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-2.0 algorithm originally published as: Liem, T. and M. Gaudry, "LEVEL: The L-2.0 program for BC-DAUHESEQ regressions – Box-Cox Directed AUtoregressive HEteroskedastic Single EQuation models", Publication CRT-972, Centre de recherche sur les transports, Université de Montréal, June 1994, and 1998. This document is complemented by a user guide for the program: Tran, C-L., and Gaudry, M., "L-2.1 program user guide with databases, examples and Tablex outputs", Publication AJD-109, Agora Jules Dupuit, Université de Montréal, April 2008.

PUBLICATION
AJD-108

LEVEL: The L-2.1 estimation procedures for BC-DAUHESEQ (Box-Cox Directed AUtoregressive HEteroskedastic Single EQuation) regression

Cong-Liem Tran[1] and Marc Gaudry[1,2]

[1] Agora Jules Dupuit, Université de Montréal.
[2] Département de sciences économiques, Université de Montréal.
(B.P. 6128, Succursale centre-ville, Montréal, QC, Canada H3C 3J7)

Publication AJD-108 Agora Jules Dupuit
Université de Montréal
April 2008, revised September 2008

ABSTRACT

L-2.1 is a program designed to deal with the specification of the functional form in the generalized single-equation regression model where the functional form of heteroskedasticity of the residuals, which may also be autocorrelated using an R-Koyck process for each autoregressive order considered, can be estimated. This process is the analog of the usual Koyck transformation on a time series, but extended to a matrix R which is called a contiguity matrix in the context of spatial econometrics and which can be constructed according to any set of rules defined or DIRECTED by the analyst. For each order of correlation ℓ, the R-Koyck process only adds to the usual autoregressive parameter ρ_ℓ a proximity parameter π_ℓ which measures the relative weight of close and remote neighbors in an infinite distributed lag structure. A maximum likelihood procedure is used to estimate jointly all the parameters including the regression coefficients, the coefficients in the heteroskedasticity function, the autocorrelation and proximity parameters as well as those of the Box-Cox transformations on the dependent, independent and heteroskedasticity variables.

Keywords: Box-Cox transformation; Heteroskedasticity; R-Koyck lag; Distributed lags; Contiguity matrix; Spatial autocorrelation; Directed autocorrelation.

RÉSUMÉ

L-2.1 est un programme destiné à l'étude de la forme fonctionnelle du modèle de régression multiple où la forme de l'hétéroscédasticité des erreurs, qui peuvent être aussi autocorrelées selon un processus R-Koyck associé à chaque ordre autorégressif considéré, peut être estimée. Ce processus est l'analogue de la transformation usuelle de Koyck d'une série chronologique, mais étendue à une matrice R qui s'appelle matrice de contiguité dans le contexte de l'économétrie spatiale et qui peut être obtenue en appliquant des règles de construction définies ou DIRIGÉES par le chercheur. Pour chaque ordre de corrélation ℓ, le processus R-Koyck n'ajoute au paramètre d'autocorrélation ρ_ℓ qu'un paramètre de proximité π_ℓ qui mesure l'importance relative des voisins proches et éloignés selon une structure de retards échelonnés de degré infini. Une procédure de maximum de vraisemblance est utilisée pour estimer conjointement tous les paramètres incluant les coefficients de régression, les coefficients dans la fonction d'hétéroscédasticité, les paramètres d'autocorrélation et de proximité ainsi que ceux des transformations de Box-Cox sur les variables dépendante et indépendantes, et celles de l'hétéroscédasticité.

Mots-clés: Transformation Box-Cox; Hétéroscédasticité; Retard R-Koyck; Retards échelonnés; Matrice de contiguité; Autocorrélation spatiale; Autocorrélation dirigée.

ZUSAMMENFASSUNG

Das L-2.1 Programm wurde entworfen für Spezifikation der funktionalen Form einer generalisierten Gleichung in Regressionsmodellen, wobei die Functionsform der Heteroskedastizität der Residuen, welche auch autokorreliert sein können unter Verwendung eines R-Koyck Prozesses für jede berücksichtigte Autokorrelationsordnung, geschätzt werden kann. Dieser Prozess stellt das Analogon zu der bekannten Koyck Transformation bei Zeitreihen dar, der jedoch zusätzlich um eine Matrix R (auch 'Nachbarschaftmatrix' genannt) zur Erfassung räumlichen ökonometrischer Beziehungen angereichert wurde. Für jede Korrelationsordung ℓ fügt der R-Koyck Prozess zu dem üblichen Autoregressionsparameter ρ_ℓ ein Lagegunstparameter π_ℓ hinzu, der das relative Gewicht von nahen und fernen Nachbarn in einer infinit verteilen 'lag Struktur' ('zeitliche / räumliche Interdependenz') angibt Für die Schätzung wird eine Maximum Likelihood Prozedur verwendet, die es erlaubt alle Parameter, d. h. die Regressionskoeffizienten, die Koeffizienten in der Heteroskedastizitätsfunktion, die Autokorrelations- une Lagegunstparameter sowie die der Box-Cox Transformationen der unabhängigen, abhängigen und Heteroskedastizitätsvariablen, gemeinsam zu bestimmen.

Stichworte: Dirigierte Autokorrelation; räumliche Autokorrelation; Heteroskedastizität; Box-Cox-Transformation; Verteilte lags; Nachbarschaftsmatrix; R-Koyck-lag; AR-C-D-Verfahren; Elastizität; Residuenstruktur.

Contents

1. INTRODUCTION

In applied regression analysis, perhaps the most important aspect of model specification is the choice of the functional forms of the dependent and independent variables. 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). Special cases of the Box-Cox transformation include the linear and logarithmic forms often encountered in econometric studies.

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 render the transformed dependent variable more nearly homoskedastic. To deal with this problem, models which estimate the flexible functional forms of the variables should include a simultaneous correction of heteroskedasticity. We use here the very general specification of Gaudry and Dagenais (1979).

Another important case is the problem of autocorrelation of the residuals, notably with spatial data which are used in different domains such as agriculture, earth sciences, epidemiology, geography and economics. To describe the correlations among the "nearest" points, a Boolean (0-1) matrix R used as a weight matrix can be specified using some equidistance rules which can be based on "natural" (e.g. time, distance) or "directed" (e.g. socioeconomic variables such as income, religion, population) factors, in the sense defined by Gaudry and Blum (1988). Here we follow a generalization of this approach, described in Blum, Bolduc and Gaudry (1990; 1995/1996), whereby an R-Koyck process which is an analog of the usual Koyck transformation on a time series, but extended to the weight matrix R, allows to estimate the relative impact of close and remote neighbors in an infinite distributed lag structure, by using a new parameter, called proximity parameter, in addition to the usual autocorrelation parameter ρ for each autocorrelation order.

The L-2.1 program allows the maximum likelihood estimation of the parameters associated with the functional forms of the dependent and independent variables, as well as those specified in the heteroskedasticity function and the autoregressive structure of the residuals. Gaudry, Mandel and Rothengatter (1994) use this algorithm to estimate their various models with different data sets and R matrix construction rules, namely with "natural", "directed" and mixed orderings. Blum and Gaudry (1990) use only different "directed" orderings in their demonstration of the importance of functional form and heteroskedasticity in obtaining expected regression signs.

2. STATISTICAL MODEL

2.1 Model

For a sample of N observations, the regression equation with Box-Cox transformations (BCT) on the dependent and independent variables, Y_t and X_{kt}'s, and with autoregressive and heteroskedastic errors is as follows:

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

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

$$v_t = \sum_{\ell=1}^{2} \rho_\ell \left(\sum_{n=1}^{N} \tilde{r}_{\ell,tn} v_n \right) + w_t \quad , \quad (-1 < \rho_\ell < 1) \quad . \tag{3}$$

The vector of residuals $u = \{u_t\}$ is assumed to be heteroskedastic with mean $E(u) = 0$ and covariance matrix $E(uu') = \Omega = diag(\omega_{11}, \ldots, \omega_{NN})$, where $\omega_{tt} = E(u_t^2) = f(Z_t)E(v_t^2)$ and $f(Z_t)$ is a positive function of a vector of variables, $Z_t = (Z_{1t}, \ldots, Z_{Mt})$, 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 be distinct.

