TRIO Model Types, Version 2.0

by

Marc Gaudry, Marcel Dagenais
Richard Laferrière and Tran Liem

with

the collaboration of
Ulrich Blum, Sergio Jara-Diaz, Benedikt Mandel,
Juan de Dios Ortuzar, Werner Rothengatter
and Michael Wills

Agora Jules Dupuit, Département de sciences économiques, Université de Montréal, C.P. 6128, Succursale Centre-ville, Montréal, Canada H3C 3J7.

This paper is based on work financed in part by Transport Canada, and which benefitted directly or indirectly from research funded by the National Sciences and Engineering Research Council and by the Social Sciences and Humanities Research Council of Canada, as well as by the Actions Structurantes and FCAR programs of the Quebec Government. Lucie L’Heureux helped decisively in the coordination of all tasks required for the production of this, as well as other, TRIO related documents.

Agora Jules Dupuit — Publication #104
Centre de recherche sur les transports — Publication #904

April 1993, June 2005
Abstract

This manual is part of the TRIO regression program documentation listed on the last page of this text. It contains three major sections. The first presents the theoretical regression extensions used in TRIO. The second demonstrates the properties, and notably the usefulness, of these methods with selected applications. The third describes in detail all available model types and specifications, as well as the corresponding estimation techniques, that are programmed in TRIO.

Key-words: regression analysis, Box-Cox transformations, logit model, dogit model, inverse power transformation logit model, autocorrelation, heteroskedasticity, SACOD index, transportation demand, mode choice analysis, Montreal, Santiago, Canada, Germany.

Résumé

Ce manuel fait partie de la documentation du progiciel d’analyse de régression TRIO dont on trouvera la liste en dernière page de ce texte, qui comprend trois parties importantes. Dans la première, on présente les caractéristiques spécifiques des prolongements des méthodes de régression incluses dans TRIO. Dans la seconde, on examine les propriétés de ces méthodes, et notamment leur utilité, par des applications choisies. La troisième décrit en détails tous les genres de modèles et de formulations, ainsi que les méthodes d’estimation correspondantes, qui sont programmés dans TRIO.

Mots-clés : analyse de régression, transformations Box-Cox, modèle logit, modèle dogit, modèle transformation puissance inverse appliqué au logit, autocorrélation, hétéroscédasticité, indice SACOD, demande de transport, choix du mode de transport, Montréal, Santiago, Canada, Allemagne.
## CONTENTS

### I - Regression Theory and Methods Used in TRIO, Version 2.0

1. FUNCTIONAL FORM: properties of the Box-Cox transformation

2. FUNCTIONAL FORM: generalizations of the Logit model

3. STOCHASTIC SPECIFICATION

### II - Properties and Selected Applications of Model Types Available in TRIO, Version 2.0

1. LEVEL models

2. SHARE models

3. PROBABILITY models

### III - Feasible Model Types and Estimation Techniques Programmed in TRIO, Version 2.0

1. LEVEL models

2. SHARE models
   - [14] Liem, T.C., Gaudry, M. and R. Laferrière, "SHARE: The S-1 to S-5 Programs for the Standard and Generalized BOX-COX LOGIT and DOGIT and for the Linear and Box-Tukey INVERSE POWER TRANSFORMATION-LOGIT Models with Aggregate Data".

3. PROBABILITY models

---

OTHER TRIO DOCUMENTATION .................................................. See last page
FOREWORD – THE COSTS AND RISKS OF COMPLEXITY

From explanatory data analysis to simple regression with ROOT models

TRIO makes it possible for a user to perform exploratory data analysis as a preliminary to formal multivariate regression analysis. As a part of this step, the user can examine various descriptive statistics pertaining to the variables. But it is also possible to carry out a graphical analysis based in particular on scatter diagrams. These diagrams allow more than a visual inspection of variables by themselves or in pairs: if pairs of variables are shown on screen, the user can ask to see on the same screen, interactively and without building a formal model (that would be done with the variant editor), the pairwise correlation between the variables, the intercept and slope of regressing one variable against the other, and the actual regression line with its standard error. These computations can be performed interactively, in effect as a part of the easy production of graphs, on the complete set of observations available, or for subsets of observations: for instance, the user might choose to examine the graph and to compute the desired values only if one or more other variables have certain values. In that case, the user carries out conditional correlation and conditional univariate regression analyses on the pair of variables shown on screen.

