Cahier No. 7814

THE USE OF BOX-COX TRANSFORMATIONS IN REGRESSION MODELS WITH HETEROSKEDASTIC AUTOREGRESSIVE RESIDUALS

By

Marc J.I. Gaudry and Marcel G. Dagenais

This research was partly supported by the Research and development Centre of Transport Canada and by the F.C.A.C. program of the *Ministère de l'Éducation du Québec*.

Université de Montréal
Centre de recherche sur les transports - Publication # 106

Août 1978 - Révisé en novembre 1978

ABSTRACT

The purpose of this paper is to show that direct and inverse power Box-Cox transformations can be useful to characterize the functional form of the heteroskedasticity of residuals both in linear models and in models where the variables themselves are subjected to direct Box-Cox transformations. The formulation of the resulting regression models as maximum likelihood problems is briefly extended to take the presence of multiple autocorrelation simultaneously into account.

CONTENTS

1. INTRODUCTION

The systematic introduction of flexible functional form considerations in the generalized single equation regression model is relatively recent. For expository purposes, three stages can be distinguished.

A. Classical practice

Consider the following model written in the usual notation for each observation as

$$y_t = \sum_k \beta_k X_{k_t} + u_t \quad , \qquad t=1,\dots,n, \tag{1A}$$

or in matrix notation as

$$y = X\beta + u, \tag{1B}$$

where X is a nonstochastic $n \times k$ matrix of rank k, β is a $k \times 1$ vector of parameters and u is the $n \times 1$ vector of disturbances with $E(u) = 0$ and $E(uu') = \Omega = \psi^2 V$ (V positive definite).

Classical practice consists in making *specific* assumptions about the form of the fixed or the stochastic part of the model; little *systematic* attention has been given however to the interdependence between these assumptions. Typically, a linear or log-linear specification of the fixed part is assumed. It is not always clear whether the latter form is used because the model is believed to be truly log-linear or because the logarithmic transformation is expected to make the residuals approximately homoskedastic.

Assumptions about the form of Ω normally distinguish between the diagonal and off-diagonal elements of the matrix. The analysis of heteroskedasticity generally consists of specifying Ω as follows:

$$E(u_t \, u_{t'}) = f(X_{1_t} \ldots X_{k_t})\,\psi^2, \qquad t=t' \quad , \qquad (2A)$$

$$= 0 \qquad , \qquad t \neq t' \quad , \qquad (2B)$$

and is centered on trying particular variants of (2A), usually with a single variable. The most frequently used of these single variable formulations is:

$$E(u_t^2) = \psi^2 \, X_{k_t}^2 \qquad (3)$$

and its presence can be identified with the Goldfeld-Quandt (1965) parametric test. Many variants have been specified by Glejser (1969) and others and studied independently from the form of the fixed part by Goldfeld and Quandt (1972). The analysis of correlation between residuals leads generally to specify:

$$E(u_t u_{t'}) = \psi^2 \quad , \qquad t=t' \quad , \qquad (4A)$$

$$\neq 0 \quad , \qquad t \neq t' \quad ; \qquad (4B)$$

temporal autocorrelation is the most frequent and it is expressed as

$$u_t = \sum_{\ell=1}^{r} \rho_\ell \, u_{t-\ell} + e_t \qquad , \qquad t=1,\ldots,n \quad , \qquad (5)$$

where the autoregressive process is assumed to satisfy standard stationarity assumptions and the e_t are presumed to be identically and independently distributed with zero mean and covariance matrix $\sigma^{**2}I$. The form of the autoregressive structure is often modified independently from the form of the fixed part in actual practice.

