error of $Y_t$ for the linear and logarithmic cases of $Y_t$

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 12.2 and 12.3 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  $\rho_\ell$ 's included in  $A_t$  equal to zero.

#### 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 :

$$\begin{aligned}
(77) \quad \lim_{w_t^* \rightarrow -\infty} \sigma(Y_t) &= \lim_{w_t^* \rightarrow -\infty} \left\{ \int_{w_t^*}^{\infty} Y_t^2(w) \varphi(w) dw - \left[ \int_{w_t^*}^{\infty} Y_t(w) \varphi(w) dw \right]^2 \right\}^{\frac{1}{2}} \\
&= \left\{ (1 + A_t)^2 \int_{-\infty}^{\infty} \varphi(w) dw + \int_{-\infty}^{\infty} w^2 \varphi(w) dw + 2(1 + A_t) \int_{-\infty}^{\infty} w \varphi(w) dw \right. \\
&\quad \left. - \left[ (1 + A_t) \int_{-\infty}^{\infty} \varphi(w) dw + \int_{-\infty}^{\infty} w \varphi(w) dw \right]^2 \right\}^{\frac{1}{2}} \\
&= \left[ \int_{-\infty}^{\infty} w^2 \varphi(w) dw \right]^{\frac{1}{2}} = \sigma_w
\end{aligned}$$

where  $Y_t(w) = 1 + A_t + w$ ,  $\int_{-\infty}^{\infty} \varphi(w) dw = 1$ ,  $\int_{-\infty}^{\infty} w \varphi(w) dw = 0$  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  $f(Z_t)^{\frac{1}{2}}$ :

$$\begin{aligned}
 \lim_{w_t^* \rightarrow -\infty} \sigma(Y_t) &= \lim_{w_t^* \rightarrow -\infty} \left\{ \int_{w_t^*}^{\infty} Y_t^2(w) \varphi(w) dw - \left[ \int_{w_t^*}^{\infty} Y_t(w) \varphi(w) dw \right]^2 \right\}^{\frac{1}{2}} \\
 &= \left\{ \left[ 1 + f(Z_t)^{\frac{1}{2}} A_t \right]^2 \int_{-\infty}^{\infty} \varphi(w) dw + f(Z_t) \int_{-\infty}^{\infty} w^2 \varphi(w) dw \right. \\
 (78) \quad &\quad \left. + 2 \left[ 1 + f(Z_t)^{\frac{1}{2}} A_t \right] f(Z_t)^{\frac{1}{2}} \int_{-\infty}^{\infty} w \varphi(w) dw \right. \\
 &\quad \left. - \left[ \left[ 1 + f(Z_t)^{\frac{1}{2}} A_t \right] \int_{-\infty}^{\infty} \varphi(w) dw + f(Z_t)^{\frac{1}{2}} \int_{-\infty}^{\infty} w \varphi(w) dw \right]^2 \right\}^{\frac{1}{2}} \\
 &= \left[ f(Z_t) \int_{-\infty}^{\infty} w^2 \varphi(w) dw \right]^{\frac{1}{2}} = f(Z_t)^{\frac{1}{2}} \sigma_w.
 \end{aligned}$$

where  $Y_t(w) = 1 + f(Z_t)^{\frac{1}{2}}(A_t + w)$ .

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

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

$$(79) \quad \sigma(Y_t) = \exp(A_t) \left\{ \int_{-\infty}^{\infty} \exp(2w) \varphi(w) dw - \left[ \int_{-\infty}^{\infty} \exp(w) \varphi(w) dw \right]^2 \right\}^{\frac{1}{2}}.$$

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

$$(80) \quad \sigma(Y_t) = \exp \left[ f(Z_t)^{\frac{1}{2}} A_t \right] \left\{ \int_{-\infty}^{\infty} \exp \left[ 2f(Z_t)^{\frac{1}{2}} w \right] \varphi(w) dw - \left[ \int_{-\infty}^{\infty} \exp \left[ f(Z_t)^{\frac{1}{2}} w \right] \varphi(w) dw \right]^2 \right\}^{\frac{1}{2}}.$$

#### D. Derivatives and elasticities of $\sigma(Y_t)$ for the linear and logarithmic cases of $Y_t$

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  $D_{X_{jt}}^\sigma$  and the elasticity  $\eta_{X_{jt}}^\sigma$  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  $\lim_{w_t^* \rightarrow -\infty} \int_{w_t^*}^{\infty} [Y_t(w) - E(Y_t)]\varphi(w)dw$  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  $D_{X_{jt}}^\sigma$  and  $\eta_{X_{jt}}^\sigma$  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  $D_{X_{jt}}^\sigma$  is a function of  $X_{jt}$  and  $\sigma(Y_t)$ , but the elasticity  $\eta_{X_{jt}}^\sigma$  is a function of  $X_{jt}$  only and it is identical to both elasticities  $\eta_{X_{jt}}^e$  and  $\eta_{X_{jt}}^s$ .
- **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  $D_{X_{jt}}^\sigma$  and the elasticity  $\eta_{X_{jt}}^\sigma$  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  $D_{Z_{mt}}^\sigma$  and the elasticity  $\eta_{Z_{mt}}^\sigma$  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  $D_{Z_{mt}}^\sigma$  and the elasticity  $\eta_{Z_{mt}}^\sigma$  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 derivatives and elasticities for the linear and logarithmic cases of  $Y_t$ .

| STATISTIC                                                                                                                                            | CASE                     | HOMOSKEDASTICITY<br>(No $f(Z_t)$ involved)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                | HETEROSKEDASTICITY<br>$X_{jt} \neq Z_{mt}$ <span style="float:right"><math>X_{jt} = Z_{mt}</math></span>                                                                                                                                                    |               |
|------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------|
| STANDARD<br>ERROR<br><br>$\sigma(Y_t)$                                                                                                               | $\lambda_y = 1$          | $\sigma(Y_t) = \left\{ \int_{w_t^*}^{\infty} Y_t^2(w) \varphi(w) dw - \left[ \int_{w_t^*}^{\infty} Y_t(w) \varphi(w) dw \right]^2 \right\}^{1/2}$ <div>where <math>Y_t(w) = 1 + A_t + w</math><span style="float:right">where <math>Y_t(w) = 1 + f(Z_t)^{1/2}(A_t + w)</math></span></div>                                                                                                                                                                                                                                          |                |                                                                                                                                                                                                                                                             |               |
|                                                                                                                                                      | $\lambda_y = 0$          | <div><div><div>LIMIT CASE</div><math>\lim_{w_t^* \rightarrow -\infty} \sigma(Y_t) = \sigma_w</math></div><div><div>LIMIT CASE</div><math>\lim_{w_t^* \rightarrow -\infty} \sigma(Y_t) = f(Z_t)^{1/2} \sigma_w</math></div></div> $\sigma(Y_t) = \left\{ \int_{-\infty}^{\infty} Y_t^2(w) \varphi(w) dw - \left[ \int_{-\infty}^{\infty} Y_t(w) \varphi(w) dw \right]^2 \right\}^{1/2}$ <div>where <math>Y_t(w) = \exp(A_t + w)</math><span style="float:right">where <math>Y_t(w) = \exp[f(Z_t)^{1/2}(A_t + w)]</math></span></div> |                |                                                                                                                                                                                                                                                             |               |
| DERIVATIVE<br><br>$D_{X_{jt}}^{\sigma}$                                                                                                              | $\lambda_y = 1$          | $D_{X_{jt}}^{\sigma} = \frac{\beta_j X_{jt}^{\lambda_{xj}-1}}{\sigma(Y_t)} \int_{w_t^*}^{\infty} [Y_t(w) - E(Y_t)] \varphi(w) dw$ <div><div>LIMIT CASE</div><math>\lim_{w_t^* \rightarrow -\infty} D_{X_{jt}}^{\sigma} = 0</math></div>                                                                                                                                                                                                                                                                                             |                | <div>Same as left column<br/>+ <math>D_{Z_{mt}}^{\sigma}(\lambda_y = 1)</math></div> <div><div>LIMIT CASE</div><math>\lim_{w_t^* \rightarrow -\infty} D_{X_{jt}}^{\sigma} = \frac{1}{2} \delta_m Z_{mt}^{\lambda_{zm}-1} f(Z_t)^{1/2} \sigma_w</math></div> |               |
|                                                                                                                                                      | $\lambda_y = 0$          | $D_{X_{jt}}^{\sigma} = \beta_j X_{jt}^{\lambda_{xj}-1} \sigma(Y_t)$                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                | <div>Same as left column<br/>+ <math>D_{Z_{mt}}^{\sigma}(\lambda_y = 0)</math></div>                                                                                                                                                                        |               |
|                                                                                                                                                      | $D_{Z_{mt}}^{\sigma}$    | $\lambda_y = 1$                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | Not applicable | $D_{Z_{mt}}^{\sigma}(\lambda_y = 1)$                                                                                                                                                                                                                        | Same as above |
|                                                                                                                                                      |                          | $\lambda_y = 0$                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | Not applicable | $D_{Z_{mt}}^{\sigma}(\lambda_y = 0)$                                                                                                                                                                                                                        | Same as above |
| ELASTICITY<br><br>$\eta_{X_{jt}}^{\sigma}$                                                                                                           | $\lambda_y = 1$          | $\eta_{X_{jt}}^{\sigma} = \frac{\beta_j X_{jt}^{\lambda_{xj}}}{\text{Var}(Y_t)} \int_{w_t^*}^{\infty} [Y_t(w) - E(Y_t)] \varphi(w) dw$ <div><div>LIMIT CASE</div><math>\lim_{w_t^* \rightarrow -\infty} \eta_{X_{jt}}^{\sigma} = 0</math></div>                                                                                                                                                                                                                                                                                     |                | <div>same as left column<br/>+ <math>\eta_{Z_{mt}}^{\sigma}(\lambda_y = 1)</math></div> <div><div>LIMIT CASE</div><math>\lim_{w_t^* \rightarrow -\infty} \eta_{X_{jt}}^{\sigma} = \frac{1}{2} \delta_m Z_{mt}^{\lambda_{zm}}</math></div>                   |               |
|                                                                                                                                                      | $\lambda_y = 0$          | $\eta_{X_{jt}}^{\sigma} = \eta_{X_{jt}}^{\epsilon} = \eta_{X_{jt}}^{\dot{\sigma}} = \beta_j X_{jt}^{\lambda_{xj}}$                                                                                                                                                                                                                                                                                                                                                                                                                  |                | <div>Same as left column<br/>+ <math>\eta_{Z_{mt}}^{\sigma}(\lambda_y = 0)</math></div>                                                                                                                                                                     |               |
|                                                                                                                                                      | $\eta_{Z_{mt}}^{\sigma}$ | $\lambda_y = 1$                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | Not applicable | $\eta_{Z_{mt}}^{\sigma}(\lambda_y = 1)$                                                                                                                                                                                                                     | Same as above |
|                                                                                                                                                      |                          | $\lambda_y = 0$                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | Not applicable | $\eta_{Z_{mt}}^{\sigma}(\lambda_y = 0)$                                                                                                                                                                                                                     | Same as above |
| where $A_t = \sum_{\ell} \rho_{\ell} Y_{t-\ell}^* + \sum_k \beta_k X_{kt}^{**}$ ,                                                                    |                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                |                                                                                                                                                                                                                                                             |               |
| $D_{Z_{mt}}^{\sigma}(\lambda_y = 1) = \frac{Z_{mt}^{-1}}{\sigma(Y_t)} \int_{w_t^*}^{\infty} [Y_t(w) - E(Y_t)] H_{Z_{mt}}(w) \varphi(w) dw,$          |                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                |                                                                                                                                                                                                                                                             |               |
| $D_{Z_{mt}}^{\sigma}(\lambda_y = 0) = \frac{Z_{mt}^{-1}}{\sigma(Y_t)} \int_{-\infty}^{\infty} Y_t(w) [Y_t(w) - E(Y_t)] H_{Z_{mt}}(w) \varphi(w) dw,$ |                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                |                                                                                                                                                                                                                                                             |               |
| $\eta_{Z_{mt}}^{\sigma}(\lambda_y = 1) = \frac{1}{\text{Var}(Y_t)} \int_{w_t^*}^{\infty} [Y_t(w) - E(Y_t)] H_{Z_{mt}}(w) \varphi(w) dw,$             |                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                |                                                                                                                                                                                                                                                             |               |
| $\eta_{Z_{mt}}^{\sigma}(\lambda_y = 0) = \frac{1}{\text{Var}(Y_t)} \int_{-\infty}^{\infty} Y_t(w) [Y_t(w) - E(Y_t)] H_{Z_{mt}}(w) \varphi(w) dw$     |                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                |                                                                                                                                                                                                                                                             |               |
| and $H_{Z_{mt}}(w) = \frac{1}{2} \delta_m Z_{mt}^{\lambda_{zm}} \left[ f(Z_t)^{1/2}(A_t + w) - \sum_k \beta_k X_{kt}^{(\lambda_{xk})} \right]$ .     |                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                |                                                                                                                                                                                                                                                             |               |