Normally, the next step would involve formal model building with a simple specification: we recommend that the user start with elementary forms, which we call ROOTS. The current ROOT model types are Ordinary Least Squares (OLS) and the Linear Logit (LIN-LOGIT). These procedures are well known and guaranteed to yield a unique maximum of the measure of adjustment used (the value of the logarithm of the likelihood of observing the sample) – intuitively, to maximize the possible “fit” that can be achieved for the given ROOT specification independently of the number of explanatory variables specified in these ROOT forms.

From ROOT models to progressively more complex extensions

As soon as one builds progressively more complex extensions that contain the ROOT as a special nested special case obtained by setting the functional form or stochastic specification additional parameters to special values, then the guarantee of unimodality of the criterion function is, as a rule, lost. This means that, as soon as sufficient additional complexity is attempted, the user will be dealing with another order of magnitude of complexity. Many of the user friendly features of TRIO are designed to ease this task – for instance with the variant editor and the variant creation and initialization mechanisms that purport both to allow fast modifications of specifications (including numerical criteria) and to ensure that every variant is unique and fully reproducible.

This uncertainty occurs whether the user extends the ROOT model by allowing for flexibility of the functional form (in L-1.4, S-1/S-5 and P-2), or whether corrections of the structure of residuals are attempted (in L-1.4). Without trying to be precise in a textbook sense, we wish to give to the user a feel for the kinds of uncertainties that arise. Some are of a theoretical nature; others depend on the practical use of specifications because such use can reveal helpful constraints that the maximization procedure should obey; others arise because of numerical reasons in contexts that the design of the algorithms has not anticipated in view of the very large numbers of specifications that the allowed extensions make possible. Let us consider them in turn.
Theory: functional form. It is important for the user to remember that direct Box-Cox transformations (BCT) applied to the dependent variable of models are invariant to a scalar transformation of that variable only if a constant term is present among the regressors (Schlesselman, 1971): for this reason, TRIO automatically generates regression constants that cannot be removed by the user. It is to maintain invariance of the BCT to changes in units of measurement that the program also generates “associated dummy variables” when a direct Box-Cox transformation is applied to a continuous variable that contains some null observations (these variables are called quasi-dummy variables in the documentation), but these additional constants can be removed if they cause multicollinearity. In some cases, the user might remember that the BCT is invariant to a power transformation of the transformed variable (Ch. 2), a property that can be extremely useful: for instance, in Ch. 8 an index, called SACOD, is defined and used as a dependent variable, thereby allowing for an efficient use of answers to multiple-question surveys that detects significant interaction among the various indicators of the latent variable.

The powerful classificatory properties of the BCT shown in Ch. 1 are generally based on the scalar invariance property of the direct BCT. Although that article helped to reestablish the value of the forgotten shift parameter present in the original paper by Box and Cox (1964), by showing its empirical importance and renaming the second form of the BCT the Box-Tukey transformation (BTT), one cannot say that the properties of the shift parameter are well known – no doubt because the meaning of shift parameters applied to variables (over and beyond freeing one from natural coordinates) – is unclear. In TRIO, the direct and inverse BTT are applied not to variables but to functions of the logit model, and admits for the shift parameter a very clear interpretation. The user should note, however, that the invariance properties of the inverse BCT and BTT to either scalar or power transformations have not been studied to date except for the former in Chapter 14 and in the reference manual of the PROBABILITY algorithm P-2/P-6 (Publication BETA-9802/CRT-97-57) not included in TRIO Version 2.0.