B. Systematic analysis of the form of the fixed part

A major departure from "trial and error" form specification in models with positive variables was made possible by the use of monotonic transformations studied successively by Anscombe and Tukey (1954), referred to in Tukey (1957), Box and Tidwell (1962) and Box and Cox (1964). The general specification of the problem becomes

$$y^{(\lambda_y)} = \sum_k \beta_k X_k^{(\lambda_{xk})} + u \quad , \tag{6}$$

where the Box-Cox transformation of any variable z, if z denotes the dependent variable y or the k^{th} non-Bolean independent variable X_k, is defined for the parameter λ as

$$z^{(\lambda)} = \begin{cases} \frac{z^\lambda - 1}{\lambda} \quad , & \lambda \neq 0 \quad , \\ \ln z \quad , & \lambda = 0 \quad . \end{cases} \tag{7A}$$

The inverse of this transformation is clearly

$$z = \begin{cases} (\lambda z^{(\lambda)} + 1)^{1/\lambda} \quad , & \lambda \neq 0 \quad , \\ \exp\,(z^{(\lambda)}) \quad , & \lambda = 0 \end{cases} \tag{7B}$$

The direct power transformation (7A) has typically been used without joint examination of the impact of that transformation on the form of the stochastic part of the model (e.g. Kau and Sirmans (1976), Welland (1976), Heckman and Polachek (1974), Spitzer (1976, 1977)). Zarembka (1968,1974) briefly considered first order autoregressive disturbances but without joint hypothesis testing on autocorrelation structure and the form of the fixed part of the model.

C. Joint specification of the fixed and stochastic parts

The first systematic Box-Cox analysis of the structure of the stochastic part of a model simultaneously with that of the functional form of its fixed part was apparently made by Savin and White (1978) for first order autocorrelation and extended to multiple autocorrelation by Gaudry and Wills (1977). Little attention has been paid to the problem of the joint determination of heteroskedasticity and functional form of the fixed part despite the knowledge that Box-Cox parameters appear to compensate for heteroskedasticity and may be more sensitive to its presence than to the presence of autocorrelation. Zarembka (1971) has devised an approximate measure of the sensitivity of functional form estimates to the eventual presence of a particular form of heteroskedasticity.

In this paper, we intend to formulate a fairly general specification of the form of heteroskedasticity and to show that it can be analyzed jointly with the form of the fixed part of the model. We will also extend the analysis to incorporate multiple autocorrelation structures of the residuals.

2. THE FORM OF HETEROSKEDASTICITY IN A LINEAR MODEL

A. A general form for heteroskedastic disturbances

Assume that the covariance matrix of the u's of a linear model is given by

$$E(uu') = \psi^2 \begin{bmatrix} f(X_1) & 0 & \dots & 0 & \dots & 0 \\ 0 & \ddots & & & & \vdots \\ \vdots & & & f(X_t) & & 0 \\ \vdots & & & & \ddots & \vdots \\ 0 & \dots & & 0 & \dots & f(X_n) \end{bmatrix} = \Omega \tag{8}$$

where

$$f(X_t) = \{\lambda_u [\delta_o + \sum_k \delta_k X_{k_t}^{(\lambda_{ek})}] + 1\}^{1/\lambda_u} \tag{9}$$

in which direct Box-Cox transformations are applied to the X_k explanatory variables of the error variance and an inverse power transformation is applied to the contents of the square parentheses. The constants δ_o and ψ are necessary to preserve (Schlesselman, 1971) the invariance of the transformation to units of measurement of the X's and y. Equation (9) contains a great number of special cases of traditional interest.[@]

Setting $\lambda_u = 1$ yields

$$\tilde{\tilde{\omega}}_{tt} = \tilde{\tilde{\delta}}_o + \sum_k \tilde{\tilde{\delta}}_k X_{k_t}^{(\lambda_{ek})}, \text{ where } \tilde{\tilde{\delta}}_o = (\delta_o + 1)\,\psi^2,\ \tilde{\tilde{\delta}}_k = \delta_k\,\psi^2, \tag{10}$$

which includes the linear form $(\lambda_{ek} = 1 \text{ for all } k)$ and many nonlinear forms (square root if the $\lambda_{ek} = 1/2$ and square power if they equal 2) as special cases.

