96

Reprinted from

TRANSPORTATION RESEARCH, PART B

Vol. 15*B*, *No.* 2, *pp.* 97–103

THE INVERSE POWER TRANSFORMATION LOGIT AND DOGIT MODE CHOICE MODELS

MARC J. I. GAUDRY

Centre de recherche sur les transports, Département de sciences économiques,
Université de Montréal, Montréal, Canada

PERGAMON PRESS
NEW YORK · TORONTO · OXFORD · PARIS
FRANKFURT · SYDNEY
1981

Transpn Res.-B, Vol. 15B, pp. 97–103, 1981
Printed in Great Britain.

0191–2615/81/020097–07$02.00/0

THE INVERSE POWER TRANSFORMATION LOGIT AND DOGIT MODE CHOICE MODELS†

MARC J. I. GAUDRY
Centre de recherche sur les transports et Département de sciences économiques, Université de Montréal

(*Received* 3 *July* 1979; *in revised form* 26 *November* 1979)

Abstract—This paper explores the properties of inverse Box-Cox and Box-Tukey transformations applied to the exponential functions of logit and dogit mode choice models. It is suggested that inverse power transformations allow for the introduction of modeler ignorance in the models and solve the "thin equal tails" problem of the logit model; it is also shown that they allow for asymmetry of response functions in both logit and dogit models by introducing alternative-specific parameters which make cross elasticities of demand among alternatives generally asymmetric. In the dogit model, modeler ignorance and consumer captivity remain conceptually distinct. Standard logit and dogit models appear as very special "perfect knowledge" cases in broad spectra of models which also include, among others, the reciprocal extreme value or log-Weibull variants. These improvements over the simple symmetric-thin-equal-tail-perfect-knowledge logit and the symmetric-pure-captivity dogit are achieved at the cost of introducing at the most two new parameters per alternative considered in the original logit and dogit mode choice models.

1. INTRODUCTION: TWO PROBLEMS OF MODE CHOICE ANALYSIS

The purpose of this paper is to suggest a way of approaching two problems which occur in the explanation of the choice of travel mode. The first problem, illustrated in Fig. 1, is that of the systematic underestimation of the choice probabilities of travel modes at the "end" of the distribution in logit models: this problem arises when consumers choose certain modes despite their apparent unattractiveness. We will call such systematic underestimation (or overestimation) the "thin equal tails" problems of the logit mode choice model. The second problem, illustrated in Fig. 2 by the shape of the hypothetical response curve (no observations are shown), is that of the symmetry of the consumers' potential response function in mode choice contexts. Asymmetry means that the tails of the distribution should not be approached in the same way for each mode considered because consumers' reactions to various modes can exhibit a certain amount of stickiness or mode-specificity in the choice of mode. The response curve in Fig. 2 reflects the fact that consumers can be induced to abandon one mode for another only if the difference in modal attributes between them reaches a certain level beyond which they react—perhaps in nearly step-wise fashion—to such differences.

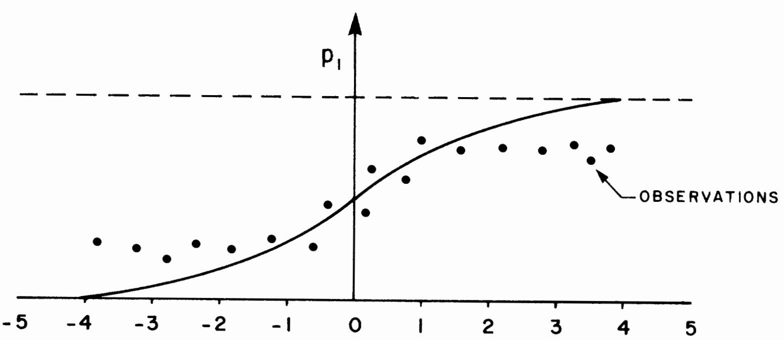

Fig. 1. Response curve of binomial symmetric logit model.

†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*.

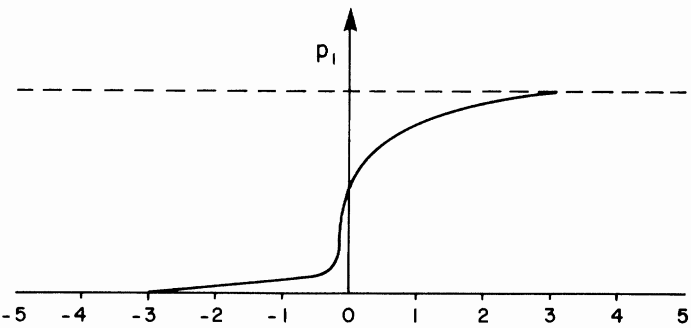

Fig. 2. Response curve of a binomial asymmetric model.