Moreover, as the unconditional asymptotic t-statistics of the linear coefficients of the regression problem depend upon the units of measurement (Spitzer, 1984; Dagenais and Dufour, 1991), TRIO also computes the t-statistics conditionally upon the estimated values of the Box-Cox transformations: we recommend that, as a rule, the user routinely select those.

In addition to these mathematical properties, the statistical properties of the maximum likelihood estimators are also the objects of current work. For L-1.4, the formal examination of the asymptotic properties is finessed (see Ch. 7) by assuming that the data contain no limit (mass point) observations and by verifying this assumption a posteriori in a computation of the probability that each observation be at the limits. Our algorithm does not use the scaled version of the Box-Cox model in which the scaling factor associated with each variable corresponds to its sample geometric mean (such a formulation can be seen in Ch. 1, Eq. (8)), a procedure that would be full of pitfalls (see Dagenais and Dufour, 1992), unless such a scaling were merely used as a shortcut in an algorithm that would terminate with an “unscaled” iteration to establish the point of convergence. The asymptotic properties of the S-1/S-5 estimators have been proved in Ch. 10 for the LIN-IPT-LOGIT, but not for the other families. The asymptotic properties of the P-2 estimator have not been studied but, as the use of the BCT on explanatory variables of a logit model is reasonably straightforward, we do not expect biases to be revealed in future studies of these properties. The bibliographies of Ch. 13, 14 and 15 contain many standard
references on the statistical properties of estimators.

In model types where the BCT is applied to variables, we recommend that users always start from the ROOT model and first extent it with one BCT on all variables of interest, then with two, and so on, in order to stay in control and ease the tasks of ensuring that the global maximum has been found.

**Theory: stochastic specification.** Currently, L-1.4 allows the user to estimate simultaneously direct BCT applied to one or many variables or groups of variables, as well as multiple order autocorrelation and heteroskedasticity. Simply stated, the model is a straightforward extension of the classical regression model in two ways.

Firstly, we require that a vector of autoregressive (VAR) parameters be used even for residual errors that are clearly autoregressive moving average processes (ARMA), as indicated by the Box-Jenkins analysis provided. The reason for increasing the order of the AR scheme to model truly ARMA residuals is computational convenience. In addition, note the multivariate generalization of the univariate AR case amounts to imposing on all variables of the model the same AR structure. This may seem restrictive to those who wish to apply a distinct AR structure to each variable in order to isolate the white noise residuals or “innovations”, and then apply regression to these residuals in a distinct second step. However, our procedure has the advantage of estimating the regression coefficients and the autoregressive structure simultaneously, a process that may be more robust if many of the variables contain errors of observation (Dagenais, 1993): clearly, whitewashing data series that contain observation errors runs the risk that the residuals of each variable will primarily consist of errors of observation. One would then expect regression made on such “innovations” to be very sensitive to minute changes in specifications of the first step one-by-one whitewashing schemes (in practice not always themselves unique or straightforward) and to the second step choice of “regressors”.

Although it took a very long time to find examples of multiple admissible maxima in time series (Dufour et al., 1980, 1983) and the literature is remarkably ambiguous concerning the possibility of their existence, the TRIO user should always carefully verify a preferred variant before accepting it as a final “reference” for the regression problem at hand. Our work on generalizations of correlation based on the idea of direct ordering of residuals (Ch. 6 and Blum et al., 1996) led to the L-2.0 algorithm (Liem et al., 1998) not included in TRIO Version 2.0.