Due to computational costs, the vector of residuals $v = \{v_t\}$ is assumed to follow a first or a second order autoregressive process with mean $E(v) = 0$ and covariance matrix $E(vv') = \sigma_w^2 (P'P)^{-1}$, where $P = I - \sum_{\ell=1}^{2} \rho_\ell \widetilde{R}_\ell$ and $\widetilde{R}_\ell$ whose typical element is $\tilde{r}_{\ell,tn}$ is a spatial analog of the usual Koyck distributed lag structure in time series, proposed by Blum, Bolduc and Gaudry (1990; 1995/1996).

The vector of residuals $w = \{w_t\}$ is assumed to have a mean $E(w) = 0$ and a covariance matrix $E(ww') = \sigma_w^2 I$.

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 which is defined as a power transformation with a parameter λ on any positive real variable V_t:

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

A special option in the program allows a nonnegative independent variable called "quasi-dummy" to be transformed by BCT only for the positive observations of the variable but not for the null observations. To compensate for this effect, an associated real dummy which takes a value of 1 if the observation of the quasi-dummy is positive and a value of 0 otherwise, is automatically introduced into the regression equation.

In (2) the functional form of heteroskedasticity for $f(Z_t)$ is assumed to be, as proposed by Gaudry and Dagenais (1979):

$$f(Z_t) = \exp\left[\sum_{m=1}^{M} \delta_m Z_{mt}^{(\lambda_{zm})}\right] = \prod_{m=1}^{M} \exp\left[\delta_m Z_{mt}^{(\lambda_{zm})}\right] \quad . \tag{5}$$

This form is commonly called "multiplicative" heteroskedasticity since it can also be expressed as a product of M exponential functions. Hence, the logarithm of the variance of u_t is a linear combination of $Z_{mt}^{(\lambda_{zm})}$'s:

$$\ln E\big(u_t^2\big) = \ln\big[f(Z_t)E\big(v_t^2\big)\big] = \sum_{m=1}^{M} \delta_m Z_{mt}^{(\lambda_{zm})} + \ln E\big(v_t^2\big) \tag{6}$$

In (3) the elements $\tilde{r}_{\ell,tn}$'s of $\widetilde{R}_\ell$ can be explicitly defined by using a matrix form of (3):

$$v = \sum_{\ell=1}^{2} \rho_\ell \widetilde{R}_\ell v + w \quad . \tag{7}$$

Each matrix $\widetilde{R}_\ell$ is a Koyck distributed lag applied to a matrix $\bar{R}_\ell$:

$$\widetilde{R}_\ell = \pi_\ell\big[I - (1-\pi_\ell)\bar{R}_\ell\big]^{-1}\bar{R}_\ell\,, \quad (0 < \pi_\ell \leq 1) \tag{8}$$

where $\bar{R}_\ell$ is a row normalized matrix of a Boolean matrix R_ℓ whose typical element $r_{\ell,tn}$ is equal to one if there is a correlation between two residuals v_t and v_n $(t \neq n)$, and equal to zero otherwise, and π_ℓ is a new parameter to be defined in the next section.

In spatial econometrics, the matrix R_ℓ is called a contiguity matrix since the observation unit being usually a zone, we want to specify a correlation between a residual of a given zone and each of the residuals of the contiguous zones or neighbors, according to some specific equidistance rules which can be based on "natural" (e.g. time, distance) or "directed" (e.g. socioeconomic variables such as income, religion) factors. Mixed orders are possible: for example, the first order matrix R_1 can be based on distance, whereas the second order matrix R_2 on income.

2.2 Near and Distant Effects

Following the Blum-Bolduc-Gaudry approach, the Koyck process applied to a row normalized matrix $\bar{R}_\ell$ represents a sum of increasing powers of $\bar{R}_\ell$ weighted by an infinite normalized series of geometrically declining nonnegative weights $a_{\ell c}$, $(c = 1, \ldots, \infty)$:

1. The powers of $\bar{R}_\ell$ generate a sequence of contiguity matrices $\left(\bar{R}_\ell, \bar{R}_\ell{}^2, \ldots, \bar{R}_\ell{}^c\right)$ allowing backward and forward linkages arising through the first degree contiguity matrix $\bar{R}_\ell$.
   For example, the ith diagonal element of $\bar{R}_\ell{}^2$ can be written as:

$$\bar{r}_{\ell,ii}^{\,2} = \bar{r}_{\ell,i\cdot}\,\bar{r}_{\ell,\cdot i} = \sum_{t=1}^{N} \bar{r}_{\ell,it}\,\bar{r}_{\ell,ti} = \sum_{t\neq i} \bar{r}_{\ell,it}\,\bar{r}_{\ell,ti} \tag{9}$$

where $\bar{r}_{\ell,i\cdot}$ and $\bar{r}_{\ell,\cdot i}$ denote respectively the ith row and the ith column of the matrix $\bar{R}_\ell$. Each term in the sum above represents the product of two effects, one associated with a forward linkage from zone i to zone t and the other with a backward linkage from zone t to zone i. Note that the sum contains only $N-1$ terms since the omitted term corresponds to the square of the diagonal element $\bar{r}_{\ell,ii}$ which is null by definition.
An off-diagonal element of $\bar{R}_\ell{}^2$ can be expressed as:

$$\bar{r}^{\,2}_{\ell,ij} = \bar{r}_{\ell,i\cdot}\,\bar{r}_{\ell,\cdot j} = \sum_{t=1}^{N} \bar{r}_{\ell,it}\,\bar{r}_{\ell,tj} = \sum_{t\neq i,j} \bar{r}_{\ell,it}\,\bar{r}_{\ell,tj} \tag{10}$$

where $\bar{r}_{\ell,i\cdot}$ and $\bar{r}_{\ell,\cdot j}$ denote respectively the ith row and the jth column of the matrix $\bar{R}_\ell$. Clearly, the off-diagonal element represents the total effect of $N-2$ terms, each of which is a compound effect of two forward linkages, the first from zone i to zone t, and the second from zone t to zone j. In this sense, $\bar{R}_\ell{}^2$ defines a second degree of neighborliness or proximity (neighbors of neighbors).

2. The nonnegative weights $a_{\ell c}$'s in the infinite normalized series are geometrically decreasing:

$$a_{\ell c} = a_\ell^{c-1}(1-a_\ell) \quad , \quad (0 \leq a_\ell < 1) \quad . \tag{11}$$

The normalization ensures that the sum of the weights $a_{\ell c}$ is always equal to unity:

$$\sum_{c=1}^{\infty} a_{\ell c} = \sum_{c=1}^{\infty} a_\ell^{c-1}(1-a_\ell) = \frac{1-a_\ell}{1-a_\ell} = 1 \quad . \tag{12}$$

The Koyck process applied to the contiguity matrix $\bar{R}_\ell$ can then be written as:

$$\begin{aligned}
\widetilde{R}_\ell &= \sum_{c=1}^{\infty} a_{\ell c}\bar{R}_\ell{}^c \\
&= (1-a_\ell)\sum_{c=1}^{\infty} a_\ell^{c-1}\bar{R}_\ell{}^c \\
&= (1-a_\ell)\Big[\bar{R}_\ell + a_\ell\bar{R}_\ell{}^2 + a_\ell^2\bar{R}_\ell{}^3 + \ldots\Big] \\
&= (1-a_\ell)\Big[I + a_\ell\bar{R}_\ell + a_\ell^2\bar{R}_\ell{}^2 + \ldots\Big]\bar{R}_\ell \\
&= (1-a_\ell)\big[I - a_\ell\bar{R}_\ell\big]^{-1}\bar{R}_\ell \quad .
\end{aligned} \tag{13}$$

Note that using the row normalized matrix $\bar{R}_\ell$ ensures that the weighted sum of increasing powers of $\bar{R}_\ell$ will eventually converge to a finite matrix. Moreover it also guarantees a convex likelihood function over the stable unit interval of ρ_ℓ, as Bolduc (1987) has shown.

To obtain a more intuitive measure that increases rather than decreases with the level of proximity, a change of parameter is performed on a_ℓ:

$$\pi_\ell = 1 - a_\ell \quad , \quad (0 < \pi_\ell \leq 1) \tag{14}$$

where the new parameter π_ℓ is called a proximity parameter.