## 2.4 Derivatives and elasticities of the skewness of the dependent variable

### A. General case

Since an independent variable  $X_{jt}$  can also appear as an heteroskedasticity variable  $Z_{mt}$  in the heteroskedasticity function  $f(Z_t)$ , two cases for the derivatives and elasticities of  $\gamma(Y_t)$  can be considered:

**1. The independent variable  $X_{jt}$  also appears in  $f(Z_t)$ .** The derivative and the elasticity of  $\gamma(Y_t)$  with respect to an independent variable  $X_{jt}$  can be obtained as:

$$(81) \quad \begin{aligned} D_{X_{jt}}^{\gamma} &= \frac{\partial \gamma(Y_t)}{\partial X_{jt}} = \frac{1}{\sigma^3(Y_t)} \left[ \frac{\partial \mu_3(Y_t)}{\partial X_{jt}} - \gamma(Y_t) \frac{\partial \sigma^3(Y_t)}{\partial X_{jt}} \right] \\ \eta_{X_{jt}}^{\gamma} &= \frac{\partial \gamma(Y_t)}{\partial X_{jt}} \frac{X_{jt}}{\gamma(Y_t)} \end{aligned}$$

Using (65), the derivative of  $\mu_3(Y_t)$  with respect to  $X_{jt}$  can be written as:

$$(82) \quad \frac{\partial \mu_3(Y_t)}{\partial X_{jt}} = \frac{\partial E(Y_t^3)}{\partial X_{jt}} - \left\{ \frac{\partial E(Y_t)}{\partial X_{jt}} \left[ 3E(Y_t^2) - 2(E(Y_t))^2 \right] + E(Y_t) \left[ 3 \frac{\partial E(Y_t^2)}{\partial X_{jt}} - 4E(Y_t) \frac{\partial E(Y_t)}{\partial X_{jt}} \right] \right\}$$

The derivative of  $\sigma^3(Y_t)$  with respect to  $X_{jt}$  is:

$$(83) \quad \frac{\partial \sigma^3(Y_t)}{\partial X_{jt}} = 3\sigma^2(Y_t) \frac{\partial \sigma(Y_t)}{\partial X_{jt}}$$

where

$$(84) \quad \frac{\partial \sigma(Y_t)}{\partial X_{jt}} = \frac{1}{2\sigma(Y_t)} \left[ \frac{\partial E(Y_t^2)}{\partial X_{jt}} - 2E(Y_t) \frac{\partial E(Y_t)}{\partial X_{jt}} \right]$$

In (82) and (84), the derivatives of the first three noncentral moments of  $Y_t$  with respect to  $X_{jt}$  are given by:

$$(85) \quad \frac{\partial E(Y_t^r)}{\partial X_{jt}} = r \int_{R_w} [Y_t(w)]^{r-\lambda_y} \left[ \beta_j X_{jt}^{\lambda_{xj}-1} + X_{jt}^{-1} H_{X_{jt}}(w) \right] \varphi(w) dw, \quad (r = 1, 3)$$

where  $Y_t(w)$  is defined in (51) and the heteroskedasticity term  $H_{X_{jt}}(w)$  in (71). Note that for the special case where  $X_{jt}$  is not specified in  $f(Z_t)$ , then the heteroskedasticity term  $H_{X_{jt}}(w)$  is null.

**2. The heteroskedasticity variable  $Z_{mt}$  appears only in  $f(Z_t)$ .** The derivative and the elasticity of  $\gamma(Y_t)$  with respect to  $Z_{mt}$  are analogous to (81) with  $X_{jt}$  replaced by  $Z_{mt}$ :

$$(86) \quad \begin{aligned} D_{Z_{mt}}^{\gamma} &= \frac{\partial \gamma(Y_t)}{\partial Z_{mt}} = \frac{1}{\sigma^3(Y_t)} \left[ \frac{\partial \mu_3(Y_t)}{\partial Z_{mt}} - \gamma(Y_t) \frac{\partial \sigma^3(Y_t)}{\partial Z_{mt}} \right] \\ \eta_{Z_{mt}}^{\gamma} &= \frac{\partial \gamma(Y_t)}{\partial Z_{mt}} \frac{Z_{mt}}{\gamma(Y_t)} \end{aligned}$$

The derivatives of the first three noncentral moments of  $Y_t$  with respect to  $Z_{mt}$  are given by:

$$(87) \quad \frac{\partial E(Y_t^r)}{\partial Z_{mt}} = r Z_{mt}^{-1} \int_{R_w} [Y_t(w)]^{r-\lambda_y} H_{Z_{mt}}(w) \varphi(w) dw, \quad (r = 1, 3)$$

where  $H_{Z_{mt}}(w)$  is defined in (72).

### B. Derivatives and elasticities of $\gamma(Y_t)$ for the linear and logarithmic cases of $Y_t$

For the linear case of the dependent variable, the skewness of  $Y_t$  is zero, since the third central moment of  $Y_t$  is equal to zero:

$$(88) \quad \mu_3(Y_t) = E[Y_t - E(Y_t)]^3 = E \left[ Y_t - \sum_k \beta_k X_{kt}^{(\lambda_{xk})} \right]^3 = E(u_t^3) = 0$$

where for simplicity, we suppose that  $Y_t = \sum_k \beta_k X_{kt}^{(\lambda_{xk})} + u_t$  without heteroskedasticity and autocorrelation. Since the residuals  $u_t$ 's are normally and independently distributed, all the odd-order moments of  $u_t$  are zero, in particular the first and third-order moments. Hence the derivative  $\partial \gamma(Y_t) / \partial X_{jt}$  as well as the elasticity  $\eta_{X_{jt}}^\gamma$  are all zero in the linear case of  $Y_t$ . For the logarithmic case of the dependent variable, there are no closed forms of the skewness  $\gamma(Y_t)$ , the derivatives  $D_{X_{jt}}^\gamma$  and  $D_{Z_{mt}}^\gamma$ , and the elasticities  $\eta_{X_{jt}}^\gamma$  and  $\eta_{Z_{mt}}^\gamma$ . They should be computed numerically using (64) and (81)-(87).

All these cases are summarized in Table 5.

TABLE 5 Explicit forms of the skewness of  $Y_t$  and its derivatives and elasticities for the linear and logarithmic cases of  $Y_t$ .

| STATISTIC                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | CASE            | HOMOSKEDASTICITY<br>(No $f(Z_t)$ involved)                                                                                                                                                                                                      |                                                                       | HETEROSKEDASTICITY<br>$X_{jt} \neq Z_{mt}$ <span style="float:right;"><math>X_{jt} = Z_{mt}</math></span> |  |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------|--|
| SKEWNESS<br><br>$\gamma(Y_t)$                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | $\lambda_y = 1$ | 0                                                                                                                                                                                                                                               |                                                                       |                                                                                                           |  |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | $\lambda_y = 0$ | $\gamma(Y_t) = \frac{\mu_3(Y_t)}{\sigma^3(Y_t)} = \frac{E(Y_t^3) - E(Y_t)\{3E(Y_t^2) - 2[E(Y_t)]^2\}}{\{E(Y_t^2) - [E(Y_t)]^2\}^{3/2}}$ <p style="text-align:center;">where <math>E(Y_t^r)</math>, <math>r = 1, 3</math>, is given in (66).</p> |                                                                       |                                                                                                           |  |
| DERIVATIVE<br><br>$D_{X_{jt}}^\gamma$                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | $\lambda_y = 1$ | 0                                                                                                                                                                                                                                               |                                                                       | 0                                                                                                         |  |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | $\lambda_y = 0$ | $D_{X_{jt}}^\gamma$                                                                                                                                                                                                                             |                                                                       | $D_{X_{jt}}^\gamma + D_{Z_{mt}}^\gamma$                                                                   |  |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | $\lambda_y = 1$ | Not applicable                                                                                                                                                                                                                                  | 0                                                                     | 0                                                                                                         |  |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | $\lambda_y = 0$ | Not applicable                                                                                                                                                                                                                                  | $D_{Z_{mt}}^\gamma$                                                   | $D_{X_{jt}}^\gamma + D_{Z_{mt}}^\gamma$                                                                   |  |
| ELASTICITY<br><br>$\eta_{X_{jt}}^\gamma$                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | $\lambda_y = 1$ | 0                                                                                                                                                                                                                                               |                                                                       | 0                                                                                                         |  |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | $\lambda_y = 0$ | $\eta_{X_{jt}}^\gamma = D_{X_{jt}}^\gamma \frac{X_{jt}}{\gamma(Y_t)}$                                                                                                                                                                           |                                                                       | $\eta_{X_{jt}}^\gamma + \eta_{Z_{mt}}^\gamma$                                                             |  |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | $\lambda_y = 1$ | Not applicable                                                                                                                                                                                                                                  | 0                                                                     | 0                                                                                                         |  |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | $\lambda_y = 0$ | Not applicable                                                                                                                                                                                                                                  | $\eta_{Z_{mt}}^\gamma = D_{Z_{mt}}^\gamma \frac{Z_{mt}}{\gamma(Y_t)}$ | $\eta_{X_{jt}}^\gamma + \eta_{Z_{mt}}^\gamma$                                                             |  |
| <div>where</div> $D_{X_{jt}}^\gamma = \frac{\partial \gamma(Y_t)}{\partial X_{jt}} = \frac{1}{\sigma^3(Y_t)} \left[ \frac{\partial \mu_3(Y_t)}{\partial X_{jt}} - \gamma(Y_t) \frac{\partial \sigma^3(Y_t)}{\partial X_{jt}} \right]$ <div>and</div> $D_{Z_{mt}}^\gamma = \frac{\partial \gamma(Y_t)}{\partial Z_{mt}} = \frac{1}{\sigma^3(Y_t)} \left[ \frac{\partial \mu_3(Y_t)}{\partial Z_{mt}} - \gamma(Y_t) \frac{\partial \sigma^3(Y_t)}{\partial Z_{mt}} \right]$ <div>as given in (81)</div> <div>as given in (86) .</div> |                 |                                                                                                                                                                                                                                                 |                                                                       |                                                                                                           |  |

## 2.5 Ratios of derivatives of the moments of the dependent variable

In the three preceding sections, we have considered the derivatives of the three moments of  $Y_t$  with respect to an independent variable  $X_{jt}$ , namely  $\partial E(Y_t)/\partial X_{jt}$  in (70),  $\partial \sigma(Y_t)/\partial X_{jt}$  in (73) and  $\partial \gamma(Y_t)/\partial X_{jt}$  in (81). Using these derivatives, three different relations can be obtained for the ratios of two derivatives of the moments of  $Y_t$ . Consider them in turn:

The ratio of the derivatives of a same moment with respect to two different independent variables  $X_{it}$  and  $X_{jt}$  gives the Marginal Rate of Substitution (MRS) between these two variables:

$$(89) \quad MRS_{X_i, X_j} = \frac{\partial X_{jt}}{\partial X_{it}} = \frac{\partial E(Y_t)/\partial X_{it}}{\partial E(Y_t)/\partial X_{jt}} = \frac{\partial \sigma(Y_t)/\partial X_{it}}{\partial \sigma(Y_t)/\partial X_{jt}} = \frac{\partial \gamma(Y_t)/\partial X_{it}}{\partial \gamma(Y_t)/\partial X_{jt}}.$$

This MRS can be shown below to be independent from any moment chosen for the derivatives.