Autocorrelation of residuals is easy to understand as a case of missing explanatory variables or, in time series, as implicit adjustment for a truly dynamic phenomenon (Spanos, 1987-88). It is much harder to find an adequate meaning for heteroskedasticity. L-1.4 allows the user to specify a very general correction, as shown in Ch. 5 and demonstrated in Ch. 7. However, we have not used the procedure as frequently as that for autocorrelation. It is also clear that, when the user wishes to specify a reasonably free form for the explanatory variables used in the heteroskedasticity formula, the risk of multicollinearity increases because all observations are divided by a nonlinear form, which may considerably change their relative values and make one observation stand out. Although the specific form programmed in TRIO rules out negative variances, the user should apply the procedure with care, remembering that, since the first version of L-1.4 became available in 1979, we have often found in practice and stated, for instance in Section 5.3.2, of Ch. 7, that correcting for heteroskedasticity sometimes produced worse results than not correcting (even when heteroskedasticity is present), a result also noted by others (Mishkin, 1990). Clearly, it should be all the more important to correct for heteroskedasticity.
that the BCT cannot be used without changing error variance – one would certainly want two sets of instruments to control for two objectives (as explained in Ch. 7). But the difficulty of thinking of other good mispecification reasons why heteroskedasticity would appear may explain our predicament, and the state-of-the-art. In practice, users should make very progressive extensions of their models when they try specifications of the heteroskedasticity function. They must also remember that, when they use an explanatory variable in the heteroskedasticity function, elasticity computations and other TRIO formulas will take due account of that presence, over and above taking into account the effect of its eventual presence as a normal regressor.

**Practice and programs.** In practice, transformations on variables or on functions make possible a large number of combinations – so large, in fact, that we have not tried every type of specification that is feasible with the programs. Let us then give, for each of the three algorithms, an idea of our experience as well as an idea of areas where the possibilities of TRIO have yet to be fully explored. As all programs were first developed with double precision calculations on a CDC Cyber machine, and as the results were reproduced with adequate precision on 32-bit machines, we can be reasonably sure that the procedures are precise enough. The use of numerical constraints also ensures that, to our knowledge, we avoid exceeding the precision of the computers. This is of particular concern in some of the S-1/S-5 model types that involve simultaneous use of overlapping nonlinear transformations (e.g. an exponential function transformed by an inverse BTT).

**L-1.4** described in Ch. 13 has been used extensively with both time series and cross sectional data sets. In the former case, we have in particular built multiple-equation applications where each of the equations has a common BCT on both dependent and independent variables, as well as an equation-specific autoregressive scheme (Gaudry, 1984), but where we rapidly reached diminishing returns when we used two or three BCT. In cross sections, we have shown that, if one used the proper BCT and a direct ordering of observations to estimate a first order autocorrelation process, one found that compulsory contributions to Social Pension Insurance in Germany led to partially offsetting behaviour on the part of savers (i.e. to some disaving), in contrast with non optimal linear or multiplicative forms that implied a complementarity between private and public savings (Blum and Gaudry, 1990). In that case, the optimal number of BCT was generally (depending of the socioeconomic group considered) two, although more could be estimated without difficulty.

**P-2** described in Ch. 15 has been used in a number of transportation applications of the BOX-COX LOGIT. As a rule, one could estimate 3 to 6 BCT without difficulty, but diminishing returns occurred rapidly after the first BCT. In every urban passenger case, 6 modes or more were considered and, in all cases, one could accept neither the linear nor the logarithmic transformations of the variables. In the case of the Paris 1976 data set (Gaudry, 1985) and of the Paris 1985 data set pertaining to access to Paris airports, the optimal value of the BCT assigned to network variables was 0.5; in the former case, two of the explanatory variables of the transit authority’s reference model coefficients changed signs in a more reasonable direction. In the Santiago case described in Ch. 11, the introduction of nonlinearity in the utility function corrected the unreasonable linear results. In intercity applications to passengers (Ch. 12) and to freight (Picard and Gaudry, 1993) the optimal results were also different from the linear results and, in the former case, implied quite different forecasts of High Speed Rail demand for Germany from those of the linear model.
As we have not yet carried out a large number of analyses with the Generalized Box-Cox Logit, the user should make sure that complexity in this case is introduced very gradually with constraints that prevent the BCT used on variables present in all utility functions from being equal to one another.