The introduction of π_ℓ for each order ℓ in (8) allows an endogenization of the relative importance of near and distant effects:

1. If π_ℓ is equal to 1: $\widetilde{R}_\ell$ is equal to $\bar{R}_\ell$, indicating that only the adjacent neighbors will have an impact on the correlations between the associated residuals. This is the classical case considered by Ord (1975).
2. If π_ℓ tends towards 0: the limiting form of $\widetilde{R}_\ell$ is a matrix with identical rows provided that the contiguity matrix $\bar{R}_\ell$ be general, i.e. non block diagonal. In this case, the near effect is reduced at its minimum in favor of the distant effect.. Note that if $\bar{R}_\ell$ is a block diagonal matrix, then the limiting form of $\widetilde{R}_\ell$ is a matrix with as many subsets of identical rows as there are diagonal blocks in $\bar{R}_\ell$.

2.3 Likelihood Function

Rewrite the model (1)-(2)-(3) in matrix form:

$$Y^{(\lambda_y)} = X^{(\lambda_x)}\beta + u \tag{15}$$

$$u = Hv \tag{16}$$

$$v = \sum_{\ell=1}^{2} \rho_\ell \widetilde{R}_\ell v + w \quad \Leftrightarrow \quad Pv = w \tag{17}$$

where $H = \mathrm{diag}\left[f(Z_1)^{1/2}, \ldots, f(Z_N)^{1/2}\right]$ and $P = I - \sum_{\ell=1}^{2} \rho_\ell \widetilde{R}_\ell$.

Following (3), under the assumption that the residuals w_t's are normally and independently distributed with zero mean and variance σ_w^2, the likelihood function associated with w is:

$$\mathcal{L}(w) = \left(2\pi\sigma_w^2\right)^{-N/2} \exp\left[-\frac{1}{2\sigma_w^2} w'w\right] \quad . \tag{18}$$

Changing variables successively from w to v, v to u, and u to Y, the likelihood function associated with Y is:

$$\mathcal{L}(Y) = \mathcal{L}(w)\left|\frac{\partial w}{\partial v}\right|\left|\frac{\partial v}{\partial u}\right|\left|\frac{\partial u}{\partial Y}\right| \tag{19}$$

where each Jacobian is defined as the absolute value of the determinant of the matrix of derivatives:

$$|\partial w/\partial v| = |\partial(Pv)/\partial v| = |\det P| \tag{20}$$

$$|\partial v/\partial u| = \left|\partial\left(H^{-1}u\right)/\partial u\right| = \left|\det H^{-1}\right| = \prod_t \left[f(Z_t)\right]^{-1/2} \tag{21}$$

and

$$|\partial u/\partial Y| = \left|\partial\left[Y^{(\lambda_y)} - X^{(\lambda_x)}\beta\right]/\partial Y\right| = \left|\partial Y^{(\lambda_y)}/\partial Y\right| = \prod_t Y_t^{\lambda_y - 1} \tag{22}$$

.

Taking the logarithm of (19), we obtain the log-likelihood function:

$$L = -\frac{N}{2}\ln\left(2\pi\sigma_w^2\right) - \frac{1}{2\sigma_w^2}w'w + \ln\mid\det P\mid - \frac{1}{2}\sum_t \ln f(Z_t) + (\lambda_y - 1)\sum_t \ln Y_t \tag{23}$$

2.4 Concentrated Log-likelihood Function

Premultiply both sides of (15) by PH^{-1}, we get the model in terms of the transformed variables $Y^{\star\star}$ and $X^{\star\star}$:

$$Y^{\star\star} = X^{\star\star}\beta + w \tag{24}$$

where $Y^{\star\star} = PH^{-1}Y^{(\lambda_y)}$ and $X^{\star\star} = PH^{-1}X^{(\lambda_x)}$.

Since this model is just linear in the β-coefficients, we can concentrate the log-likelihood function (23) on β and σ_w^2 by equating the first derivatives of the function with respect to these parameters to zero:

$$\frac{\partial L}{\partial \beta} = -\frac{1}{2\sigma_w^2}\frac{\partial w'w}{\partial \beta} = \frac{1}{\sigma_w^2}\left(X^{\star\star\prime}Y^{\star\star} - X^{\star\star\prime}X^{\star\star}\beta\right) = 0 \tag{25}$$

$$\frac{\partial L}{\partial \sigma_w^2} = -\frac{N}{2\sigma_w^2} + \frac{1}{2\sigma_w^4}w'w = \frac{1}{2\sigma_w^2}\left(-N + \frac{1}{\sigma_w^2}w'w\right) = 0 \tag{26}$$

and solving these equations for β and σ_w^2:

$$\hat{\beta} = \left(X^{\star\star\prime}X^{\star\star}\right)^{-1}X^{\star\star\prime}Y^{\star\star} \tag{27}$$

$$\hat{\sigma}_w^2 = \frac{1}{N}w'w = \frac{1}{N}(Y^{\star\star} - X^{\star\star}\beta)'(Y^{\star\star} - X^{\star\star}\beta) \quad . \tag{28}$$

Replacing β by $\hat{\beta}$ in the expression for $\hat{\sigma}_w^2$ gives another value of σ_w^2:

$$\bar{\sigma}_w^2 = \frac{1}{N}\left(Y^{\star\star} - X^{\star\star}\hat{\beta}\right)'\left(Y^{\star\star} - X^{\star\star}\hat{\beta}\right) \quad . \tag{29}$$

Substitution of $\hat{\beta}$ and $\bar{\sigma}_w^2$ in (23) yields the concentrated log-likelihood function:

$$\bar{L} = -\frac{N}{2}[1 + \ln(2\pi)] - \frac{N}{2}\ln\bar{\sigma}_w^2 + \ln\mid\det P\mid - \frac{1}{2}\sum_t \ln f(Z_t) + (\lambda_y - 1)\sum_t \ln Y_t \tag{30}$$

which now depends only on the parameters $\lambda_y, \lambda_x, \lambda_z, \delta, \rho$ and π.

3. COMPUTATIONAL ASPECTS

3.1 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 $\lambda_y, \lambda_x, \lambda_z, \delta, \rho$ and π.

In computing $\bar{L}$ at each iteration for the general case where two autocorrelation orders are involved and each matrix $\widetilde{R}_\ell$ is not symmetric due to the row normalization of the contiguity matrix R_ℓ which is usually symmetric, the logarithm of the Jacobian of the transformation from w to v, $\ln |\det P|$, cannot be evaluated with an equivalent form analogous to the one obtained in the special case where only one autocorrelation order is considered and the matrix $\widetilde{R}$ is symmetric so that it can always be diagonalized. In this simple case considered for example by Blum, Bolduc and Gaudry (1990; 1995/1996), the equivalent form of $\ln \mid \det P \mid$ can be written as $\sum_t \ln [1 - \rho\gamma_t]$ where the eigenvalues γ_t's of $\widetilde{R}$ can be quickly computed from the relation $\gamma_t = \pi\lambda_t/[1 - (1 - \pi)\lambda_t]$ with the eigenvalues λ_t's of $\bar{R}$ evaluated just once, before the iterations begin. For the general case, we must use an algorithm which inverts the nonsymmetric matrix P while giving as a by-product the determinant of P, since the inverse of P is also needed in the computation of the derivatives of $\ln \mid \det P \mid$ with respect to the ρ and π-parameters.