The ratio of the derivatives of two different moments  $m_r$  and  $m_s$  with respect to a same independent variable  $X_{jt}$  gives the Marginal Rate of Substitution (MRS) between those two moments:

$$(90) \quad MRS^{m_r, m_s} = \frac{\partial m_r}{\partial m_s} = \frac{\partial m_r / \partial X_{jt}}{\partial m_s / \partial X_{jt}}, \quad (r \neq s = 1, 3 \text{ and } j = 1, K).$$

where  $m_1$ ,  $m_2$  and  $m_3$  represent respectively  $E(Y_t)$ ,  $\sigma(Y_t)$  and  $\gamma(Y_t)$ . It will be shown below that this MRS does not depend on any  $X_{jt}$  with respect to which the derivatives of the moments are taken.

For two different moments  $m_r$  and  $m_s$ , and two different independent variables  $X_{it}$  and  $X_{jt}$ , we have the following relation for the ratio of derivatives:

$$(91) \quad MRS_{X_i, X_j}^{m_r, m_s} = \frac{\partial m_r / \partial X_{it}}{\partial m_s / \partial X_{jt}} = \frac{\partial m_r}{\partial m_s} \frac{\partial X_{jt}}{\partial X_{it}} = MRS^{m_r, m_s} \cdot MRS_{X_i, X_j}.$$

It is a combination of the two preceding cases, i.e. a product of a MRS of moments and a MRS of variables. But the economic interpretation of this case remains to be found.

We therefore focus on the first two ratios and on the elasticity of the second one.

### A. Marginal Rate of Substitution between two independent variables

The marginal rate of substitution between two independent variables  $X_{it}$  and  $X_{jt}$  can be computed from the ratio of the derivatives of the first moment  $E(Y_t)$  with respect to  $X_{it}$  and  $X_{jt}$ . It will be shown below that this MRS does not depend on any moment chosen for the derivatives in the ratio. Since the independent variables  $X_{it}$  and  $X_{jt}$  may also appear in the heteroskedasticity function  $f(Z_t)$ , two cases of the MRS between  $X_{it}$  and  $X_{jt}$  can be considered:

**$f(Z_t)$  is not a function of  $X_{it}$  and  $X_{jt}$ .** Using (85) for the expressions of the derivatives of  $E(Y_t)$  with respect to  $X_{it}$  and  $X_{jt}$ , where the heteroskedasticity terms  $H_{X_{it}}(w)$  and  $H_{X_{jt}}(w)$

are null, the MRS can be written as:

$$(92) \quad MRS_{X_i, X_j} = \frac{\partial X_{jt}}{\partial X_{it}} = \frac{\partial E(Y_t)/\partial X_{it}}{\partial E(Y_t)/\partial X_{jt}} = \frac{\int_{R_w} [Y_t(w)]^{1-\lambda_y} \beta_i X_{it}^{\lambda_{xi}-1} \varphi(w) dw}{\int_{R_w} [Y_t(w)]^{1-\lambda_y} \beta_j X_{jt}^{\lambda_{xj}-1} \varphi(w) dw} = \frac{\beta_i X_{it}^{\lambda_{xi}-1}}{\beta_j X_{jt}^{\lambda_{xj}-1}}.$$

For example, in the context of travel demand, the marginal rate of substitution between travel time (T) and cost (C), also known as the value of time, can be computed as:

$$(93) \quad \frac{\partial C_t}{\partial T_t} = \frac{\beta_T T_t^{\lambda_T-1}}{\beta_C C_t^{\lambda_C-1}}.$$

$f(Z_t)$  is a function of both  $X_{it}$  and  $X_{jt}$ . Using (85) again for the expressions of the derivatives of  $E(Y_t)$  with respect to  $X_{it}$  and  $X_{jt}$ , where the heteroskedasticity terms  $H_{X_{it}}(w)$  and  $H_{X_{jt}}(w)$  are nonnull, the MRS cannot be obtained in closed form, but should be computed numerically:

$$(94) \quad MRS_{X_i, X_j} = \frac{\partial X_{jt}}{\partial X_{it}} = \frac{\partial E(Y_t)/\partial X_{it}}{\partial E(Y_t)/\partial X_{jt}} = \frac{\int_{R_w} [Y_t(w)]^{1-\lambda_y} \left[ \beta_i X_{it}^{\lambda_{xi}-1} + X_{it}^{-1} H_{X_{it}}(w) \right] \varphi(w) dw}{\int_{R_w} [Y_t(w)]^{1-\lambda_y} \left[ \beta_j X_{jt}^{\lambda_{xj}-1} + X_{jt}^{-1} H_{X_{jt}}(w) \right] \varphi(w) dw}.$$

Two special subcases of (94) can be noted:

- If only  $X_{it}$  appears in  $f(Z_t)$ , then  $H_{X_{jt}}(w)$  is null.
- If only  $X_{jt}$  appears in  $f(Z_t)$ , then  $H_{X_{it}}(w)$  is null.

Now we turn to the proof for the double equality of the ratios of the derivatives of the moments of  $Y_t$ :

$$(95) \quad \frac{\partial E(Y_t)/\partial X_{it}}{\partial E(Y_t)/\partial X_{jt}} = \frac{\partial \sigma(Y_t)/\partial X_{it}}{\partial \sigma(Y_t)/\partial X_{jt}}$$

and

$$(96) \quad \frac{\partial \sigma(Y_t)/\partial X_{it}}{\partial \sigma(Y_t)/\partial X_{jt}} = \frac{\partial \gamma(Y_t)/\partial X_{it}}{\partial \gamma(Y_t)/\partial X_{jt}}.$$

Using (73) for the expressions of the derivatives of  $\sigma(Y_t)$  with respect to  $X_{it}$  and  $X_{jt}$ , the ratio of these derivatives can be written as:

$$(97) \quad \begin{aligned} \frac{\partial \sigma(Y_t)/\partial X_{it}}{\partial \sigma(Y_t)/\partial X_{jt}} &= \frac{\frac{1}{2\sigma(Y_t)} \left[ \frac{\partial E(Y_t^2)}{\partial X_{it}} - 2E(Y_t) \frac{\partial E(Y_t)}{\partial X_{it}} \right]}{\frac{1}{2\sigma(Y_t)} \left[ \frac{\partial E(Y_t^2)}{\partial X_{jt}} - 2E(Y_t) \frac{\partial E(Y_t)}{\partial X_{jt}} \right]} = \frac{\frac{\partial E(Y_t)}{\partial X_{it}} \left[ \frac{\partial E(Y_t^2)}{\partial X_{it}} \frac{\partial X_{it}}{\partial E(Y_t)} - 2E(Y_t) \right]}{\frac{\partial E(Y_t)}{\partial X_{jt}} \left[ \frac{\partial E(Y_t^2)}{\partial X_{jt}} \frac{\partial X_{jt}}{\partial E(Y_t)} - 2E(Y_t) \right]} \\ &= \frac{\partial E(Y_t)/\partial X_{it}}{\partial E(Y_t)/\partial X_{jt}} \end{aligned}$$

since  $\frac{\partial E(Y_t^2)}{\partial X_{it}} \frac{\partial X_{it}}{\partial E(Y_t)} = \frac{\partial E(Y_t^2)}{\partial X_{jt}} \frac{\partial X_{jt}}{\partial E(Y_t)} = \frac{\partial E(Y_t^2)}{\partial E(Y_t)}.$

Using (81) for the expressions of the derivatives of  $\gamma(Y_t)$  with respect to  $X_{it}$  and  $X_{jt}$ , the ratio of these derivatives can be expressed as:

$$(98) \quad \frac{\partial \gamma(Y_t)/\partial X_{it}}{\partial \gamma(Y_t)/\partial X_{jt}} = \frac{\frac{1}{\sigma^3(Y_t)} \left[ \frac{\partial \mu_3(Y_t)}{\partial X_{it}} - \gamma(Y_t) \frac{\partial \sigma^3(Y_t)}{\partial X_{it}} \right]}{\frac{1}{\sigma^3(Y_t)} \left[ \frac{\partial \mu_3(Y_t)}{\partial X_{jt}} - \gamma(Y_t) \frac{\partial \sigma^3(Y_t)}{\partial X_{jt}} \right]} = \frac{\frac{\partial \sigma^3(Y_t)}{\partial X_{it}} \left[ \frac{\partial \mu_3(Y_t)}{\partial X_{it}} \frac{\partial X_{it}}{\partial \sigma^3(Y_t)} - \gamma(Y_t) \right]}{\frac{\partial \sigma^3(Y_t)}{\partial X_{jt}} \left[ \frac{\partial \mu_3(Y_t)}{\partial X_{jt}} \frac{\partial X_{jt}}{\partial \sigma^3(Y_t)} - \gamma(Y_t) \right]}.$$

By noting that

$$(99) \quad \frac{\partial \sigma^3(Y_t)/\partial X_{it}}{\partial \sigma^3(Y_t)/\partial X_{jt}} = \frac{3\sigma^2(Y_t)\partial \sigma(Y_t)/\partial X_{it}}{3\sigma^2(Y_t)\partial \sigma(Y_t)/\partial X_{jt}} = \frac{\partial \sigma(Y_t)/\partial X_{it}}{\partial \sigma(Y_t)/\partial X_{jt}}$$

and

$$(100) \quad \frac{\frac{\partial \mu_3(Y_t)}{\partial X_{it}} \frac{\partial X_{it}}{\partial \sigma^3(Y_t)}}{\frac{\partial \mu_3(Y_t)}{\partial X_{jt}} \frac{\partial X_{jt}}{\partial \sigma^3(Y_t)}} = \frac{\partial \mu_3(Y_t)}{\partial X_{jt}} \frac{\partial X_{jt}}{\partial \sigma^3(Y_t)} = \frac{\partial \mu_3(Y_t)}{\partial \sigma^3(Y_t)},$$

we obtain the equality of the ratios of the derivatives:

$$(101) \quad \frac{\partial \gamma(Y_t)/\partial X_{it}}{\partial \gamma(Y_t)/\partial X_{jt}} = \frac{\partial \sigma(Y_t)/\partial X_{it}}{\partial \sigma(Y_t)/\partial X_{jt}}.$$

## B. Marginal Rate of Substitution between two moments

Since there are three moments of  $Y_t$ , we have six possible combinations of two different moments for the ratio of the derivatives of two moments of  $Y_t$  with respect to a same independent variable  $X_{jt}$ . For each combination, we will show that this ratio is independent from any  $X_{jt}$  with respect to which the derivatives of the moments are taken:

1. **MRS between  $E(Y_t)$  and  $\sigma(Y_t)$ .** Using (73) for the expression of the derivative  $\partial \sigma(Y_t)/\partial X_{jt}$ , we can write:

$$(102) \quad \frac{\partial E(Y_t)}{\partial \sigma(Y_t)} = \frac{\partial E(Y_t)/\partial X_{jt}}{\partial \sigma(Y_t)/\partial X_{jt}} = \frac{\partial E(Y_t)/\partial X_{jt}}{\left[ \partial E(Y_t^2)/\partial X_{jt} - 2E(Y_t)\partial E(Y_t)/\partial X_{jt} \right] / [2\sigma(Y_t)]}.$$

Dividing both the numerator and the denominator by  $\partial E(Y_t)/\partial X_{jt}$ , we obtain:

$$(103) \quad MRS^{E,\sigma} = \frac{\partial E(Y_t)}{\partial \sigma(Y_t)} = \frac{2\sigma(Y_t)}{\partial E(Y_t^2)/\partial E(Y_t) - 2E(Y_t)}.$$

Clearly, this expression does not depend on any  $X_{jt}$  with respect to which the derivatives of the two moments are taken initially.