Another family of special cases is obtained by setting $\lambda_u = 0$. By (7B) one obtains

$$\tilde{\omega}_{tt} = \psi^2 \exp(\delta_o + \sum_k \delta_k X_{k_t}^{(\lambda_{ek})}) \equiv \sigma^2[\exp(\sum_k \delta_k X_{k_t}^{(\lambda_{ek})})], \text{where } \sigma^2 = \psi^2 e^{\delta_o}, \tag{11}$$

[@]Note that a still more general specification of Equation (9) would use, instead of $(X_1,\dots,X_k)$, a set of exogeneous variables $(Z_1,\dots,Z_{k'})$ where the $Z_{k'}$ may differ from the X_k of the fixed part.

which includes the multivariate multiplicative model $(\lambda_{ek}=0$ for all k)

$$\hat{\hat{\omega}}_{tt} = \sigma^2 \prod_k X_{k_t}^{\delta_k} \qquad (12)$$

found in Dagum and Dagum (1974) and the classical multivariate multiplicative form $(\delta_k = 2$ for all k):

$$\hat{\omega}_{tt} = \sigma^2 \prod_k X_{k_t}^2 \quad . \qquad (13)$$

Park's (1966) specification is obtained by setting all δ_k but one equal to zero in (11) and by setting the remaining $\lambda_{ek} = 0$. Imposing the further constraint that the retained $\delta_k = 2$ yields the classical form (3).

The subclass defined by (11) is more useful than the subclass defined by (10) because the latter can produce negative elements in Ω. Form (11) also leaves considerable room for interactions among the variables most likely to explain the error variance, typically a small subset of the k explanatory variables. For these reasons, and in order to simplify the presentation, we shall use form (11) in the derivation of the likelihood function. A similar derivation with form (9) would be straightforward.

B. The likelihood function for a linear model

If, in model (1), u is assumed to be distributed normally with mean vector 0 and covariance matrix (8) specified as in (11), the likelihood function can be written as

$$L = \left(\frac{1}{\sqrt{2\pi}}\right)^n |\tilde{\Omega}|^{-1/2} \exp\left\{-\frac{1}{2}(y-X\beta)'\tilde{\Omega}^{-1}(y-X\beta)\right\} , \qquad (14)$$

or as

$$L = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma \sqrt{\exp(\sum_k \delta_k X_{k_t}^{(\lambda_{ek})})}} \exp\left[- \frac{(y_t - \sum_k \beta_k X_{k_t})^2}{2\,\sigma^2 \exp(\sum_k \delta_k X_{k_t}^{(\lambda_{ek})})} \right], \qquad (15)$$

from which one estimates parameters β_k, σ, δ_k and λ_{ek}.

In terms of testing for heteroskedasticity, one approach is to obtain asymptotic variances and covariances for these parameters from the matrix of second partial derivatives of the likelihood function and to perform asymptotic tests based on the normal distribution. Another approach is to apply the likelihood ratio test: if for instance L_1 denotes the value of (15) at the maximum and L_0 its value under the assumption that $f(X_t)$ has a particular form, say (12) or (13), the test statistic $-2\,\ell n(L_0/L_1)$, which has a χ^2 distribution with a number of degrees of freedom given by the additional constraints in (12) or (13), can be used to consider either of these null hypotheses. It is therefore possible to deal analytically with the form of heteroskedasticity in a linear regression model and to consider a large number of structures by specifying matrix (8) as in (9) or in more restricted ways.

3. THE FORMULATION IN A NONLINEAR MODEL

If the general Box-Cox form (6) replaces the linear form (1) but the error covariance matrix (8) is still specified as in (11), the likelihood function is now

$$L=\left(\frac{1}{\sqrt{2\pi}}\right)^n |\tilde{\Omega}|^{-1/2} \exp\left\{-\frac{1}{2}(y^{(\lambda_y)} - X^{(\lambda_x)}\beta)^{1}\,\tilde{\Omega}^{-1}\,(y^{(\lambda_y)} - X^{(\lambda_x)}\beta)\right\} |J(\lambda_y;y)|\,, \qquad (16)$$

where $|J(\lambda_y;y)|$ is the Jacobian of the transformation from $y_t^{(\lambda_y)}$ to the actually observed y_t , or

$$|J(\lambda_y;y)| = |\det \frac{\partial y_t^{(\lambda_y)}}{\partial y_t}| = \prod_{t=1}^{n} y_t^{\lambda_y -1} \quad . \tag{17}$$