S-I/S-5 described in Ch. 14 has also been tested in varying degrees, depending on the model type considered. The Generalized Box-Cox option has been subjected to limited testing. The use of BCT on explanatory variables has been thoroughly tested in the LOGIT family, and found to be as useful with aggregate data as it is with disaggregate data, but has been subjected to limited testing in the other families (Standard DOGIT, Generalized DOGIT, LIN-IPT-LOGIT and BT-IPT-LOGIT. Moreover, as the unconditional asymptotic t-statistics of the linear coefficients of the regression problem depend upon the units of measurement (Spitzer, 1984), TRIO also computes the t-statistics conditionally upon the estimated values of the Box-Cox transformations, we recommend that, as a rule, the user routinely select those. ) because, as noted in Ch. 9, the emphasis of the tests has been to probe the interest of the “extra” parameters (the $\theta$, $\phi$ and $\mu$). The user should therefore either use fixed values for the $\lambda$, or proceed with great care in combining the transformation of explanatory variables (the $\lambda$) with the other transformation parameters for these models. As an evaluation of the usefulness of all of the possible combinations will require extensive further work, it appears better to allow users to try their own combinations and accept the corresponding risk than to wait perhaps two years and decrease the chances of either numerical difficulties or unreasonable results by imposing within the program restrictions on subsets of potential specifications.

It is also true that TRIO contains some procedures that have been tested only a limited number of times. One example is the procedure found in L-1.4 to compute the elasticity of the standard error of the dependent variable (in addition to the elasticity of the expected value of the dependent variable). Another example is the procedure found in L-1.4 ans S-1/S-5 to obtain maximum likelihood forecasts of the dependent variable, that is forecasts as solutions of a maximum likelihood problem (in addition to the forecasts obtained by the usual simulation procedure).

_Caveat indagator!_

Marc Gaudry
April 13, 1993
October 22, 1993
June 21, 2005

**References**


Babin, A., Dagenais, Florian, M., Gaudry, M., Guélat, J., Lestage, P. and T. C. Liem
"TRIO, An Open Interactive Graphic System for Demand Model Estimation", Proceedings
of the World Conference Transport Research 1986, 1690-1704, The Centre for Transport
Studies, University of British Columbia, or Publication #441, Centre de recherche sur les
transports, Université de Montréal, January 1986 and March 1987.

Describes the general design concepts and the open computer based system,
shows examples of graphic output. Shows how TRIO was conceived using
the same graphical user interface as EMME/2.

Gaudry, M. et al. "Cur Cum TRIO?" Publication #901, Centre de recherche sur les
transports, Université de Montréal, April 1994.

Summarizes in English, French, German and Spanish the professional, scientific
and technical features of TRIO, notably the full integration of the four basic tasks
required for regression work: the management of information, the production of
models, the analysis of data or model results, and assistance in report generation.
Shows examples of text, model results analysed with TABLEX tables and graphic
output.

Gaudry, M., Lestage, P., Guélat, J. and P. Galvan, "TRIO Tutorial Version 2.0", Pub-
lication #902, Centre de recherche sur les transports, Université de Montréal, February
1994.

Conducts the user through a first session with TRIO. Principal steps covered include:
creation of a database, model estimation with LEVEL, SHARE and PROBABILITY
algorithms, use of the TABLEX table to examine regression results and production
of graphs.

2.0", Publication #903, Centre de recherche sur les transports, Université de Montréal,
November 1994.

Provides full description of all functions available in TRIO, as they pertain for
instance to the system environment, the database, the variable and variant editors
for all model classes, the analysis of data and model results, through tabular and
graphical methods.

2.0", Publication #904, Centre de recherche sur les transports, Université de Montréal,

Contains a summary of the regression theory and methods used in TRIO. Demon-
strates the properties and usefulness of the methods with selected applications. De-
scribes in detail all available model types or specifications, as well as the estimation
techniques, which are programmed in TRIO.

2.0", Publication #996, Centre de recherche sur les transports, Université de Montréal,

This student manual conducts the TRIO user through a first session with TRIO using
a subset of the full TRIO TUTORIAL database and especially tailored examples.
Principal steps covered include: creation of a database, model estimation with
LEVEL, SHARE and PROBABILITY algorithms, use of TABLEX interactive table
to examine regression results, and production of graphs.