2. **MRS between  $E(Y_t)$  and  $\gamma(Y_t)$ .** Using (81) for the expression of the derivative  $\partial \gamma(Y_t)/\partial X_{jt}$ , we can write:

$$(104) \quad \frac{\partial E(Y_t)}{\partial \gamma(Y_t)} = \frac{\partial E(Y_t)/\partial X_{jt}}{\partial \gamma(Y_t)/\partial X_{jt}} = \frac{\partial E(Y_t)/\partial X_{jt}}{\left[ \partial \mu_3(Y_t)/\partial X_{jt} - \gamma(Y_t)\partial \sigma^3(Y_t)/\partial X_{jt} \right] / \sigma^3(Y_t)}.$$

Dividing both the numerator and the denominator by  $\partial E(Y_t)/\partial X_{jt}$ , we obtain:

$$(105) \quad MRS^{e,\gamma} = \frac{\partial E(Y_t)}{\partial \gamma(Y_t)} = \frac{\sigma^3(Y_t)}{\partial \mu_3(Y_t)/\partial E(Y_t) - \gamma(Y_t)\partial \sigma^3(Y_t)/\partial E(Y_t)}.$$

Clearly, this expression does not depend on any  $X_{jt}$  with respect to which the derivatives of the two moments are taken initially.

3. **MRS between  $\sigma(Y_t)$  and  $\gamma(Y_t)$ .** Using (73) and (81) for the expressions of  $\partial \sigma(Y_t)/\partial X_{jt}$  and  $\partial \gamma(Y_t)/\partial X_{jt}$  respectively, we can write:

$$(106) \quad \frac{\partial \sigma(Y_t)}{\partial \gamma(Y_t)} = \frac{\partial \sigma(Y_t)/\partial X_{jt}}{\partial \gamma(Y_t)/\partial X_{jt}} = \frac{[\partial E(Y_t^2)/\partial X_{jt} - 2E(Y_t)\partial E(Y_t)/\partial X_{jt}]/[2\sigma(Y_t)]}{[\partial \mu_3(Y_t)/\partial X_{jt} - \gamma(Y_t)\partial \sigma^3(Y_t)/\partial X_{jt}]/\sigma^3(Y_t)}.$$

Dividing both the numerator and the denominator by  $\partial E(Y_t)/\partial X_{jt}$ , we obtain:

$$(107) \quad MRS^{\sigma,\gamma} = \frac{\partial \sigma(Y_t)}{\partial \gamma(Y_t)} = \frac{\sigma^2(Y_t)[\partial E(Y_t^2)/\partial E(Y_t) - 2E(Y_t)]}{2[\partial \mu_3(Y_t)/\partial E(Y_t) - \gamma(Y_t)\partial \sigma^3(Y_t)/\partial E(Y_t)]}.$$

Clearly, this expression does not depend on any  $X_{jt}$  with respect to which the derivatives of the two moments are taken initially.

We have shown that the ratio is independent from any  $X_{jt}$  with respect to which the derivatives of two moments are taken for the three cases  $\partial E(Y_t)/\partial \sigma(Y_t)$ ,  $\partial E(Y_t)/\partial \gamma(Y_t)$  and  $\partial \sigma(Y_t)/\partial \gamma(Y_t)$ . The three remaining cases  $\partial \sigma(Y_t)/\partial E(Y_t)$ ,  $\partial \gamma(Y_t)/\partial E(Y_t)$  and  $\partial \gamma(Y_t)/\partial \sigma(Y_t)$  are just the inverse ratios of the three preceding cases. Furthermore it is interesting to note that the six MRS's are interrelated. Using the chain rule, the product of (103) and (107) gives (105):

$$(108) \quad \frac{\partial E(Y_t)}{\partial \sigma(Y_t)} \frac{\partial \sigma(Y_t)}{\partial \gamma(Y_t)} = \frac{\partial E(Y_t)}{\partial \gamma(Y_t)}.$$

Each MRS on the left hand side can then be deduced:

$$(109) \quad \frac{\partial E(Y_t)}{\partial \sigma(Y_t)} = \frac{\partial E(Y_t)}{\partial \gamma(Y_t)} / \frac{\partial \sigma(Y_t)}{\partial \gamma(Y_t)}$$

and

$$(110) \quad \frac{\partial \sigma(Y_t)}{\partial \gamma(Y_t)} = \frac{\partial E(Y_t)}{\partial \gamma(Y_t)} / \frac{\partial E(Y_t)}{\partial \sigma(Y_t)}.$$

The inverse cases  $\partial \gamma(Y_t)/\partial E(Y_t)$ ,  $\partial \sigma(Y_t)/\partial E(Y_t)$  and  $\partial \gamma(Y_t)/\partial \sigma(Y_t)$  can be obtained by taking the inverse of every component in (108), (109) and (110) respectively. Although the sign of each MRS taken separately can be:

Positive, if the derivatives of the two moments with respect to a same independent variable  $X_{jt}$  have the same sign,

Negative, if they have opposite signs,

the equation (108) implies that once the signs of any two MRS's are known, the sign of the remaining one is determined. Therefore, only four out of eight ( $2^3$ ) combinations of signs of the three MRS's are feasible. Table 6 gives these four feasible combinations.

TABLE 6 Feasible combinations of signs of the MRS's among moments.

| Feasible combinations | Signs of MRS's   |                       |                  |
|-----------------------|------------------|-----------------------|------------------|
|                       | $MRS^{e,\sigma}$ | $MRS^{\sigma,\gamma}$ | $MRS^{e,\gamma}$ |
| A                     | —                | —                     | +                |
| B                     | +                | —                     | —                |
| C                     | —                | +                     | —                |
| D                     | +                | +                     | +                |

### C. Elasticities of substitution between two moments

In the preceding section, the MRS's among the three moments were defined. Since the first two moments  $E(Y_t)$  and  $\sigma(Y_t)$  are measured in the same units as the dependent variable  $Y_t$ , whereas the skewness  $\gamma(Y_t)$  is a pure number, the MRS's where the first two moments are involved, namely  $MRS^{e,\sigma}$  and  $MRS^{\sigma,e}$ , are dimensionless in contrast to the other four MRS's where the skewness  $\gamma(Y_t)$  is involved with one of the first two moments. Therefore, the elasticity of substitution between two moments  $m_r$  and  $m_s$  should be also computed to obtain a statistic which does not depend on the measurement units of  $Y_t$ :

$$(111) \quad \eta^{m_r, m_s} = MRS^{m_r, m_s} \frac{m_s}{m_r} = \frac{\partial m_r}{\partial m_s} \frac{m_s}{m_r} .$$

Since the first moment  $E(Y_t)$  and the skewness  $\gamma(Y_t)$  can be positive or negative —  $E(Y_t)$  can be negative if the dependent variable  $Y_t$  is not transformed by Box-Cox, hence  $Y_t$  does not need to be strictly positive — whereas the standard error  $\sigma(Y_t)$  is always positive, four cases for the signs of the elasticities of substitution among moments are given in Table 7, for each feasible combination of signs of the MRS's among moments reported in Table 6. Note that the first row associated with each feasible combination of signs of the MRS's corresponds to the case where all the moments are positive, hence the signs of the elasticities of substitution are identical to those of the MRS's.

TABLE 7 Feasible combinations of signs of the elasticities of substitution among moments.

| Feasible combinations of signs of MRS's | Cases | Signs of moments |          |          | Signs of the elasticities of substitution |                        |                   |
|-----------------------------------------|-------|------------------|----------|----------|-------------------------------------------|------------------------|-------------------|
|                                         |       | e                | $\sigma$ | $\gamma$ | $\eta^{e,\sigma}$                         | $\eta^{\sigma,\gamma}$ | $\eta^{e,\gamma}$ |
| A                                       | A.1   | +                | +        | +        | −                                         | −                      | +                 |
|                                         | A.2   | −                | +        | +        | +                                         | −                      | −                 |
|                                         | A.3   | +                | +        | −        | −                                         | +                      | −                 |
|                                         | A.4   | −                | +        | −        | +                                         | +                      | +                 |
| B                                       | B.1   | +                | +        | +        | +                                         | −                      | −                 |
|                                         | B.2   | −                | +        | +        | −                                         | −                      | +                 |
|                                         | B.3   | +                | +        | −        | +                                         | +                      | +                 |
|                                         | B.4   | −                | +        | −        | −                                         | +                      | −                 |
| C                                       | C.1   | +                | +        | +        | −                                         | +                      | −                 |
|                                         | C.2   | −                | +        | +        | +                                         | +                      | +                 |
|                                         | C.3   | +                | +        | −        | −                                         | −                      | +                 |
|                                         | C.4   | −                | +        | −        | +                                         | −                      | −                 |
| D                                       | D.1   | +                | +        | +        | +                                         | +                      | +                 |
|                                         | D.2   | −                | +        | +        | −                                         | +                      | −                 |
|                                         | D.3   | +                | +        | −        | +                                         | −                      | −                 |
|                                         | D.4   | −                | +        | −        | −                                         | −                      | +                 |

## 2.6 Evaluation of moments, their derivatives, rates of substitution and elasticities

### A. Evaluation of the moments of $Y_t$

In Section 2.1, the expected values of the dependent variable  $Y_t$ , as well as its standard error and skewness are computed:

For each observation  $t$  in the estimation period.

For a subset of observations included in the estimation period. Note that this option applies only to the expected values of  $Y_t$ .

Note that the calculated value of  $Y_t$  given in (48) is not computed due to the fact that the expression in the square brackets can be negative and hence cannot be raised to the power  $1/\hat{\lambda}_y$ , if  $\hat{\lambda}_y \neq 0$ .

### B. Evaluation of the derivatives and elasticities of $Y_t$

In Sections 2.2, 2.3 and 2.4, the derivatives and elasticities of the sample and expected values of  $Y_t$ , and those of the standard error and skewness of  $Y_t$  are computed:

At the sample means of the dependent and independent variables, and the heteroskedasticity variables if any, for the estimation period.

At the sample means of these variables, for a subset of observations included in the estimation period. Note that this option applies only to the sample and expected values of  $Y_t$ .

### C. 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  $\lambda$ -parameter, so that the transformed real dummy will contain only null observations.

Two cases of the quasi-dummy should be considered:

1. **Quasi-dummy not transformed by Box-Cox.** The correction of the elasticities for the positive observations 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:

$$(112) \quad \eta_j^+ = \eta_j \frac{\bar{X}_j^+}{\bar{X}_j} = \eta_j \frac{\sum_t X_{jt}^+ / N_j^+}{\sum_t X_{jt} / N} = \eta_j \frac{N}{N_j^+}$$

where  $\eta_j$  represents  $\eta_{X_{jt}}^s$ ,  $\eta_{X_{jt}}^e$ ,  $\eta_{X_{jt}}^\sigma$  or  $\eta_{X_{jt}}^\gamma$  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 a subset of observations from the total set, the correction formula is analogous.

2. **Quasi-dummy transformed by Box-Cox.** If a quasi-dummy, say  $Q_t$ , is transformed by Box-Cox, the correction will change slightly for  $\eta_{Q_t}^e$  since the expected value of  $Y_t$  evaluated at the sample means of  $X$  and  $Z$ , say  $E(Y_t)|_{\bar{X}, \bar{Z}}$ , should be computed in fact from the sample mean of a strictly positive variable  $Q_t^*$  instead of  $Q_t$  so that the well-known properties, namely  $E(Y_t)|_{\bar{X}, \bar{Z}} = \bar{Y}$  and  $\eta_{Q_t}^e|_{\bar{X}, \bar{Z}} = \eta_{Q_t^*}^e|_{\bar{X}, \bar{Z}}$  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  $Q_t^*$  such that  $Q_t^*$  is identical to  $Q_t$  for the portion of positive observations  $Q_t^+$  and  $Q_t^*$  is equal to 1 for the portion of null observations of  $Q_t$ , since  $Q_t^{(\lambda_Q)} = 0$ , if  $Q_t^* = 1$  for any value of  $\lambda_Q$ . Thus, the correction formula in this case has the following form:

$$(113) \quad \eta_{Q_t}^+ = \eta_{Q^*}^e \frac{\bar{Q}^+}{\bar{Q}^*} = \eta_{Q^*}^e \frac{\Sigma_t Q_t^+ / N_Q^+}{\Sigma_t Q_t^* / N} = \eta_{Q^*}^e \frac{N}{N_Q^+} \left( \frac{\Sigma_t Q_t^+}{N_Q^0 + \Sigma_t Q_t^+} \right)$$

where  $\eta_{Q^*}^e$  is evaluated at the sample means of  $X$  and  $Z$ , with the mean of  $Q$  replaced by the mean of  $Q^*$ , and  $N_Q^+$  and  $N_Q^0$  are respectively the numbers of positive and null observations of  $Q_t$ .