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

$$\frac{\partial \bar{L}}{\partial \lambda_y} = -\frac{1}{\bar{\sigma}_w^2} \hat{w}' P \frac{\partial \hat{v}}{\partial \lambda_y} + \sum_t \ln Y_t \tag{31}$$

$$\frac{\partial \bar{L}}{\partial \lambda_{xk}} = -\frac{1}{\bar{\sigma}_w^2} \hat{w}' P \frac{\partial \hat{v}}{\partial \lambda_{xk}} \tag{32}$$

$$\frac{\partial \bar{L}}{\partial \lambda_{zm}} = -\frac{1}{\bar{\sigma}_w^2} \hat{w}' P \frac{\partial \hat{v}}{\partial \lambda_{zm}} - \frac{1}{2}\delta_m \sum_t \frac{\partial Z_{mt}^{(\lambda_{zm})}}{\partial \lambda_{zm}} \tag{33}$$

$$\frac{\partial \bar{L}}{\partial \delta_m} = -\frac{1}{\bar{\sigma}_w^2} \hat{w}' P \frac{\partial \hat{v}}{\partial \delta_m} - \frac{1}{2} \sum_t Z_{mt}^{(\lambda_{zm})} \tag{34}$$

$$\frac{\partial \bar{L}}{\partial \rho_\ell} = -\frac{1}{\bar{\sigma}_w^2} \hat{w}' \frac{\partial P}{\partial \rho_\ell} \hat{v} + \frac{\partial \ln \mid \det P \mid}{\partial \rho_\ell} \tag{35}$$

$$\frac{\partial \bar{L}}{\partial \pi_\ell} = -\frac{1}{\bar{\sigma}_w^2} \hat{w}' \frac{\partial P}{\partial \pi_\ell} \hat{v} + \frac{\partial \ln \mid \det P \mid}{\partial \pi_\ell} \tag{36}$$

where

$$\hat{w} = Y^{\star\star} - X^{\star\star}\hat{\beta} \tag{37}$$

$$\hat{v} = H^{-1}\left[Y^{(\lambda_y)} - X^{(\lambda_x)}\hat{\beta}\right] \tag{38}$$

$$\frac{\partial \hat{v}}{\partial \lambda_y} = \left\{ \frac{1}{\sqrt{f(Z_t)}} \frac{\partial Y_t^{(\lambda_y)}}{\partial \lambda_y} \right\} \tag{39}$$

$$\frac{\partial \hat{v}}{\partial \lambda_{xk}} = \left\{ -\frac{\hat{\beta}_k}{\sqrt{f(Z_t)}} \frac{\partial X_{kt}^{(\lambda_{xk})}}{\partial \lambda_{xk}} \right\} \tag{40}$$

$$\frac{\partial \hat{v}}{\partial \lambda_{zm}} = \left\{ -\frac{1}{2} \hat{v}_t \delta_m \frac{\partial Z_{mt}^{(\lambda_{zm})}}{\partial \lambda_{zm}} \right\} \tag{41}$$

$$\frac{\partial \hat{v}}{\partial \delta_m} = \left\{ -\frac{1}{2} \hat{v}_t Z_{mt}^{(\lambda_{zm})} \right\} \tag{42}$$

$$\frac{\partial P}{\partial \rho_\ell} = -\widetilde{R}_\ell \tag{43}$$

$$\frac{\partial \ln \mid \det P \mid}{\partial \rho_\ell} = \mathrm{tr}\left(\frac{\partial P}{\partial \rho_\ell} P^{-1} \right) = -\mathrm{tr}\left(\widetilde{R}_\ell P^{-1} \right) \tag{44}$$

$$\frac{\partial P}{\partial \pi_\ell} = -\rho_\ell \frac{\partial \widetilde{R}_\ell}{\partial \pi_\ell} \tag{45}$$

$$\frac{\partial \widetilde{R}_\ell}{\partial \pi_\ell} = \left[I - \widetilde{R}_\ell \right] \left[I - (1-\pi_\ell) \bar{R}_\ell \right]^{-1} \bar{R}_\ell \tag{46}$$

$$\frac{\partial \ln \mid \det P \mid}{\partial \pi_\ell} = \mathrm{tr}\left(\frac{\partial P}{\partial \pi_\ell} P^{-1} \right) = -\rho_\ell \mathrm{tr}\left(\frac{\partial \widetilde{R}_\ell}{\partial \pi_\ell} P^{-1} \right) \quad . \tag{47}$$

The first derivatives of the BCT's on Y_t, X_{kt} and Z_{mt} with respect to λ_y, λ_{xk} and λ_{zm} in (39)-(40)-(41) respectively are given by the generic formula:

$$\frac{\partial V_t^{(\lambda)}}{\partial \lambda} = \begin{cases} \frac{1}{\lambda} \left[V_t^{\lambda} \ln V_t - V_t^{(\lambda)} \right] & , (\lambda \neq 0) \\ \frac{1}{2} \ln^2 V_t & , (\lambda = 0) \end{cases} \quad . \tag{48}$$

3.2 Asymptotic Covariance Matrix of the Parameter Estimates

At the maximum point of $\bar{L}$, hence L, the asymptotic covariance matrix of all the estimates of the model $\hat{\theta} = \left(\hat{\beta}', \hat{\sigma}_w^2, \hat{\lambda}_y, \hat{\lambda}_x', \hat{\lambda}_z', \hat{\delta}', \hat{\rho}', \hat{\pi}' \right)'$ is evaluated by the method of Berndt et al. (1974) which is based on the cross products of the first derivatives of the log-likelihood function L_t associated with each observation t.

Rewrite the log-likelihood function L as a sum of N individual log-likelihood functions L_t:

$$L = \sum_t L_t = \sum_t \left[-\frac{1}{2} \ln \left(2\pi\sigma_w^2 \right) - \frac{w_t^2}{2\sigma_w^2} + \frac{1}{N} \ln \mid \det P \mid - \frac{1}{2} \ln f(Z_t) + (\lambda_y - 1) \ln Y_t \right] \quad . \tag{49}$$

The covariance matrix of the parameter estimates $\hat{\theta}$ is given by:

$$\mathrm{Var}\left(\hat{\theta} \right) = \left[\sum_t \frac{\partial L_t}{\partial \theta} \frac{\partial L_t}{\partial \theta'} \right]_{\theta = \hat{\theta}}^{-1} \tag{50}$$

where the typical elements of the column vector $\partial L_t/\partial\theta$ are:

$$\frac{\partial L_t}{\partial \beta_k} = -\frac{1}{\sigma_w^2} w_t \frac{\partial w_t}{\partial \beta_k} \tag{51}$$

$$\frac{\partial L_t}{\partial \sigma_w^2} = \frac{1}{2}\left[\left(\frac{w_t}{\sigma_w^2}\right)^2 - \frac{1}{\sigma_w^2}\right] \tag{52}$$

$$\frac{\partial L_t}{\partial \lambda_y} = -\frac{1}{\sigma_w^2} w_t \frac{\partial w_t}{\partial \lambda_y} + \ln Y_t \tag{53}$$

$$\frac{\partial L_t}{\partial \lambda_{xk}} = -\frac{1}{\sigma_w^2} w_t \frac{\partial w_t}{\partial \lambda_{xk}} \tag{54}$$

$$\frac{\partial L_t}{\partial \lambda_{zm}} = -\frac{1}{\sigma_w^2} w_t \frac{\partial w_t}{\partial \lambda_{zm}} - \frac{1}{2}\delta_m \frac{\partial Z_{mt}^{(\lambda_{zm})}}{\partial \lambda_{zm}} \tag{55}$$

$$\frac{\partial L_t}{\partial \delta_m} = -\frac{1}{\sigma_w^2} w_t \frac{\partial w_t}{\partial \delta_m} - \frac{1}{2} Z_{mt}^{(\lambda_{zm})} \tag{56}$$

$$\frac{\partial L_t}{\partial \rho_\ell} = -\frac{1}{\sigma_w^2} w_t \frac{\partial w_t}{\partial \rho_\ell} + \frac{1}{N}\frac{\partial \ln \mid \det P \mid}{\partial \rho_\ell} \tag{57}$$

$$\frac{\partial L_t}{\partial \pi_\ell} = -\frac{1}{\sigma_w^2} w_t \frac{\partial w_t}{\partial \pi_\ell} + \frac{1}{N}\frac{\partial \ln \mid \det P \mid}{\partial \pi_\ell} \quad . \tag{58}$$

In (51) and (53)-(58), the first derivatives of w_t with respect to β_k, λ_y, λ_{xk}, λ_{zm}, δ_m, $\rho_\ell$and π_ℓ are the typical elements of the column vectors of derivatives of w with respect to these parameters:

$$\frac{\partial w}{\partial \beta_k} = -P X_k^{(\lambda_{xk})} \tag{59}$$

$$\frac{\partial w}{\partial \lambda_y} = P\frac{\partial Y^{(\lambda_y)}}{\partial \lambda_y} \tag{60}$$

$$\frac{\partial w}{\partial \lambda_{xk}} = -\beta_k P\frac{\partial X_k^{(\lambda_{xk})}}{\partial \lambda_{xk}} \tag{61}$$

$$\frac{\partial w}{\partial \lambda_{zm}} = P\frac{\partial v}{\partial \lambda_{zm}} \tag{62}$$

$$\frac{\partial w}{\partial \delta_m} = P\frac{\partial v}{\partial \delta_m} \tag{63}$$

$$\frac{\partial w}{\partial \rho_\ell} = \frac{\partial P}{\partial \rho_\ell} v = -\widetilde{R}_\ell v \tag{64}$$

$$\frac{\partial w}{\partial \pi_\ell} = \frac{\partial P}{\partial \pi_\ell} v = -\rho_\ell \frac{\partial \widetilde{R}_\ell}{\partial \pi_\ell} v \quad . \tag{65}$$

3.3 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}'s and Z_{mt}'s, an automatic scaling of these variables is performed during the iterations to avoid numerical problems which can slow down or inhibit the convergence process.