Equation (16) can be rewritten as

$$L = \prod_{t=1}^{n} \frac{y_t^{\lambda_y -1}}{\sqrt{2\pi}\,\sigma \sqrt{\exp(\sum_k \delta_k X_{k_t}^{(\lambda_{ek})})}} \exp.\left[- \frac{(y_t^{(\lambda_y)} - \sum_k \beta_k X_{k_t}^{(\lambda_{xk})})^2}{2\sigma^2 \cdot \exp(\sum_k \delta_k X_{k_t}^{(\lambda_{ek})})} \right] \tag{18}$$

which adds the parameters $\lambda_y, \lambda_{x1}, \ldots, \lambda_{xk}$ to those already present in the linear model and allows for joint tests of the form of heteroskedasticity with the form of the fixed part of the model. In practice one would probably constrain the λ parameters of the fixed part and those of the stochastic part in order to decrease the number of parameters to be estimated.

4. EXTENSION TO CONSIDER MULTIPLE AUTOCORRELATION SIMULTANEOUSLY

A. The complete formulation in the presence of first order autocorrelation

It is straightforward to extend the previous result to take into account autocorrelation simultaneously with heteroskedasticity in the nonlinear model. Reformulate the previous problem as

$$y_t^{(\lambda_y)} = \sum_k \beta_k X_{k_t}^{(\lambda_{xk})} + u_t \quad , \tag{20}$$

$$u_t = f(X_t)^{1/2} v_t \quad , \tag{21}$$

$$v_t = \rho_1 v_{t-1} + s_t \quad , \tag{22}$$

where $E(v_t^2) = \psi^2$ and s_t is normally and independently distributed with mean zero and variance $\psi^{*2} = \psi^2 (1-\rho_1^2)$.

It is well known that, if the series are stationary, the vector s can be expressed in the form

$$s = Qv \tag{23}$$

where the $n \times n$ transformation matrix Q is defined as

$$Q = \begin{bmatrix} 1-\rho_1{}^2 & 0 & \ldots & 0 \\ -\rho_1 & 1 & \ldots & 0 \\ 0 & . & \ldots & 0 \\ 0 & 0 & \ldots -\rho_1 & 1 \end{bmatrix} \tag{24}$$

and s_1 therefore corresponds to $v_1 \sqrt{1-\rho_1{}^2}$.

Hence one may write $u = Hv = HQ^{-1}s$ and $E(u\,u') = HQ^{-1}(Q^{-1})'H'\psi^{*2} = \Omega^*$ where $H = \text{diag}\,[f(X_t)^{1/2}]$. In the particular case where, as in (9), $\lambda_u = 0$, the combination of the autoregressive and heteroskedastic specifications exposed in (21) and (22) transforms (16) into

$$L = \left(\frac{1}{\sqrt{2\pi}\,\sigma^*}\right)^n |\tilde{H}|^{-\frac{1}{2}} |Q| \exp\left\{-\frac{1}{2}(y^{(\lambda_y)} - X^{(\lambda_x)}\beta)\tilde{\Omega}^{*-1}(y^{(\lambda_y)} - X^{(\lambda_x)}\beta)\right\} |J(\lambda_y;y)|, \tag{25}$$

where $\tilde{\Omega}^* = \tilde{H}Q^{-1}(Q^{-1})'\tilde{H}'\sigma^{*2}$, $\sigma^{*2} = \sigma^2(1-\rho_1{}^2)$ and $\tilde{H} = \text{diag}\left[\exp(\sum_k \delta_k X_{k_t}^{(\lambda_{ek})}/2)\right]$.