TABLE 8 Correction of the elasticities  $\eta_{X_{jt}}^s$ ,  $\eta_{X_{jt}}^e$ ,  $\eta_{X_{jt}}^\sigma$  and  $\eta_{X_{jt}}^\gamma$  at the sample means, for the positive observations of a quasi-dummy or a real dummy.

| CATEGORY                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | ELASTICITY                                                                                  |                                                                                             |                                                                                             |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------|
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | $\eta_{X_{jt}}^s$ or $\eta_{X_{jt}}^e$                                                      | $\eta_{X_{jt}}^\sigma$                                                                      | $\eta_{X_{jt}}^\gamma$                                                                      |
| (2)<br>Quasi-dummy                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                                                                                             |                                                                                             |                                                                                             |
| • No Box-Cox                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | $\eta_j \frac{N}{N_j^+}$                                                                    | $\eta_j \frac{N}{N_j^+}$                                                                    | $\eta_j \frac{N}{N_j^+}$                                                                    |
| • With Box-Cox                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | $\eta_{Q^*} \frac{N}{N_j^+} \left( \frac{\sum_t X_{jt}^+}{N_j^0 + \sum_t X_{jt}^+} \right)$ | $\eta_{Q^*} \frac{N}{N_j^+} \left( \frac{\sum_t X_{jt}^+}{N_j^0 + \sum_t X_{jt}^+} \right)$ | $\eta_{Q^*} \frac{N}{N_j^+} \left( \frac{\sum_t X_{jt}^+}{N_j^0 + \sum_t X_{jt}^+} \right)$ |
| (3) Real<br>dummy                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                                                                                             |                                                                                             |                                                                                             |
| • No Box-Cox                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | $\eta_j \frac{N}{N_j^+}$                                                                    | $\eta_j \frac{N}{N_j^+}$                                                                    | $\eta_j \frac{N}{N_j^+}$                                                                    |
| <p>where <math>\eta_j</math> represents <math>\eta_{X_{jt}}^s</math>, <math>\eta_{X_{jt}}^e</math>, <math>\eta_{X_{jt}}^\sigma</math> or <math>\eta_{X_{jt}}^\gamma</math> evaluated at the sample means, <math>\eta_{Q^*}</math> represents <math>\eta_{Q^*}^s</math>, <math>\eta_{Q^*}^e</math>, <math>\eta_{Q^*}^\sigma</math> or <math>\eta_{Q^*}^\gamma</math> evaluated at the sample means, <math>N_j^+</math> is the number of positive observations <math>X_{jt}^+</math>'s, <math>N_j^0</math> is the number of null observations of <math>X_{jt}</math> and <math>N</math> is the total number of observations (<math>N = N_j^0 + N_j^+</math>) used for estimation.</p> |                                                                                             |                                                                                             |                                                                                             |

#### D. Evaluation of the ratios of derivatives of the moments of $Y_t$ and elasticities of substitution among moments

In Section 2.5, three different relations were obtained for the ratios of two derivatives of the moments of  $Y_t$ . The first relation gives the Marginal Rate of Substitution between two independent variables, the second the Marginal Rate of Substitution (and elasticity of substitution) between two moments, and the third the product of a MRS of moments and a MRS of variables. Since the third relation cannot be interpreted in economic terms, only the first two relations are computed at the sample means of the dependent and independent variables, and the heteroskedasticity variables if any, for the estimation period.

Table 9 summarizes the evaluations provided for the derivatives and elasticities of the moments, and for the ratios of the derivatives of the moments and their elasticities.

TABLE 9 Evaluations provided for the derivatives and elasticities of the moments, and for the ratios of the derivatives of the moments and their elasticities.

| Statistics                                 | Equation #<br>reference | Evaluation provided |        |                                                      |
|--------------------------------------------|-------------------------|---------------------|--------|------------------------------------------------------|
|                                            |                         | At the sample means |        | Corrections for<br>Quasi-dummies<br>and Real Dummies |
|                                            |                         | Full set            | Subset |                                                      |
| Derivative                                 |                         |                     |        |                                                      |
| $D_{X_{jt}}^s$                             | (67)                    | ✓                   | ✓      |                                                      |
| $D_{X_{jt}}^e$                             | (70)                    | ✓                   | ✓      |                                                      |
| $D_{X_{jt}}^\sigma$                        | (73)                    | ✓                   |        |                                                      |
| $D_{X_{jt}}^\gamma$                        | (81)                    | ✓                   |        |                                                      |
| Elasticity                                 |                         |                     |        |                                                      |
| $\eta_{X_{jt}}^s$                          | (67)                    | ✓                   | ✓      | ✓                                                    |
| $\eta_{X_{jt}}^e$                          | (70)                    | ✓                   | ✓      | ✓                                                    |
| $\eta_{X_{jt}}^\sigma$                     | (73)                    | ✓                   |        | ✓                                                    |
| $\eta_{X_{jt}}^\gamma$                     | (81)                    | ✓                   |        | ✓                                                    |
| MRS between variables                      |                         |                     |        |                                                      |
| $MRS_{X_i,X_j}$                            | (89)                    | ✓                   |        |                                                      |
| MRS between moments                        |                         |                     |        |                                                      |
| $MRS^{m_r,m_s}$                            | (90)                    | ✓                   |        |                                                      |
| Elasticity of substitution between moments |                         |                     |        |                                                      |
| $\eta^{m_r,m_s}$                           | (111)                   | ✓                   |        |                                                      |

## 2.7 Student's t-statistics

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

At the initial step of the maximization procedure, only the t-statistics for the estimated  $\beta$ -coefficients **conditional** upon the initial or fixed values of  $\lambda_y, \lambda_x, \lambda_z, \delta$  and  $\rho$  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  $Z$ .

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  $\sigma_w^2$  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  $\beta$  and  $\delta$ -coefficients due to the application of the Box-Cox transformations on these variables. For the  $\lambda$ -parameters, apart from the value 0 which corresponds to the 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  $\lambda_y, \lambda_x$  and  $\lambda_z$  are also computed under the null hypothesis that the parameter is equal to **one**.
- b. To obtain **invariant** t-tests for the  $\beta$  and  $\delta$ -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  $\beta, \delta, \rho$  and  $\sigma_w^2$  **conditionally** upon the estimates of  $\lambda_y, \lambda_x$  and  $\lambda_z$  by using the conditional covariance matrix of  $\beta, \delta, \rho$  and  $\sigma_w^2$  which is equal to the inverse of a submatrix derived from the matrix  $\left[ \sum_t \frac{\partial L_t}{\partial \Pi} \frac{\partial L_t}{\partial \Pi'} \right]_{\Pi=\hat{\Pi}}$  in (38) by deleting the rows and columns associated with  $\lambda_y, \lambda_x$  and  $\lambda_z$  (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.

## 2.8 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  $Y_t, E(Y_t)$ , defined in section 2.1 and the  $R_L^2$  measure is based on the likelihood ratio test statistic  $\Lambda$ . Since both measures are just nonlinear extensions of the standard linear regression case, they are called *pseudo- $R^2$*  measures.

### A. Pseudo-(E)- $R^2$

Instead of computing the usual  $R^2$  measure defined for a standard linear regression model:

$$(114) \quad R^2 = 1 - \frac{\sum_t [Y_t - \hat{Y}_t]^2}{\sum_t (Y_t - \bar{Y})^2}$$

which can give a negative value in the nonlinear case where the dependent variable  $Y_t$  is transformed by a Box-Cox, since the sum of squares of residuals can be greater than the total sum of squares of  $Y_t$  in deviation around its sample mean  $\bar{Y}$ , we compute the square of the simple correlation coefficient between  $Y_t$  and  $E(Y_t)$  called the Pearson  $R_E^2$  like Laferrière (1999):

— unadjusted:

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

(115)

— adjusted:

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

where  $K^*$  is the total number of estimated parameters in  $\Pi^*$  which includes all the  $\beta$ -coefficients and a set of estimated parameters in  $(\lambda_y, \lambda_x, \lambda_z, \delta, \rho_1)$ ; and  $\bar{Y}$  and  $\overline{E(Y)}$  are respectively the means of  $Y_t$  and  $E(Y_t)$ :  $\bar{Y} = \sum_t Y_t / N$  and  $\overline{E(Y)} = \sum_t E(Y_t) / N$ . This  $R_E^2$  measure will always give values within the range 0 – 1, and will be identical to the usual  $R^2$  measure if the dependent variable  $Y_t$  is linear (Johnston, 1984).

### B. Pseudo-(L)- $R^2$

— unadjusted:

$$(116) \quad R_L^2 = 1 - \Lambda^{2/N}$$

— adjusted:

$$(117) \quad \bar{R}_L^2 = 1 - \frac{N-1}{N-K^*} (1 - R_L^2)$$

where  $\Lambda$  is the ratio of the likelihood function  $\mathcal{L}$  when maximized with respect to the regression constant  $\beta_0$  only, to the one with respect to  $\Pi^*$  as defined above:

$$(118) \quad \Lambda = \max_{\beta_0} \mathcal{L} / \max_{\Pi^*} \mathcal{L}.$$

The  $R_L^2$  measure always remains inside the interval 0 – 1 since the maximum of  $\mathcal{L}$  associated with  $\beta_0$  — which corresponds to the most restrictive model where no independent variables except for the constant term are specified — is necessarily smaller than the maximum associated with  $\Pi^*$  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  $R_E^2$  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:

$$(119) \quad \Lambda^{2/N} = \frac{\sum_t (Y_t^{**} - \hat{Y}_t^{**})^2}{\sum_t (Y_t - \bar{Y})^2} \left[ \sum_t \frac{Y_t^{\lambda_y-1}}{f(Z_t)^{1/2}} \right]^{-2/N} = \frac{\sum_t (Y_t - \hat{Y}_t)^2}{\sum_t (Y_t - \bar{Y})^2}$$

where  $\hat{Y}_t^{**}$  is the calculated value of  $Y_t^{**}$  given in (46).

### C. Moments of $Y_t$ observed and estimated, and of $E(Y_t)$ estimated

In a regression model where a Box-Cox transformation is used on the dependent variable  $Y_t$ , one is interested to see how much the transformation will make the distribution of the expected values of  $Y_t$  close to the distribution of the observed values of  $Y_t$ . The first three moments (mean, standard error and skewness) computed for the observed and estimated  $Y_t$ 's and also for the expected values of  $Y_t$  given by  $E(Y_t)$  constitute a set of goodness-of-fit measures. Note that in column B of Table 10, the three moments of the estimated  $Y_t$ 's are evaluated at the sample means of the variables.

TABLE 10 Moments of  $Y_t$  observed and estimated, and of  $E(Y_t)$  estimated.

| MOMENT         | (A)<br>MOMENT OF $Y_t$<br>OBSERVED                                                | (B)<br>MOMENT OF $Y_t$<br>ESTIMATED AT<br>THE SAMPLE<br>MEANS | (C)<br>MOMENT OF $E(Y_t)$<br>ESTIMATED                                                                     |
|----------------|-----------------------------------------------------------------------------------|---------------------------------------------------------------|------------------------------------------------------------------------------------------------------------|
| Mean           | $\bar{Y} = \frac{1}{N} \sum_t Y_t$                                                | $E(Y_t)$                                                      | $\overline{E(Y)} = \frac{1}{N} \sum_t E(Y_t)$                                                              |
| Standard error | $s_Y = \sqrt{\frac{\sum_t (Y_t - \bar{Y})^2}{N-1}}$                               | $\sigma(Y_t)$                                                 | $s_{E(Y)} = \sqrt{\frac{\sum_t [E(Y_t) - \overline{E(Y)}]^2}{N-1}}$                                        |
| Skewness       | $\gamma_Y = m_3 / s_Y^3$<br>where<br>$m_3 = \frac{\sum_t (Y_t - \bar{Y})^3}{N-1}$ | $\gamma(Y_t)$                                                 | $\gamma_{E(Y)} = m_3^* / s_{E(Y)}^3$<br>where<br>$m_3^* = \frac{\sum_t [E(Y_t) - \overline{E(Y)}]^3}{N-1}$ |

## 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:

$$(120) \quad X'X = \sum_{i=1}^k \gamma_i p_i p_i'$$

where  $p_i$  is the  $(K \times 1)$  eigenvector associated with the  $i$ -th eigenvalue  $\gamma_i$  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  $\gamma_{max}$  and  $\gamma_{min}$  are respectively the greatest and smallest eigenvalues of the  $\gamma_i$ '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**:  $\eta_i = (\gamma_{max}/\gamma_i)^{1/2}, i = 1, \dots, K$ . Note that if  $\gamma_i$  is equal to  $\gamma_{min}$ , then  $\eta_i$  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  $b_k$  ( $k = 1, \dots, K$ ) is performed:

$$(121) \quad \text{Var}(b_k) = \sigma^2 \sum_{j=1}^K (p_{kj}^2 / \gamma_j)$$

The proportion of  $\text{Var}(b_k)$  associated with any  $\gamma_i$  is then computed:

$$(122) \quad \Pi_{ik} = (p_{ki}^2 / \gamma_i) / \sum_{j=1}^K (p_{kj}^2 / \gamma_j)$$

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  $\gamma_i$ 's.