Each variable V_t is scaled as follows:

$$\widetilde{V}_t = g_v V_t \tag{66}$$

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

In general, the β and δ-coefficients and their unconditional t-values based on the covariance matrix $\mathrm{Var}\left(\hat{\theta}\right)$ in (50) are not invariant with respect to the change of units in Y_t, X_{kt}'s and Z_{mt}'s, but the remaining parameters λ_y, λ_x, λ_z, ρ and π as well as their unconditional t-values remain invariant since they are pure numbers.

Note that the concentrated log-likelihood function $\bar{L}$, hence L, and their error variance σ_w^2 are affected by the scaling of Y_t only. Therefore the concentrated log-likelihood values which are listed during the iterations in terms of the scaled units of Y_t would not be the same as if they were in the original units of Y_t, unless the scaling factor of Y_t happens to be equal to one for a particular data set. When a test for stability of the β-coefficients over various data sets is performed in terms of the log-likelihood values or the error variance estimates, these quantities should be in the original units of Y_t, or more generally in a system of units which is common to all data sets used.

3.4 Constraints on the ρ and π-parameters

Although the constraints on the ρ and π-parameters, namely $-1 < \rho_\ell < 1$ and $0 < \pi_\ell \leq 1$, can be explicitly implemented as such, the process of maximizing the concentrated log-likelihood function $\bar{L}$ will work more smoothly if a parameter transformation exists so that the new parameter is not constrained. In other words, the constraints are implicitly incorporated through the parameter change.

The following parameter transformations, namely the Fisher's z-transformation for ρ_ℓ and the Logit transformation for π_ℓ, are used:

$$z_\ell = \frac{1}{2}\ln\frac{1+\rho_\ell}{1-\rho_\ell} \quad \Leftrightarrow \quad \rho_\ell = \tanh z_\ell \quad , \quad (-\infty < z_\ell < +\infty) \tag{67}$$

and

$$\psi_\ell = \ln\frac{\pi_\ell}{1-\pi_\ell} \quad \Leftrightarrow \quad \pi_\ell = \frac{\exp(\psi_\ell)}{1+\exp(\psi_\ell)} \quad , \quad (-\infty < \psi_\ell < +\infty) \quad . \tag{68}$$

Therefore, during the iterations, the maximization of $\bar{L}$ with the DFP algorithm is performed in terms of the z and ψ-parameters. The derivatives of $\bar{L}$ with respect to these new parameters can be written as:

$$\frac{\partial \bar{L}}{\partial z_\ell} = \frac{\partial \bar{L}}{\partial \rho_\ell}\frac{\partial \rho_\ell}{\partial z_\ell} \tag{69}$$

and

$$\frac{\partial \bar{L}}{\partial \psi_\ell} = \frac{\partial \bar{L}}{\partial \pi_\ell}\frac{\partial \pi_\ell}{\partial \psi_\ell} \tag{70}$$

where $\partial\rho_\ell/\partial z_\ell = \ \text{sech}^2 z_\ell = 1 - \tanh^2 z_\ell = 1 - \rho_\ell^2$ and $\partial\pi_\ell/\partial\psi_\ell = \pi_\ell(1-\pi_\ell)$.

Even if the ρ-parameters are numerically constrained on the stable unit interval]-1,1[, sometimes the ρ_ℓ's may tend to go outside the prescribed interval. In this case, the program fails to converge and will stop at one of the bounds beyond which a maximum point of the likelihood function is likely to be found. For example, if both π-parameters are fixed at 1 so that each $\widetilde{R}_\ell$ reduces to $\bar{R}_\ell$ and if both ρ-parameters get more and more identical, the equation of the residual v in (7) can be rewritten as:

$$v = \big(\rho_1\bar{R}_1 + \rho_2\bar{R}_2\big)v + w = \rho^\star\big(\bar{R}_1 + \bar{R}_2\big)v + w \tag{71}$$

where $\rho^\star$ is a parameter representing both ρ_ℓ's in the limit case where they are exactly identical. Since the sum of the row normalized matrices $\bar{R}_1$ and $\bar{R}_2$ gives a new matrix that is not row normalized, there is no guarantee that the likelihood function will remain convex over the unit interval of $\rho^\star$, i.e. the maximum of the function may lie outside the interval.

Another problem concerns the identification of the ρ-parameters when the original contiguity matrices R_1 and R_2, although based on different construction rules, may be very similar:

1. If π_1 is different from π_2, then the Koyck-transformed matrices $\widetilde{R}_1$ and $\widetilde{R}_2$ remain distinct and the ρ-parameters are identified.
2. If the π-parameters become more and more identical, then $\widetilde{R}_1$ and $\widetilde{R}_2$ get more and more similar, which can lead to a nonidentification problem for the ρ_ℓ's.

4. ESTIMATION RESULTS

4.1 Expected Value of the Dependent Variable

Following the Dagenais-Gaudry-Liem (1987) approach, the dependent variable Y_t is assumed to be censored both downwards and upwards, $\varepsilon \leq Y_t \leq \nu$, where ε and ν are respectively the lower and upper censoring points common to all observations in the sample. The generalization of Tobin's (1958) model to a doubly censored dependent variable yields the following expression for the expected value of Y_t conditional on the observed values of the other Y_n's $(n \neq t)$:

$$E(Y_t) = \varepsilon \int_{-\infty}^{w_t(\varepsilon)} \varphi(w)dw + \int_{w_t(\varepsilon)}^{w_t(\nu)} Y_t(w)\varphi(w)dw + \nu \int_{w_t(\nu)}^{\infty} \varphi(w)dw \quad , \quad (\varepsilon \text{ and } \nu > 0) \tag{72}$$

where

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

$$\varphi(w) = \frac{1}{\sqrt{2\pi\,\sigma_w^2}} \exp\left(-\frac{w^2}{2\sigma_w^2}\right) \quad . \tag{73}$$

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

$$Y_t(w) = \begin{cases} \left\{1 + \lambda_y \left[\sum_k \beta_k X_{kt}^{(\lambda_{xk})} + \frac{\sqrt{f(Z_t)}}{p_{tt}}\left(w - \sum_{n\neq t} p_{tn} v_n\right)\right]\right\}^{1/\lambda_y} & , (\lambda_y \neq 0) \\ \exp\left[\sum_k \beta_k X_{kt}^{(\lambda_{xk})} + \frac{\sqrt{f(Z_t)}}{p_{tt}}\left(w - \sum_{n\neq t} p_{tn} v_n\right)\right] & , (\lambda_y = 0) \end{cases} \tag{74}$$

   where p_{tt} and p_{tn} are the typical diagonal and off-diagonal elements of matrix P, and $v_n = \left(Y_n^{(\lambda_y)} - \sum_k \beta_k X_{kn}^{(\lambda_{xk})}\right) / \sqrt{f(Z_t)}$.

3. and the integration bounds $w_t(\varepsilon)$ and $w_t(\nu)$ are given by the generic formula:

$$w_t(\xi) = \begin{cases} \frac{p_{tt}}{\sqrt{f(Z_t)}}\left(\xi^{(\lambda_y)} - \sum_k \beta_k X_{kt}^{(\lambda_{xk})}\right) + \sum_{n\neq t} p_{tn} v_n & , (\lambda_y \neq 0) \\ \frac{p_{tt}}{\sqrt{f(Z_t)}}\left(\ln \xi - \sum_k \beta_k X_{kt}^{(\lambda_{xk})}\right) + \sum_{n\neq t} p_{tn} v_n & , (\lambda_y = 0) \end{cases} \tag{75}$$

   where ξ can be ε or ν.