Since $|Q| = (1-\rho_1{}^2)^{1/2}$, (25) can also be expressed as

$$L = (1-\rho_1^{\,2})^{1/2} \frac{y_1^{\lambda_y - 1}}{\sqrt{2\pi}\,\sigma^*\sqrt{\exp(\sum_k \delta_k X_{k_1}^{(\lambda_{ek})})}} \exp\left[-\frac{(1-\rho_1)^2 (y_1^{(\lambda_y)} - \sum_k \beta_k X_{k_1}^{(\lambda_{xk})})^2}{2\sigma^{*2} \exp(\sum_k \delta_k X_{k_1}^{(\lambda_{ek})})}\right] \cdot$$

$$\prod_{t=2}^{n} \frac{y_t^{\lambda_y - 1}}{\sqrt{2\pi}\,\sigma^*\sqrt{\exp(\sum_k \delta_k X_{k_t}^{(\lambda_{ek})})}} \exp\left[-\frac{1}{2\sigma^{*2}}\left[\frac{(y_t^{(\lambda_y)} - \sum_k \beta_k X_{k_t}^{(\lambda_{xk})})}{\sqrt{\exp(\sum_k \delta_k X_{k_t}^{(\lambda_{ek})})}} - \rho_1 \frac{(y_{t-1}^{(\lambda_y)} - \sum_k \beta_k X_{k_{t-1}}^{(\lambda_{xk})})}{\sqrt{\exp(\sum_k \delta_k X_{k_{t-1}}^{(\lambda_{xk})})}}\right]^2\right] \tag{26}$$

which is analogous to (18) except that the terms which refer to the first observation have been distinguished from the others.

B. A simplified formulation in the presence of multiple autocorrelation

One may wish to avoid the added complication presented by the first observation. Disregarding the first observation makes no difference asymptotically but may be inefficient in some problems as Beach and MacKinnon (1978) have shown for a linear homoskedastic model. Neglecting the first observation means that the first row of Q in (24) is ignored and (26) simplifies to

$$L = \prod_{t=2}^{n} \frac{(y_t^{\lambda_y - 1})^{\sim}}{\sqrt{2\pi}\,\sigma^*} \exp\left[-\frac{1}{2\sigma^{*2}} (y_t^{(\lambda_y)\tilde{*}} - \sum_k \beta_k X_{k_t}^{(\lambda_{xk})\tilde{*}})^2\right] \tag{27}$$

with

$$z_{\tilde{t}} = \frac{z_t}{\sqrt{\exp(\sum_k \delta_k X_{k_t}^{(\lambda_{ek})})}} \text{ and } z_{t}^{\tilde{*}} = \frac{z_t}{\sqrt{\exp(\sum_k \delta_k X_{k_t}^{(\lambda_{ek})})}} - \rho_1 \frac{z_{t-1}}{\sqrt{\exp(\sum_k \delta_k X_{k_{t-1}}^{(\lambda_{ek})})}} .$$

In the presence of the multiple stationary autoregressive scheme (5), this simplification can, as in Gaudry and Wills (1977), be extended to higher orders of autocorrelation in order to avoid the considerable computational burden which a full formulation of the likelihood function would imply. This extension of the simplified likelihood simply requires dropping as many observations as the highest order of autocorrelation requires. Expressions (27) and (28) become respectively

$$L = \prod_{t=1+r}^{n} \frac{(y_t^{\lambda} y^{-1})^{\sim}}{\sqrt{2\pi}\,\sigma^{**}} \exp\left\{ -\frac{1}{2\sigma^{**2}} \left(y_t^{(\lambda_y)} - \sum_k \beta_k X_{k_t}^{(\lambda_{xk})} \right)^{\widetilde{**}} \right\} , \tag{29}$$

with $z_t^{\sim}$ defined as before, $\sigma^{**} > 0$ and

$$z_t^{\widetilde{**}} = \frac{z_t}{\sqrt{\exp(\sum_k \delta_k X_{k_t}^{(\lambda_{ek})})}} - \sum_{\ell=1}^{n} \rho_\ell \frac{z_{t-\ell}}{\sqrt{\exp(\sum_k \delta_k X_{k_{t-\ell}}^{(\lambda_{ek})})}} . \tag{30}$$

Expression (29) includes autoregressive parameters $\rho_1 \cdots \rho_\ell \cdots \rho_r$ in addition to the parameters contained in (18). It permits the functional form of the fixed part of a model to be determined simultaneously with the structure or form of both autocorrelation and heteroskedasticity.