TABLE 11 Eigenvalues of  $X'X$ , Condition Indexes of  $X$  and Proportions of  $\text{Var}(b_k)$ .

| EIGENVALUE                    | CONDITION            | VARIANCE-DECOMPOSITION PROPORTIONS |                      |       |                      |
|-------------------------------|----------------------|------------------------------------|----------------------|-------|----------------------|
|                               | INDEX                | Var(b <sub>1</sub> )               | Var(b <sub>2</sub> ) | . . . | Var(b <sub>K</sub> ) |
| $\gamma_1$ ( $\gamma_{min}$ ) | $\eta_1 = \kappa(X)$ | $\Pi_{11}$                         | $\Pi_{12}$           | . . . | $\Pi_{1K}$           |
| $\gamma_2$                    | $\eta_2$             | $\Pi_{21}$                         | $\Pi_{22}$           | . . . | $\Pi_{2K}$           |
| .                             | .                    | .                                  | .                    | . . . | .                    |
| .                             | .                    | .                                  | .                    | . . . | .                    |
| $\gamma_K$ ( $\gamma_{max}$ ) | $\eta_K = 1$         | $\Pi_{K1}$                         | $\Pi_{K2}$           | . . . | $\Pi_{KK}$           |

The sum of the proportions  $\Pi_{ik}$ 's in each column associated with  $\text{Var}(b_k)$  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  $\Pi_{ik}$ '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  $\Pi_{ik}$ 's can be individually small across the high  $\eta_i$ '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  $\tilde{w}_t$  that is produced by a differencing operation applied to the estimated residuals  $\hat{e}_t$ 's which can be the  $\hat{u}_t$ 's,  $\hat{v}_t$ 's or  $\hat{w}_t$ 's depending on which model type is specified:

$$(123) \quad \tilde{w}_t = (1 - B)^d (1 - B^s)^D \hat{e}_t, \quad (t = 1 + r, \dots, n)$$

where  $B$  is the backward shift operator defined as  $B\hat{e}_t = \hat{e}_{t-1}$ , hence  $B^m \hat{e}_t = \hat{e}_{t-m}$ ,  $d$  is the degree of consecutive differencing ( $0 \leq d \leq 3$ ),  $D$  is the degree of seasonal differencing ( $0 \leq D \leq 3$ ) and  $s$  is the period of seasonal differencing ( $1 \leq s \leq 31$ ). An additional constraint for  $d, D$  and  $s$  is  $d + sD \leq n - r$ . For example, if only the residuals  $\hat{e}_t$ '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  $\tilde{w}_t$ . (For a complete description and use, see Box and Jenkins (1976)).

#### A. 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  $\tilde{w}_t$ , the autocorrelation function with lag  $k$  is defined as:

$$(124) \quad \tilde{\rho}_k = \frac{E[(\tilde{w}_t - \mu_{\tilde{w}})(\tilde{w}_{t+k} - \mu_{\tilde{w}})]}{\sqrt{E[(\tilde{w}_t - \mu_{\tilde{w}})^2]E[(\tilde{w}_{t+k} - \mu_{\tilde{w}})^2]}} = \frac{Cov(\tilde{w}_t, \tilde{w}_{t+k})}{\sigma_{\tilde{w}}^2} = \frac{\tilde{\gamma}_k}{\tilde{\gamma}_0}, \left(k = 0, 1, \dots, \tilde{K}\right)$$

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

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

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

$$(125) \quad \tilde{r}_k = \frac{c_k}{c_0}, \left(k = 0, 1, \dots, \tilde{K}\right)$$

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

$$(126) \quad c_k = \sum_{t=1+r}^{n-k} (\tilde{w}_t - \tilde{\bar{w}})(\tilde{w}_{t+k} - \tilde{\bar{w}})/(n-r),$$

$\tilde{\bar{w}}$  is the sample mean of  $\tilde{w}_t$  computed as  $\sum_{t=1+r}^n \tilde{w}_t / (n-r)$  and  $c_0$  is the sample variance of  $\tilde{w}_t$

computed as  $\sum_{t=1+r}^n (\tilde{w}_t - \tilde{\bar{w}})^2 / (n-r)$ . 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.

## B. Standard error of the autocorrelation estimate

To check whether the theoretical autocorrelation  $\tilde{\rho}_k$  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:

$$(127) \quad \text{Var}(\tilde{r}_k) \simeq \frac{1}{N} \sum_{v=-\infty}^{+\infty} (\tilde{\rho}_v^2 + \tilde{\rho}_{v+k}\tilde{\rho}_{v-k} - 4\tilde{\rho}_k\tilde{\rho}_v\tilde{\rho}_{v-k} + 2\tilde{\rho}_v^2\tilde{\rho}_k^2).$$

If all the autocorrelations  $\tilde{\rho}_v$  are zero for  $v > q$ , 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  $\tilde{r}_k$  can be expressed as:

$$(128) \quad \text{Var}(\tilde{r}_k) \simeq \frac{1}{N} \left( 1 + 2 \sum_{v=1}^q \tilde{\rho}_v^2 \right), \quad (k > q).$$

To obtain an estimate of  $\text{Var}(\tilde{r}_k)$ , say  $\widehat{\text{Var}}(\tilde{r}_k)$ , the sample autocorrelations  $\tilde{r}_v$  ( $v = 1, 2, \dots, q$ ) are substituted for  $\tilde{\rho}_v$ . The square root of this estimate is referred as the **large-lag standard error**. For example, if  $q = 0$ , i.e. the series  $\tilde{w}_t$  is assumed to be completely random, then for all lags, the large-lag standard error reduces to:

$$(129) \quad \sqrt{\widehat{\text{Var}}(\tilde{r}_k)} \simeq \frac{1}{\sqrt{N}}$$

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

A second step would be to select the first lag  $q$  at which an autocorrelation coefficient was significant and then use (128) to test for significant autocorrelations at lags  $k$  longer than  $q$ . For example, if  $\tilde{r}_1$  were significant,  $\tilde{r}_2, \tilde{r}_3, \dots$  could be tested using the standard error  $[(1 + 2\tilde{r}_1^2)/N]^{1/2}$ . For moderate  $N$ , the distribution of an estimated autocorrelation coefficient, whose theoretical value is zero, is approximately Normal (Anderson, 1942): on the hypothesis that  $\tilde{\rho}_k$  is zero, the estimate  $\tilde{r}_k$  divided by its standard error will follow approximately a unit Normal distribution.

### C. 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  $\ell \leq p$  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  $j$ th coefficient in an autoregressive process of order  $\ell$  so that  $\phi_{\ell\ell}$  is the last coefficient. The  $\phi_{\ell j}$ 's can be shown to satisfy the set of equations:

$$(130) \quad \tilde{\rho}_j = \phi_{\ell 1}\tilde{\rho}_{j-1} + \phi_{\ell 2}\tilde{\rho}_{j-2} + \dots + \phi_{\ell\ell}\tilde{\rho}_{j-\ell}, \quad (j = 1, 2, \dots, \ell)$$

leading to the Yule-Walker equations:

$$(131) \quad \begin{bmatrix} 1 & \tilde{\rho}_1 & \tilde{\rho}_2 & \dots & \tilde{\rho}_{\ell-1} \\ \tilde{\rho}_1 & 1 & \tilde{\rho}_1 & \dots & \tilde{\rho}_{\ell-2} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \tilde{\rho}_{\ell-1} & \tilde{\rho}_{\ell-2} & \tilde{\rho}_{\ell-3} & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_{\ell 1} \\ \phi_{\ell 2} \\ \vdots \\ \phi_{\ell\ell} \end{bmatrix} = \begin{bmatrix} \tilde{\rho}_1 \\ \tilde{\rho}_2 \\ \vdots \\ \tilde{\rho}_\ell \end{bmatrix}.$$

The coefficient  $\phi_{\ell\ell}$ , regarded as a function of the lag  $\ell$ , is called the **partial autocorrelation function**. It measures the correlation between  $\tilde{w}_t$  and  $\tilde{w}_{t-\ell}$  given  $\tilde{w}_{t-1}, \dots, \tilde{w}_{t-(\ell-1)}$ . The estimated partial autocorrelations  $\hat{\phi}_{\ell j}$ 's can be obtained by substituting the sample autocorrelations  $\tilde{r}_j$ 's for the  $\tilde{\rho}_j$ 's in (130) to yield:

$$(132) \quad \tilde{r}_j = \hat{\phi}_{\ell 1} \tilde{r}_{j-1} + \hat{\phi}_{\ell 2} \tilde{r}_{j-2} + \dots + \hat{\phi}_{\ell \ell} \tilde{r}_{j-\ell}, \quad (j = 1, 2, \dots, \ell)$$

and solving the resulting Yule-Walker equations.

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

$$(133) \quad \hat{\phi}_{\ell\ell} = \begin{cases} \tilde{r}_1 & \text{if } \ell = 1, \\ \frac{\tilde{r}_\ell - \sum_{j=1}^{\ell-1} \hat{\phi}_{\ell-1,j} \tilde{r}_{\ell-j}}{1 - \sum_{j=1}^{\ell-1} \hat{\phi}_{\ell-1,j} \tilde{r}_j} & \text{if } \ell = 2, 3, \dots, \tilde{L}. \end{cases}$$

where  $\hat{\phi}_{\ell j} = \hat{\phi}_{\ell-1,j} - \hat{\phi}_{\ell\ell} \hat{\phi}_{\ell-1,\ell-j}$  ( $j = 1, 2, \dots, \ell-1$ ) and  $\tilde{L}$  is the maximum lag of the partial autocorrelation function. Note that  $\tilde{L}$  should be less than or equal to  $\tilde{K}$  which is the maximum lag of the autocorrelation function used in (124).

#### D. 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:

$$(134) \quad \sqrt{\widehat{\text{Var}}(\hat{\phi}_{\ell\ell})} \simeq \frac{1}{\sqrt{N}}, \quad (\ell > p)$$

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  $\hat{\phi}_{\ell\ell}$  divided by its standard error is approximately distributed as a unit Normal deviate.