To compute $E(Y_t)$ for the whole range of Y_t, $0 < Y_t < \infty$, the limit of $w_t(\varepsilon)$ when ε tends towards zero and the limit of $w_t(\nu)$ when ν tends towards infinity should be taken in (72):

$$\lim_{\varepsilon \to 0} w_t(\varepsilon) = \begin{cases} \frac{p_{tt}}{\sqrt{f(Z_t)}}\left(-\frac{1}{\lambda_y} - \sum_k \beta_k X_{kt}^{(\lambda_{xk})}\right) + \sum_{n \neq t} p_{tn} v_n & ,(\lambda_y > 0) \\ -\infty & ,(\lambda_y = 0) \\ -\infty & ,(\lambda_y < 0) \end{cases} \tag{76}$$

and

$$\lim_{\nu \to \infty} w_t(\nu) = \begin{cases} +\infty & ,(\lambda_y > 0) \\ +\infty & ,(\lambda_y = 0) \\ \frac{p_{tt}}{\sqrt{f(Z_t)}}\left(-\frac{1}{\lambda_y} - \sum_k \beta_k X_{kt}^{(\lambda_{xk})}\right) + \sum_{n \neq t} p_{tn} v_n & ,(\lambda_y < 0) \end{cases} . \tag{77}$$

The expected value of Y_t for $0 < Y_t < \infty$ can then be written as:

$$E(Y_t) = \begin{cases} \int_{\lim_{\varepsilon \to 0} w_t(\varepsilon)}^{\infty} Y_t(w)\varphi(w)dw & ,(\lambda_y > 0) \\ \int_{-\infty}^{\infty} Y_t(w)\varphi(w)dw & ,(\lambda_y = 0) \\ \int_{-\infty}^{\lim_{\nu \to \infty} w_t(\nu)} Y_t(w)\varphi(w)dw + \infty \int_{\lim_{\nu \to \infty} w_t(\nu)}^{\infty} \varphi(w)dw & ,(\lambda_y < 0) \end{cases} . \tag{78}$$

For the first two cases, $\lambda_y > 0$ and $\lambda_y = 0$, $E(Y_t)$ has a finite value, but for the last case $\lambda_y < 0$, the second term does not have a finite value since the integral does not disappear, hence $E(Y_t)$ cannot be computed in this case, unless we select a large finite value of ν, say $\tilde{\nu}$, such that the probability that Y_t exceeds this upper limit $\tilde{\nu}$ would be negligible. Using $\tilde{\nu}$, the expected value of Y_t for $\lambda_y < 0$ can be computed as:

$$E(Y_t) = \int_{-\infty}^{w_t(\tilde{\nu})} Y_t(w)\varphi(w)dw + \tilde{\nu} \int_{w_t(\tilde{\nu})}^{\infty} \varphi(w)dw \quad . \tag{79}$$

By default, the program generates a value $\tilde{\nu} = \max_t (Y_t + 10\sigma_y)$ where σ_y is the sample standard error of Y_t, but one can specify another value if there is a numerical problem in computing $E(Y_t)$.

As a sample measure of the degree of numerical approximation of $E(Y_t)$ for the two cases $\lambda_y > 0$ and $\lambda_y < 0$, the mean probability of Y_t in the sample to be at the lower limit ε if $\lambda_y > 0$, or at the upper limit $\tilde{\nu}$ if $\lambda_y < 0$ is also computed:

$$\overline{\Pr}(Y_t \leq \varepsilon) = \frac{1}{N} \sum_t \int_{-\infty}^{\lim_{\varepsilon \to 0} w_t(\varepsilon)} \varphi(w)dw \tag{80}$$

and

$$\overline{\Pr}(Y_t \geq \tilde{\nu}) = \frac{1}{N} \sum_t \int_{w_t(\tilde{\nu})}^{\infty} \varphi(w) dw \quad . \tag{81}$$

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

4.2 Goodness-of-fit Measures

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

Pseudo-(E)-$\mathbf{R^2}$

$$\text{- unadjusted:} \qquad R_E^2 = 1 - \frac{\sum_t \left[Y_t - E(Y_t)\right]^2}{\sum_t \left(Y_t - \bar{Y}_t\right)^2} \tag{82}$$

$$\text{- adjusted:} \qquad \bar{R}_E^2 = 1 - \frac{N-1}{N-K^\star}\left(1 - R_E^2\right) \tag{83}$$

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

Pseudo-(L)-$\mathbf{R^2}$

$$\text{- unadjusted:} \qquad R_L^2 = 1 - \Lambda^{2/N} \tag{84}$$

$$\text{- adjusted:} \qquad \bar{R}_L^2 = 1 - \frac{N-1}{N-K^\star}\left(1 - R_L^2\right) \tag{85}$$

where Λ is the ratio of the likelihood function when maximized with respect to the regression constant β_0 only, to the one with respect to $\theta^\star$ as defined above:

$$\Lambda = \max_{\beta_0} \mathcal{L} \,/\, \max_{\theta^\star} \mathcal{L} \quad . \tag{86}$$

Unlike the R_E^2 measure which can be negative, the R_L^2 measure always remains inside the interval 0–1 since the maximum of $\mathcal{L}$ associated with β_0 — which corresponds to the most

restrictive model where no other independent variable than the constant term is specified, the dependent variable is in linear form and the problems of heteroskedasticity and autocorrelation are not considered — is necessarily smaller than the maximum associated with $\theta^{\star}$ which includes less restricted parameters.

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

$$\Lambda^{2/N} = \frac{\sum\limits_t \left(Y_t^{\star\star} - \hat{Y}_t^{\star\star}\right)^2}{\sum\limits_t \left(Y_t - \bar{Y}\right)^2} \left[\sum_t \frac{Y_t^{\lambda_y - 1}}{\sqrt{f(Z_t)}}\right]^{-2/N} = \frac{\sum\limits_t \left(Y_t - \hat{Y}_t\right)^2}{\sum\limits_t \left(Y_t - \bar{Y}\right)^2} \tag{87}$$

where $Y_t^{\star\star}$ is a typical element of the column vector $Y^{\star\star}$ defined in (24) and $\hat{Y}_t^{\star\star}$ is the fitted value of $Y_t^{\star\star}$.

4.3 Derivatives and Elasticities of the Dependent Variable

Two types of derivatives, hence the elasticities, of the dependent variable Y_t with respect to an independent variable X_{kt} or a heteroskedasticity variable Z_{mt} are computed in the program:

1. The first type of derivatives and elasticities is defined in terms of the sample value of Y_t which is the analog of the expression of $Y_t(w)$ given in (74), but with w replaced by w_t.
2. The second type of derivatives and elasticities is defined in terms of the expected value of Y_t given in (78). Since the model is nonlinear in Y_t, this type is more relevant than the first type.

When heteroskedasticity is present, two specifications of the variable Z_{mt} can be possible:

1. The variable Z_{mt} specified in the heteroskedasticity function $f(Z_t)$ is also used as an independent variable X_{kt}.
2. It is only specified in $f(Z_t)$ and is not included among the independent variables X_{kt}'s.

By definition, the derivatives and the elasticities can be evaluated for every observation t in the sample. To obtain global measures of the derivatives and the elasticities for the whole estimation sample, they are usually evaluated at the sample means of Y_t, X_{kt}'s and Z_{mt}'s:

1. The derivatives and elasticities based on the sample value of Y_t can be computed at the sample means.
2. But the derivatives and elasticities based on the expected value of Y_t cannot be computed at the sample means due to the special autocorrelation structure and are explicated below. Instead, only the average values of the derivatives and elasticities associated with the N observations in the sample can be computed as global measures.

• Derivatives and elasticities in terms of the sample value of $\mathbf{Y_t}$

From the regression equation (1), the sample value of Y_t can be expressed as:

$$Y_t = \begin{cases} \left[1 + \lambda_y \left(\sum_k \beta_k X_{kt}^{(\lambda_{xk})} + u_t \right) \right]^{1/\lambda_y} & ,(\lambda_y \neq 0) \\ \exp\left(\sum_k \beta_k X_{kt}^{(\lambda_{xk})} + u_t \right) & ,(\lambda_y = 0) \end{cases} \tag{88}$$

where $u_t = \sqrt{f(Z_t)} v_t$ and v_t in (3) can be rewritten as:

$$v_t = \frac{1}{p_{tt}} \left(w_t - \sum_{n \neq t} p_{tn} v_n \right) \quad . \tag{89}$$

Replacing successively in (88) the residual u_t by $\sqrt{f(Z_t)} v_t$ and v_t by its value in (89), the sample value of Y_t has the same form as $Y_t(w)$ in (74), but with w replaced by w_t.