5. CONCLUSION

The growing practice in the application of Box-Cox transformations in fields such as production theory, monetary economics and transportation modeling should eventually incorporate a procedure such as that developed in this paper which prevents from confounding problems of functional form and of heteroskedasticity and which can easily be extented to models with autoregressive residuals. Experiments are presently under way to compute the maximum likelihood solution of equation (29) for problems involving large numbers of parameters using Powell's (1964) algorithm for maximizing non linear functions without computing derivatives. The routine takes account of the fact that, for given values of the λ, ρ, δ and σ^{**}, the β which maximize $\ell n\ L$ are OLS estimates; similarly for given values of the λ, δ, β and σ^{**}, the ρ which maximize $\ell n\ L$ are also OLS estimates. The program should also evaluate the asymptotic co-variance matrix of the parameter estimates.

ACKNOWLEDGEMENTS

The Research and Development Centre of Transport Canada and the F.C.A.C. program of the *Ministère de l'Éducation du Québec* have supported the research on which this article is based. The authors thank Tran Con Liem for comments on an earlier version.

BIBLIOGRAPHY

Anscombe, F.J. and Tukey, J.W. (1954) The criticism of transformations. Paper presented before the American Statistical Association and Biometric Society, Montreal.

Beach, C.M. and MacKinnon, J.G. (1978) A maximum likelihood procedure for regression with autocorrelated errors. Econometrica 46, 61-58.

Box, G.E.P. and Cox, D.R. (1964) An analysis of transformations. Journal of the Royal Statistical Society Series B, 211-243.

Box, G.E.P. and Tidwell, P.W. (1962) Transformation of the independent variables. Technometrica 4, 531-550.

Dagum, C. et Dagum, E.B. (1974) Construction de modèles et analyse économétrique. Cahiers de l'Institut de sciences mathématiques et économiques appliquées 8. N°. 11-12.

Gaudry, M.I.J. and Wills, M.J. (1977) Estimating the functional form of travel demand models. Publication # 63. Centre de recherche sur les transports. Université de Montréal. Forthcoming in Transportation Research.

Glejser, H. (1969) A new test for heteroskedasticity. Journal of the American Statistical Association 64, 316-323.

Goldfeld, S.M. and Quandt, R.E. (1965) Some tests for homoskedasticity. Journal of the American Statistical Association 60, 539-47.

Goldfeld, S.M. and Quandt, R.E.(1965) Nonlinear Methods in Econometrics. North-Holland.

Heckman, J. and Polachek, S. (1974) Empirical evidence on the functional form of the earnings-schooling relationship. Journal of the American Statistical Association 69, 350-354.

Kau, J.B. and Sirmans, C.F. (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.

Park, R.E. (1966) Estimation with heteroskedastic error terms. Econometrica 34, 888.

Powell, M.J.D. (1964) An efficient method for finding the minimum of a function of several variables without calculating derivatives. Computer Journal 7, 155-166.

Savin, N.E. and White, K.J. (1978) Estimation and testing for functional form and autocorrelation: a simultaneous approach. Journal of Econometrics 8, 1-12.

Schlesselmann, 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.

Tukey, J.W. (1957) On the comparative anatomy of transformations. Annals of Mathematical Statistics 28, 602-632.

Welland, J.D. (1976) Cognitive abilities, schooling and earnings: the question of functional form. Working paper 76-14. Department of Economics. McMaster University. Hamilton.

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. Ch. 3 of Frontiers in Econometrics, ed. by P. Zarembka.