### 3.4 Forecasting: Maximum likelihood and simulation forecasts

#### A. Maximum likelihood forecast

To obtain the predicted values of the dependent variable over the  $p$  periods for  $t = n+1, \dots, n+p$ , 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  $\sum_{t=1+r}^{n+1} L_t = \sum_{t=1+r}^n L_t + L_{n+1}$  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:

$$(135) \quad \frac{\partial L_{n+1}}{\partial Y_{n+1}} = \begin{cases} \frac{1}{Y_{n+1}} \left[ \lambda_y - 1 - \frac{Y_{n+1}}{\sigma_w^2 f(Z_{n+1})^{1/2}} \left( \frac{u_{n+1}}{f(Z_{n+1})^{1/2}} - \sum_{\ell} \rho_{\ell} \frac{u_{n+1-\ell}}{f(Z_{n+1-\ell})^{1/2}} \right) \right] & \text{if } \lambda_y \neq 0 \\ \frac{1}{Y_{n+1}} \left[ -1 - \frac{1}{\sigma_w^2 f(Z_{n+1})^{1/2}} \left( \frac{u_{n+1}}{f(Z_{n+1})^{1/2}} - \sum_{\ell} \rho_{\ell} \frac{u_{n+1-\ell}}{f(Z_{n+1-\ell})^{1/2}} \right) \right] & \text{if } \lambda_y = 0 \end{cases}$$

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

**$\lambda_y \neq 0$ .** Two subcases must be distinguished:

- (i) 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  $Y_{n+1}^{\lambda_y}$ :

$$(136) \quad Y_{n+1}^{2\lambda_y} + B Y_{n+1}^{\lambda_y} + C = 0$$

where

$$(137) \quad B = - \left[ 1 + \lambda_y f(Z_{n+1})^{1/2} \left( \sum_k \beta_k \frac{X_{k,n+1}^{(\lambda_{xk})}}{f(Z_{n+1})^{1/2}} + \sum_{\ell} \rho_{\ell} \frac{u_{n+1-\ell}}{f(Z_{n+1-\ell})^{1/2}} \right) \right]$$

and

$$(138) \quad C = -\lambda_y(\lambda_y - 1)\sigma_w^2 f(Z_{n+1}).$$

To ensure that the predicted value of  $Y_{n+1}$ , say  $\hat{Y}_{n+1}$ , is real and positive, both roots of (136), namely  $Y_{n+1}^{\lambda_y} = (-B \pm \sqrt{B^2 - 4C})/2$ , 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  $\partial^2 L_{n+1} / \partial Y_{n+1}^2 < 0$ .

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

$$(139) \quad \frac{\partial L_{n+1}}{\partial Y_{n+1}} = -\frac{1}{\sigma_w^2 f(Z_{n+1})^{1/2}} \left( \frac{u_{n+1}}{f(Z_{n+1})^{1/2}} - \sum_{\ell} \rho_{\ell} \frac{u_{n+1-\ell}}{f(Z_{n+1-\ell})^{1/2}} \right)$$

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

$$(140) \quad \hat{Y}_{n+1} = 1 + \sum_k \beta_k X_{k,n+1}^{(\lambda_{xk})} + f(Z_{n+1})^{1/2} \sum_{\ell} \rho_{\ell} \frac{u_{n+1-\ell}}{f(Z_{n+1-\ell})^{1/2}}.$$

$\lambda_y = 0$ . Setting the first derivative equal to zero and solving for  $Y_{n+1}$  also give a unique maximum likelihood predicted value of  $Y_{n+1}$ :

$$(141) \quad \hat{Y}_{n+1} = \exp \left[ \sum_k \beta_k X_{k,n+1}^{(\lambda_{xk})} + f(Z_{n+1})^{1/2} \sum_\ell \rho_\ell \frac{u_{n+1-\ell}}{f(Z_{n+1-\ell})^{1/2}} - \sigma_w^2 f(Z_{n+1}) \right],$$

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

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 (135) – (141) with  $n + 1$  replaced by  $n + i$ .

## B. Variance of the forecast error

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

$$(142) \quad COV_{(1)} = \begin{bmatrix} COV_{(0)} & C_{(1)} \\ C'_{(1)} & \sigma_{n+1}^2 \end{bmatrix} = - \begin{bmatrix} \partial^2 L_{(1)} / \partial \Pi \partial \Pi' & \partial^2 L_{(1)} / \partial \Pi \partial Y_{n+1} \\ \partial^2 L_{(1)} / \partial Y_{n+1} \partial \Pi' & \partial^2 L_{(1)} / \partial Y_{n+1}^2 \end{bmatrix}^{-1}$$

where

$COV_{(0)}$  is the covariance matrix of  $\hat{\Pi}$  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,

$C_{(1)}$  is the column vector of covariances between the elements of  $\hat{\Pi}$  and  $\hat{Y}_{n+1}$ :

$$(143) \quad C_{(1)} = -COV_{(0)} \left[ \frac{\partial^2 L_{(1)}}{\partial \Pi \partial Y_{n+1}} \right] \left[ \frac{\partial^2 L_{(1)}}{\partial Y_{n+1}^2} \right]^{-1}$$

and  $\sigma_{n+1}^2$  is the variance of the forecast error for  $Y_{n+1}$ :

$$(144) \quad \sigma_{n+1}^2 = - \left[ \frac{\partial^2 L_{(1)}}{\partial Y_{n+1}^2} \right]^{-1} + C'_{(1)} COV_{(0)}^{-1} C_{(1)}.$$

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

$$(145) \quad COV_{(i)} = \begin{bmatrix} COV_{(i-1)} & C_{(i)} \\ C'_{(i)} & \sigma_{n+i}^2 \end{bmatrix} = - \begin{bmatrix} \partial^2 L_{(i)} / \partial \Pi_{(i-1)} \partial \Pi'_{(i-1)} & \partial^2 L_{(i)} / \partial \Pi_{(i-1)} \partial Y_{n+i} \\ \partial^2 L_{(i)} / \partial Y_{n+i} \partial \Pi'_{(i-1)} & \partial^2 L_{(i)} / \partial Y_{n+i}^2 \end{bmatrix}^{-1}$$

where

$COV_{i-1}$  is the covariance matrix of  $\hat{\Pi}_{(i-1)} = (\hat{\Pi}', \hat{Y}_{n+1}, \dots, \hat{Y}_{n+i-1})'$  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,  
 $C_{(i)}$  is the column vector of covariances between the elements of  $\hat{\Pi}_{(i-1)}$  and  $\hat{Y}_{n+i}$ :

$$(146) \quad C_{(i)} = -COV_{(i-1)} \left[ \frac{\partial^2 L_{(i)}}{\partial \Pi_{(i-1)} \partial Y_{n+i}} \right] \left[ \frac{\partial^2 L_{(i)}}{\partial Y_{n+i}^2} \right]^{-1}$$

and  $\sigma_{n+i}^2$  is the variance of the forecast error for  $Y_{n+i}$ :

$$(147) \quad \sigma_{n+i}^2 = - \left[ \frac{\partial^2 L_{(i)}}{\partial Y_{n+i}^2} \right]^{-1} + C'_{(i)} COV_{(i-1)}^{-1} C_{(i)} .$$

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  $\partial^2 L_{(i)} / \partial Y_{n+i}^2 = \partial^2 L_{n+i} / \partial Y_{n+i}^2$ , 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  $\partial^2 L_{n+i} / \partial Y_{n+i}^2$  are in the form  $Y_{n+i}^{\lambda-1} / \left[ \sigma_w^2 f(Z_{n+i})^{1/2} \right]'$  times a component which is given in Table 12 for each element of  $\Pi_{(i)} = (\Pi'_{(i-1)}, Y_{n+i})$  crossed with  $Y_{n+i}$ .

TABLE 12 Components of the second derivatives of  $L_{n+i}$  with respect to  $\Pi_{(i)}$  and  $Y_{n+i}$ 

| $\Pi_{(i)}$    | $Y_{n+i}$                                                                                                                                                                                                                                                                                                                                                               |
|----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| $\beta_k$      | $\frac{X_{k,n+i}^{(\lambda_{xk})}}{f(Z_{n+i})^{1/2}} - \sum_{\ell} \rho_{\ell} \frac{X_{k,n+i-\ell}^{(\lambda_{xk})}}{f(Z_{n+i-\ell})^{1/2}}$                                                                                                                                                                                                                           |
| $\sigma_w^2$   | $w_{n+i} / \sigma_w^2$                                                                                                                                                                                                                                                                                                                                                  |
| $\lambda_y$    | $\frac{\sigma_w^2 f(Z_{n+i})^{1/2}}{Y_{n+i}^{\lambda_y}} - w_{n+i} \ln Y_{n+i} - \left[ \frac{1}{f(Z_{n+i})^{1/2}} \frac{\partial Y_{n+i}^{(\lambda_y)}}{\partial \lambda_y} - \sum_{\ell} \frac{\rho_{\ell}}{f(Z_{n+i-\ell})^{1/2}} \frac{\partial Y_{n+i-\ell}^{(\lambda_y)}}{\partial \lambda_y} \right]$                                                            |
| $\lambda_{xk}$ | $\beta_k \left[ \frac{1}{f(Z_{n+i})^{1/2}} \frac{\partial X_{k,n+i}^{(\lambda_{xk})}}{\partial \lambda_{xk}} - \sum_{\ell} \frac{\rho_{\ell}}{f(Z_{n+i-\ell})^{1/2}} \frac{\partial X_{k,n+i-\ell}^{(\lambda_{xk})}}{\partial \lambda_{xk}} \right]$                                                                                                                    |
| $\lambda_{zm}$ | $\frac{\delta_m}{2} \left[ \left( \frac{2u_{n+i}}{f(Z_{n+i})^{1/2}} - \sum_{\ell} \rho_{\ell} \frac{u_{n+i-\ell}}{f(Z_{n+i-\ell})^{1/2}} \right) \frac{\partial Z_{m,n+i}^{(\lambda_{zm})}}{\partial \lambda_{zm}} - \sum_{\ell} \rho_{\ell} \frac{u_{n+i-\ell}}{f(Z_{n+i-\ell})^{1/2}} \frac{\partial Z_{m,n+i-\ell}^{(\lambda_{zm})}}{\partial \lambda_{zm}} \right]$ |
| $\delta_m$     | $\frac{1}{2} \left[ \left( \frac{2u_{n+i}}{f(Z_{n+i})^{1/2}} - \sum_{\ell} \rho_{\ell} \frac{u_{n+i-\ell}}{f(Z_{n+i-\ell})^{1/2}} \right) Z_{m,n+i}^{(\lambda_{zm})} - \sum_{\ell} \rho_{\ell} \frac{u_{n+i-\ell}}{f(Z_{n+i-\ell})^{1/2}} Z_{m,n+i-\ell}^{(\lambda_{zm})} \right]$                                                                                      |
| $\rho_{\ell}$  | $\frac{Y_{n+i-\ell}^{(\lambda_y)} - \sum_k \beta_k X_{k,n+i-\ell}^{(\lambda_{xk})}}{f(Z_{n+i-\ell})^{1/2}}$                                                                                                                                                                                                                                                             |
| $Y_{n+i-s}$    | $\frac{\rho_s}{f(Z_{n+i-s})^{1/2}} Y_{n+i-s}^{(\lambda_y-1)} \quad (s = 1, \dots, \min(i-1, r) \text{ and } i \geq 2)$                                                                                                                                                                                                                                                  |
| $Y_{n+i}$      | $-\left(2Y_{n+i}^{(\lambda_y)} + B\right) / \left[Y_{n+i} f(Z_{n+i})^{1/2}\right]$                                                                                                                                                                                                                                                                                      |

In Table 12, 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  $L(i) = \sum_{t=1+r}^{n+i} L_t$ , the predicted values of  $(Y_{n+1}, \dots, Y_{n+i})$  obtained by the approach above are used just as when the predicted value of  $Y_{n+i}$  is computed, the predicted values of  $(Y_{n+1}, \dots, Y_{n+i-1})$  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  $t = n+1, \dots, n+p$ :

$$(148) \quad \hat{e}_t = \left(1 - \frac{\hat{Y}_t}{Y_t^0}\right) 100$$

where  $\hat{Y}_t$  and  $Y_t^0$  are respectively the predicted and observed values of  $Y_t$ . The 95% confidence interval for  $Y_t$  is also computed:

$$(149) \quad \hat{Y}_t - 1.96\hat{\sigma}_t \leq Y_t \leq \hat{Y}_t + 1.96\hat{\sigma}_t$$

where  $\hat{\sigma}_t$  is estimated from (147).

### C. Simulation forecast

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

$$(150) \quad \tilde{Y}_t = \begin{cases} \left[1 + \lambda_y f(Z_t)^{1/2} (\Sigma_\ell \rho_\ell Y_{t-\ell}^* + \Sigma_k \beta_k X_{kt}^{**})\right]^{1/\lambda_y} & \text{if } \lambda_y \neq 0 \\ \exp \left[ f(Z_t)^{1/2} \left( \Sigma_\ell \rho_\ell \frac{\ln Y_{t-\ell}}{f(Z_{t-\ell})^{1/2}} + \Sigma_k \beta_k X_{kt}^{**} \right) \right] & \text{if } \lambda_y = 0 \end{cases}$$

where  $Y_{t-\ell}$  is replaced by  $\tilde{Y}_{t-\ell}$  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.