The ***derivative*** of Y_t with respect to an independent variable X_{kt} is:

$$D^s_{X_{kt}} = \frac{\partial Y_t}{\partial X_{kt}} = \beta_k \frac{X_{kt}^{\lambda_{xk}-1}}{Y_t^{\lambda_y - 1}} \quad ,(\forall \lambda_{xk}, \lambda_y) \quad . \tag{90}$$

The ***derivative*** of Y_t with respect to a heteroskedasticity variable Z_{mt} is:

$$D^s_{Z_{mt}} = \frac{\partial Y_t}{\partial Z_{mt}} = \frac{1}{2} \delta_m \frac{Z_{mt}^{\lambda_{zm}-1}}{Y_t^{\lambda_y - 1}} \left[Y_t^{(\lambda_y)} - \sum_k \beta_k X_{kt}^{(\lambda_{xk})} \right] \quad . \tag{91}$$

The ***elasticities*** of Y_t with respect to X_{kt} and Z_{mt} are defined as:

$$E^s_{X_{kt}} = \frac{\partial Y_t}{\partial X_{kt}} \frac{X_{kt}}{Y_t} = \beta_k \frac{X_{kt}^{\lambda_{xk}}}{Y_t^{\lambda_y}} \quad ,(\forall \lambda_{xk}, \lambda_y) \tag{92}$$

and

$$E^s_{Z_{mt}} = \frac{\partial Y_t}{\partial Z_{mt}} \frac{Z_{mt}}{Y_t} = \frac{1}{2} \delta_m \frac{Z_{mt}^{\lambda_{zm}}}{Y_t^{\lambda_y}} \left[Y_t^{(\lambda_y)} - \sum_k \beta_k X_{kt}^{(\lambda_{xk})} \right] \quad . \tag{93}$$

For the special case where Z_{mt} also appears as an independent variable X_{kt}, the total effect for the derivative is given by the sum of (90) and (91), and similarly the total effect for the elasticity is given by the sum of (92) and (93).

Clearly the derivatives and the elasticities of the sample value of Y_t depend only on the t-th observation of Y, X_k's and Z_m's, so that all the formulas (90)-(93) can be computed at the sample means of these variables.

• Derivatives and elasticities in terms of the expected value of $\mathbf{Y_t}$

Using (78) for the expressions of $E(Y_t)$ for the cases $\lambda_y > 0$ and $\lambda_y = 0$, and (79) for the case $\lambda_y < 0$, derivatives and elasticities can be obtained.

The ***derivative*** of $E(Y_t)$ with respect to an independent variable X_{kt} can be written as:

$$D^e_{X_{kt}} = \frac{\partial E(Y_t)}{\partial X_{kt}} = \begin{cases} \int\limits_{w_t^\star}^{\infty} \frac{\partial Y_t(w)}{\partial X_{kt}} \varphi(w) dw & ,(\lambda_y > 0) \\ \int\limits_{-\infty}^{\infty} \frac{\partial Y_t(w)}{\partial X_{kt}} \varphi(w) dw & ,(\lambda_y = 0) \\ \int\limits_{-\infty}^{w_t(\tilde{\nu})} \frac{\partial Y_t(w)}{\partial X_{kt}} \varphi(w) dw & ,(\lambda_y < 0) \end{cases} \tag{94}$$

where $w_t^\star = \lim\limits_{\varepsilon \to 0} w_t(\varepsilon)$ given in (76) for the case $\lambda_y > 0$ and $\partial Y_t(w)/\partial X_{kt} = \beta_k X_{kt}^{\lambda_{xk}-1} / [Y_t(w)]^{\lambda_y - 1}$.

Similarly, the ***derivative*** of $E(Y_t)$ with respect to a heteroskedasticity variable Z_{mt} has the following form:

$$D^e_{Z_{mt}} = \frac{\partial E(Y_t)}{\partial Z_{mt}} = \begin{cases} \int\limits_{w_t^\star}^{\infty} \frac{\partial Y_t(w)}{\partial Z_{mt}} \varphi(w) dw & ,(\lambda_y > 0) \\ \int\limits_{-\infty}^{\infty} \frac{\partial Y_t(w)}{\partial Z_{mt}} \varphi(w) dw & ,(\lambda_y = 0) \\ \int\limits_{-\infty}^{w_t(\tilde{\nu})} \frac{\partial Y_t(w)}{\partial Z_{mt}} \varphi(w) dw & ,(\lambda_y < 0) \end{cases} \tag{95}$$

where $\partial Y_t(w)/\partial Z_{mt} = \frac{1}{2}\delta_m Z_{mt}^{\lambda_{zm}-1} \left[\sqrt{f(Z_t)}/p_{tt}\right] \left(w - \sum\limits_{n \neq t} p_{tn} v_n\right) / [Y_t(w)]^{\lambda_y - 1}$.

The ***elasticity*** of $E(Y_t)$ with respect to X_{kt} and Z_{mt} are computed as:

$$E^e_{X_{kt}} = \frac{\partial E(Y_t)}{\partial X_{kt}} \frac{X_{kt}}{E(Y_t)} \tag{96}$$

and

$$E^e_{Z_{mt}} = \frac{\partial E(Y_t)}{\partial Z_{mt}} \frac{Z_{mt}}{E(Y_t)} \quad . \tag{97}$$

For the special case where Z_{mt} also appears as an independent variable X_{kt}, the total effect for the derivative is given by the sum of (94) and (95) , and similarly, the total effect for the elasticity is given by the sum of (96) and (97).

In (94) and (95) the derivatives $\partial Y_t(w)/\partial X_{kt}$ and $\partial Y_t(w)/\partial Z_{mt}$ are functions of the expression $Y_t(w)$ which itself depends on the sum $\sum\limits_{n \neq t} p_{tn} v_n$ which is defined only with respect to a given observation t but not defined at the sample means of the variables Y, X_k's and Z_m's. Hence the derivatives and elasticities of $E(Y_t)$ with respect to X_{kt} and Z_{mt} cannot be evaluated at the sample means of the variables. Instead the sample averages of the individual derivatives and elasticities over N observations are computed to obtain overall measures.

• Corrected elasticities for a quasi-dummy and a real dummy variable

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. 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. To compensate for the effect that only the positive observations are 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 introduced into the regression equation.

Category 3: Real dummy which has only two values: 0 and a positive constant, for example a dummy associated with a quasi-dummy defined above.

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

The correction is based on the fact that one is interested in the effect on the dependent variable only when an activity or a phenomenon represented by the quasi-dummy or the real dummy really occurs, i.e. only when the observations of the dummy are positive.

Two types of corrections should be distinguished:

1. **Elasticity of the sample value of $\mathbf{Y_t}$:** Since this type of elasticity can be evaluated at the sample means, the following correction formula for the elasticity at the sample means (Dagenais et al., 1987) is used:

$$E_k^{s+} = E_k^s \frac{\bar{X}_k^+}{\bar{X}_k} = E_k^s \frac{\sum\limits_t X_{kt}^+ / N_k^+}{\sum\limits_t X_{kt}/N} = E_k^s \frac{N}{N_k^+} \tag{98}$$

where E_k^s represents the elasticity $E_{X_{kt}}^s$ evaluated at the sample means, $\bar{X}_k$ is the sample mean of a quasi-dummy or a real dummy X_k computed from the total set of N observations

used in estimation and $\bar{X}_k^+$ is the sample mean computed from the subset of N_k^+ positive observations.