**$\lambda_y \neq 0$ .** Two subcases should be considered:

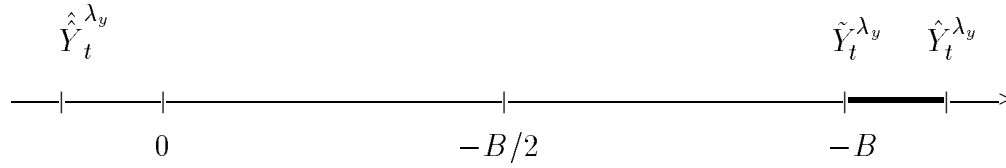
**Subcase  $\lambda_y \neq 1$**  : If both  $\tilde{Y}_t$  and  $\hat{Y}_t$  are positive, the bias  $\tilde{Y}_t - \hat{Y}_t$ , or equivalently  $\tilde{Y}_t^{\lambda_y} - \hat{Y}_t^{\lambda_y}$ , can be shown to be negative if  $\lambda_y < 0$  or  $\lambda_y > 1$  and positive if  $0 < \lambda_y 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

For the first period of forecast  $t = n+1$ , if the two roots of (136) are denoted by  $\hat{Y}_t^{\lambda_y} = (-B + \sqrt{B^2 - 4C})/2$  and  $\tilde{Y}_t^{\lambda_y} = (-B - \sqrt{B^2 - 4C})/2$ , then replacing  $-B$  by its value which is equal to  $\tilde{Y}_t^{\lambda_y}$  in the greater root  $\hat{Y}_t^{\lambda_y}$  yields:

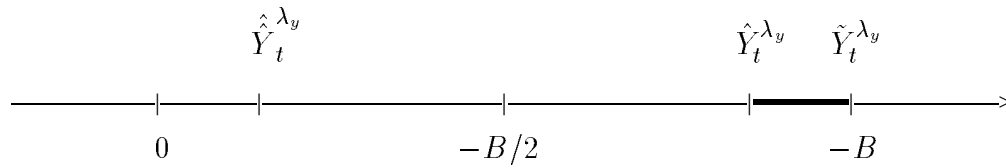
$$(151) \quad \hat{Y}_t^{\lambda_y} = \tilde{Y}_t^{\lambda_y} \left(1 + \sqrt{1 - 4C/\tilde{Y}_t^{2\lambda_y}}\right)/2.$$

If  $\lambda_y < 0$  or  $\lambda_y > 1$ , then  $C$  is negative,  $\sqrt{1 - 4C/\tilde{Y}_t^{2\lambda_y}} > 1$  and the bias  $\tilde{Y}_t^{\lambda_y} - \hat{Y}_t^{\lambda_y}$  is negative. If  $0 < \lambda_y < 1$ , then  $C$  is positive,  $\sqrt{1 - 4C/\tilde{Y}_t^{2\lambda_y}} < 1$  and the bias  $\tilde{Y}_t^{\lambda_y} - \hat{Y}_t^{\lambda_y}$  is positive. Furthermore, since  $\tilde{Y}_t^{\lambda_y} = \hat{Y}_t^{\lambda_y} + \hat{\tilde{Y}}_t^{\lambda_y}$ , the bias  $\tilde{Y}_t^{\lambda_y} - \hat{Y}_t^{\lambda_y}$  will be given by the smaller root  $\hat{\tilde{Y}}_t^{\lambda_y}$ . These results can be illustrated by the following diagrams:

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



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



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

**Subcase  $\lambda_y = 1$**  : In this case, the bias  $\tilde{Y}_t - \hat{Y}_t$  is null, i.e. the simulated value of the dependent variable  $\tilde{Y}_t$  is identical to the maximum likelihood predicted value  $\hat{Y}_t$  given in (140):

$$\begin{aligned}
 \tilde{Y}_t &= 1 + f(Z_t)^{1/2} \left( \sum_{\ell} \rho_{\ell} Y_{t-\ell}^* + \sum_k \beta_k X_{kt}^{**} \right) \\
 (152) \quad &= 1 + f(Z_t)^{1/2} \left[ \sum_{\ell} \rho_{\ell} \left( Y_{t-\ell}^* - \sum_k \beta_k X_{k,t-\ell}^* \right) + \sum_k \beta_k X_{kt}^* \right] \\
 &= 1 + \sum_k \beta_k X_{kt}^{(\lambda_x t)} + f(Z_t)^{1/2} \sum_{\ell} \rho_{\ell} \frac{u_{t-\ell}}{f(Z_{t-\ell})^{1/2}} \\
 &= \hat{Y}_t
 \end{aligned}$$

$\lambda_y = 0$ . In this case, the bias  $\tilde{Y}_t - \hat{Y}_t$  is positive:

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

since  $\hat{Y}_t$  contains in the argument of the exponential (141) an additional negative term equal to  $-\sigma_w^2 f(Z_t)$ .

#### 4. REFERENCES

- Abramowitz, M. and I.A. Stegun (1965). *Handbook of Mathematical Functions*. Dover Publications, New York.
- Amemiya, T. (1977). A Note on a Heteroscedastic Model. *Journal of Econometrics*, **6**, 365–370; and “Corrigenda”. *Journal of Econometrics*, **8**, 275.
- Anderson, R.L. (1942). Distribution of the Serial Correlation Coefficient. *Annals of Mathematical Statistics*, **13**, 1.
- Appelbaum, E. (1979). On the Choice of Functional Forms. *International Economic Review*, **20**, 449–458.
- 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.
- Belsley, D., E. Kuh and R.E. Welsh (1980). *Regression Diagnostics*. John Wiley, New York.
- 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.
- Berndt, E.K. and M.S. Khaled (1979). Parametric Productivity Measurement and Choice among Flexible Functional Form. *Journal of Political Economy*, **87**, 1220–1245.
- Blum, U. and M. Gaudry (1999). “SNUS-2.5, a Multimoment Analysis of Road Demand, Accidents and their Severity in Germany, 1968–1989”. Publication CRT-99-07, Centre de recherche sur les transports, Université de Montréal; Working Paper No. 99-xx, Bureau d’Économie Théorique et Appliquée, Université Louis Pasteur, Strasbourg; Dresdner Beiträge zur Volkswirtschaftslehre Nr. 4/99, Dresden University of Technology, 50 p., May 1999.
- Gaudry, M. and S. Lassarre, (eds), *Structural Road Accident Models: The International DRAG Family*, Elsevier Science, Oxford, Ch.3.
- Box, G.P. and D.R. Cox (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society, Series B*, **26**, 211–243.
- Box, G.P. and G.M. Jenkins (1976). *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, Revised Edition.

- 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.
- 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.
- 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.
- Egy, D. and K. Lahiri (1979). On Maximum Likelihood Estimation of Functional Form and Heteroscedasticity. *Economics Letters*, 2, 155–159.
- Fletcher, R. and M.J.D. Powell (1963). A Rapidly Convergent Descent Method for Minimization. *Computer Journal*, 6, 163–168.
- Froehlich, B.R. (1973). Some Estimators for a Random Coefficient Regression Model. *Journal of the American Statistical Association*, 68, 329–334.
- Gaudry, M. *et al.* (1993, 1994, 1995, 1996, 1997). Cur cum TRIO? Publication CRT-901. Centre de recherche sur les transports, Université de Montréal, 20 p.
- Gaudry, M.J.I. and M.G. Dagenais (1978). The Use of Box-Cox Transformations in Regression Models with Heteroskedastic Autoregressive Residual. Cahier no. 7814, Département de sciences économiques, Université de Montréal.
- Gaudry, M.J.I. and M.G. Dagenais (1979). Heteroscedasticity and the Use of Box-Cox Transformations. *Economic Letters*, 2, 225–229.
- Gaudry, M.J.I. and J.J. Wills (1978). Estimating the Functional Form of Travel Demand Models. *Transportation Research*, 12, 253–289.
- Geary, R.C. (1966). A Note on Residual Heterovariance and Estimation Efficiency in Regression. *American Statistician*, 20, (4), 30–31.
- Glejser, H. (1969). A New Test for Heteroscedasticity. *Journal of the American Statistical Association*, 64, 316–323.
- Goldfeld, S.M. and R.E. Quandt (1972). *Nonlinear Methods in Econometrics*, North-Holland, Amsterdam.
- Harvey, A.C. (1974). Estimation of Parameters in a Heteroscedastic Regression Model. *European Meeting of the Econometric Society*, Grenoble, France.
- Harvey, A.C. (1976). Estimating Regression Models with Multiplicative Heteroscedasticity. *Econometrica*, 44, 461–465.
- 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.
- 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.
- 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.
- Johnston, J. (1984). *Econometric Methods*, Third Edition, McGraw-Hill, New York.
- 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.

- 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.
- Kmenta, J. (1971). *Elements of Econometrics*, Macmillan, New York.
- Laferrière, R. (1999). Apply the Pearson Coefficient as a Measure of  $R^2$  in Nonlinear Models. Publication CRT-99-25, Centre de recherche sur les transports, Université de Montréal.
- Liem, T.C. (1979). A Program for Box-Cox Transformations in Regression Models with Heteroskedastic and Autoregressive Residuals. Publication CRT-134, Centre de recherche sur les transports, and Cahier #7917, Département de sciences économiques, Université de Montréal, 60 p.. Refereed in *The American Statistician*, Vol. 34, no. 2, p. 121, May 1980.
- Liem, T.C., Dagenais, M. and M. Gaudry (1983). L-1.1: A Program for Box-Cox Transformations in Regression Models with Heteroskedastic and Autoregressive Residuals. Publication CRT-301, Centre de recherche sur les transports, and Cahier #8314, Département de sciences économiques, Université de Montréal, 70 p..
- Liem, T.C., Dagenais, M. and M. Gaudry (1987, 1990, 1993). LEVEL: The L-1.4 program for BC-GAUHESEQ regression —Box-Cox Generalized Autoregressive Heteroskedastic Single Equation models. Publication CRT-510, Centre de recherche sur les transports, Université de Montréal, 41 p.
- Liem, T.C., Dagenais, M., Gaudry, M. und R. Koblo (1986). Ein Programm für Box-Cox Transformationen in Regressionsmodellen mit heteroskedastischen und autoregressiven Residuen. Publication CRT-301-D, Centre de recherche sur les transports, Université de Montréal, and Discussion paper 4/86, Institut für Wirtschaftspolitik und Wirtschaftsforschung, Universität Karlsruhe, 44 p.
- Park, R.E. (1966). Estimation with Heteroskedastic Error Terms. *Econometrica*, **34**, 888.
- Quenouille, M.H. (1949). Approximate Tests of Correlation in Time Series. *Journal of the Royal Statistical Society*, Series B, **11**, 68.
- 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.
- Schlesselman, J. (1971). Power Families: A Note on the Box and Cox Transformation. *Journal of the Royal Statistical Society*, Series B, **33**, 307–311.
- Spitzer, J.J. (1976). The Demand for Money, the Liquidity Trap, and Functional Forms. *International Economic Review*, **17**, 220–227.
- Spitzer, J.J. (1977). A Simultaneous Equations System of Money Demand and Supply Using Generalized Functional Forms. *Journal of Econometrics*, **5**, 117–128.
- Theil, H. (1971). *Principles of Econometrics*, John Wiley, New York.
- Tobin, J. (1958). Estimation of Relationships for Limited Dependent Variables. *Econometrica*, **25**, 24–36.
- Welland, J.D. (1976). Cognitive Abilities, Schooling and Earnings: the Question of Functional Form. Working paper 76-14, Department of Economics, McMaster University, Hamilton.
- White, K.J. (1972). Estimation of the Liquidity Trap with a Generalized Functional Form. *Econometrica*, **40**, 193–199.

- Zarembka, P. (1968), Functional Form in the Demand for Money. *Journal of the American Statistical Association*, **63**, 502–511.
- Zarembka, P. (1974). Transformation of Variables in Econometrics. P. Zarembka (ed.), *Frontiers in Econometrics*. Academic Press, New York.