2. **Elasticity of the expected value of $\mathbf{Y_t}$:** Since the overall value of this type of elasticity is computed as an average of the elasticities $E^e_{X_{kt}}$'s evaluated from N observations in the estimation sample, the corrected elasticity for the positive observations of a quasi-dummy or a real dummy is just an average of the elasticities computed from the subset of N_k^+ positive observations of the dummy:

$$E_k^{e+} = \frac{1}{N_k^+} \sum_{t, X_{kt}>0} E^e_{X_{kt}} \quad . \tag{99}$$

4.4 Student-t Statistics

Two types of t-statistics are computed in the program:

1. **Unconditional t-statistics**: these statistics are computed with the standard errors based on the covariance matrix of all the parameter estimates $\hat{\theta}$ given in (50). Due to the presence of the BCT's on Y, X_k's and Z_m's, the t-tests associated with the β and δ-coefficients are not invariant with respect to the change of units in the variables, whereas those associated with the remaining parameters λ_y, λ_x, λ_z, ρ and π are.
2. **Conditional t-statistics**: these statistics are computed conditionally on the estimated values of the Box-Cox parameters λ_y, λ_x and λ_z , i.e. using the standard errors based on the covariance matrix of the subset of parameter estimates $\hat{\theta}^\star$ excluding the Box-Cox parameters:

$$\text{Var}\left(\hat{\theta}^\star\right) = \left[\sum_t \frac{\partial L_t}{\partial \theta^\star} \frac{\partial L_t}{\partial \theta^{\star\prime}}\right]^{-1}_{\theta^\star=\hat{\theta}^\star} \tag{100}$$

where $\hat{\theta}^\star = \left(\hat{\beta}', \hat{\sigma}_w^2, \hat{\delta}', \hat{\rho}', \hat{\pi}'\right)'$.

The conditional t-tests associated with the β and δ-coefficients are then invariant and can be used in practice to test for the significance of these coefficients, instead of the unconditional ones which are not reliable.

For all the estimated parameters, the standard null hypothesis $\mathbf{H_0}: \ \hat{\theta} = 0$ or $\hat{\theta}^\star = 0$ is used. For the estimated Box-Cox parameters $\hat{\lambda}_y$, $\hat{\lambda}_x$ and $\hat{\lambda}_z$, it is interesting to test against two different values of the parameter:

$$\mathbf{H_0}: \begin{cases} \hat{\lambda} = 0 & : \text{Logarithmic transformation.} \\ \hat{\lambda} = 1 & : \text{Linear transformation.} \end{cases} \tag{101}$$

Finally, for the estimated proximity parameters $\hat{\pi}$, two important values of the parameter to be tested against are:

$$\mathbf{H_0}: \begin{cases} \hat{\pi} = 0 & : \text{Full impact of the distant neighbors,} \\ & \quad \text{and minimum impact of the nearest neighbors.} \\ \hat{\pi} = 1 & : \text{Full impact of the nearest neighbors only.} \end{cases} \tag{102}$$

Table 1 Eigenvalues of $X'X$, Condition Indexes of X and Proportions of Var(b_k).

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

4.5 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 ($X^{\star\star}$ and $Y^{\star\star}$) 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 $X^{\star\star}$ at the maximum of the log-likelihood function.

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

$$X^{'}X \;=\; \sum_{i=1}^{k} \gamma_i p_i p_{i,}^{'} \tag{103}$$

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

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

1. **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 γ_i's. This number measures the sensitivity of b to changes in $X^{'}X$ or $Y^{'}Y$ in linear systems represented by the normal equations $X^{'}Xb \;=\; X^{'}Y$.

2. **Condition indexes**: $\eta_i = (\gamma_{\max}/\gamma_i)^{1/2}, i = 1, ..., K$. Note that if γ_i is equal to $\gamma_{\min}$, then η_i has a maximum value which corresponds to the condition number of X: $\eta_{\max} = \kappa(X)$.

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

$$\text{Var}(b_k) = \sigma^2 \sum_{j=1}^{K} (p_{kj}^2/\gamma_j) \tag{104}$$

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

$$\Pi_{ik} = (p_{ki}^2/\gamma_i) / \sum_{j=1}^{K} (p_{kj}^2/\gamma_j) \tag{105}$$

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

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

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

5. SIMULATION

A separate utility program which is available only for the Sun, but not for the PC, can be used for simulation with two types of samples:

1. Sample used in the estimation of the model (1)-(2)-(3).
2. Enlarged sample which includes the estimation sample.

5.1 Simulation with the Estimation Sample

A four-step procedure is used to compute the simulated value of the dependent variable Y_t after a change in the independent variables X_{kt}'s in (1):

1. Given the parameter estimates $\hat{\theta}$ based on a sample of N observations, compute the expected value of Y_t $(t = 1, \ldots, N)$, say $E_0(Y_t)$, before a change in the independent variables, using the formulas (78) and (79).
2. Following a change in one or more independent variables X_k's, compute the new expected value of Y_t, say $E_1(Y_t)$, that may be called the FITTED VALUE BASED FORECAST.
3. Compute the change in the expected value of Y_t:
$$\Delta(Y_t) = E_1(Y_t) - E_0(Y_t) \quad . \tag{106}$$
4. Using the pivot method commented on by Laferrière (1994), compute the simulated value of Y_t after the change as:
$$\widetilde{Y}_t = Y_t + \Delta(Y_t) \tag{107}$$
which is based on the observed value of Y_t, and hence may be called the OBSERVED VALUE BASED FORECAST.

5.2 Simulation with an Enlarged Sample

Using an enlarged sample implies that bigger contiguity matrices R_1 and R_2 are created with a changed structure: the new observations outside the estimation sample will introduce new forward and backward linkages among themselves and also with the observations within the estimation sample.

Depending on the size of the enlarged sample, two versions of the simulation program are available:

1. The first version is general since it can be used with any estimated value of the proximity parameter π_ℓ in the range $0 < \pi_\ell \leq 1$, but it involves the inversion of a big matrix in order to obtain the R-Koyck matrix $\widetilde{R}_\ell$ in (8) for each order of correlation ℓ. Therefore this version will be limited to an enlarged sample of 2000 observations.
2. The second version is used for a big or very big sample including more than 2000 observations, that may be possible only if there is no matrix inversion to be made, i.e. only if π_ℓ is fixed at 1, which corresponds to the classical case considered by Ord (1975).

To compute the simulated value of Y_t after a change in the independent variables X_k's, the same four-step procedure given in the case of an estimation sample is applied to the enlarged sample. Note that the same parameter estimates $\hat{\theta}$ based on a small sample of N observations of the previous section are used to compute the expected values $E_0(Y_t)$ and $E_1(Y_t)$ for the observations in the enlarged sample.

6. REFERENCES

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.

Blum, U.C.H. and M.J.I. Gaudry (1990), "The Impact of Social Security Contributions on Savings", *Jahrbuch für Sozialwissenschaft* 41, 2, 217–242.

Blum, U., D. Bolduc and M. Gaudry (1990), "From Correlation to Distributed Contiguities: a Family of AR-C-D Autocorrelation Processes", Centre de Recherche sur les Transports, Publication #734, Université de Montréal, *Jahrbuch für Regionalwissenschaft* 16/17, 99–128. 1995/1996.

Bolduc, D. (1987), "On the Estimation of Models with Generalized SAR(1) Processes on the Residuals of a Regression", Centre de Recherche sur les Transports, Publication #508, Université de Montréal.

Box, G.P. and D.R. Cox (1964), "An Analysis of Transformations", *Journal of the Royal Statistical Society*, Series B, 26, 211–243.

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*, 21B, 6, 443–477.

Fletcher, R. and M.J.D. Powell (1963), "A Rapidly Convergent Descent Method for Minimization", *Computer Journal*, 6, 163–168.

Gaudry, M.J.I. and U. Blum (1988), "An Example of Correlation among Residuals in Directly Ordered Data", *Economic Letters*, 26, 335–340.

Gaudry, M.J.I. and M.G. Dagenais (1979), "Heteroscedasticity and the Use of Box-Cox Transformations", *Economic Letters*, 2, 225–229.

Gaudry, M., B. Mandel and W. Rothengatter (1994), "Introducing Spatial Competition through an Autoregressive Contiguous Distributed (AR-C-D) Process in Intercity Generation-Distribution Models within a Quasi-Direct-Format (QDF)", Centre de Recherche sur les Transports, Publication #971, Université de Montréal.

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.

Laferrière, R. (1994), "A Travel Demand Forecasting Dilemma: To Pivot or Not?", *The Logistics and Transportation Review*, Vol. 30, no. 1, p. 21.

Ord, J.K. (1975), "Estimation Methods for Models of Spatial Interaction", *Journal of the American Statistical Association*, Vol. 70, no. 349, 120–126.

Tobin, J. (1958), "Estimation of Relationships for Limited Dependent Variables", *Econometrica*, 25, 24–36.

Zarembka, P. (1974), “Transformation of Variables in Econometrics”, Ch. 3 in P. Zarembka (ed.), *Frontiers in Econometrics*, Academic Press, New York.