In order to formulate approaches to these problems, we shall apply inverse Box–Cox and Box–Tukey monotonic transformations to the exponential functions of logit and dogit models. We will suggest that inverse power transformations (IPT) introduce what amounts to imperfect modeler information or "ignorance" in the logit model which, in its standard formulation, is a special case implicitly based on the assumption of perfect modeling knowledge. We will argue that, in the dogit subclass, it is further possible to distinguish between modeler *ignorance of* the characteristics of alternatives and consumer *captivity to* these alternatives. We will also show that IPT further allow for asymmetry in the shape of the response function in both logit and dogit models.

In order not to burden our exposition unnecessarily, we will concentrate on the formal description of IPT models and minimize the use of practical examples and the references to the travel demand literature of which this article is a part.

2. INVERSE POWER TRANSFORMATIONS (IPT) OF INTEREST

(a) *The inverse Box-Cox (IBC) and Box-Tukey (IBT) transformations*

The monotonic transformations of interest here have been studied in the past by Anscombe and Tukey (1954), referred to in Tukey (1957), Box and Cox (1964), Zarembka (1974) and other. The Box–Tukey transformation (λ_z, μ_z) of any strictly positive non-Boolean variable z is defined as

$$z^{(\lambda_z, \mu_z)} = \begin{cases} \dfrac{(z-\mu_z)^{\lambda_z}-1}{\lambda_z}, & \lambda \neq 0,\ z-\mu_z > 0, \quad (1) \\ \ln(z-\mu_z), & \lambda \to 0,\ z-\mu_z > 0. \quad (1') \end{cases}$$

The standard Box–Cox transformation is a special case of (1)–(1′) obtained by setting $\mu_z = 0$. The inverse Box–Tukey (IBT) transformation $(\lambda_x, \mu_x)^{-1}$ of any variable x is

$$x^{(\lambda_x, \mu_x)^{-1}} = \begin{cases} (\lambda_x x + 1)^{\lambda_x^{-1}} + \mu_x & \lambda \neq 0, \quad (2) \\ e^x + \mu_x, & \lambda \to 0. \quad (2') \end{cases}$$

where $\lim[(\lambda_x x + 1)^{\lambda_x^{-1}} + \mu_x] = (e^x + \mu_x)$ as $\lambda_x \to 0$.

(b) *Inverse power transformation logit (IPTL) and dogit (IPTD) applications*

The first subclass of models in which we will use IPT functions is that of logit models defined as

$$p_i = \frac{e^{V_i}}{\sum_j e^{V_j}}, \qquad i, j = 1, \ldots, N, \tag{3}$$

where p_i: choice probability or market share of the ith alternative; V_i: a function of K independent variables associated with the ith alternative, for instance $V_i = \sum_k \beta_k X_{ik}$. In standard terminology, this function is called the representative utility of the ith alternative.

We will apply inverse power transformations to the exponential terms of the logit model, thus defining the IPTL subclass of models characterized most generally as

$$p_i = \frac{[(\lambda_i e^{V_i} + 1)^{\lambda_i^{-1}} + \mu_i]}{\Sigma_j [(\lambda_j e^{V_j} + 1)\lambda_j^{-1} + \mu_j]}, \quad i, j = 1, \ldots, N, \tag{4}$$

$$\left.\begin{array}{lll} \text{where} & \mu_i > -(\lambda_i e^{V_i} + 1)^{\lambda_i^{-1}}, & \text{if } 0 \le \lambda_i \le 1, \\ & |(\lambda_i e^{V_i} + 1)^{\lambda_i^{-1}}|, & \text{if } \lambda_i > 1. \end{array}\right\} \tag{4'}$$

We shall limit our comments to cases in which $0 \le \lambda_i \le 1$ because values of $\lambda_i > 1$ have less intuitive appeal.

Similarly, we will use IPT functions in the dogit model (Gaudry and Dagenais, 1979) defined as:

$$p_i = \frac{e^{V_i} + \theta_i \Sigma_j e^{V_j}}{(1 + \Sigma_j \theta_j) \Sigma_j e^{V_j}}, \quad i, j = 1, \ldots, N, \tag{5}$$

where $\theta_i \ge 0$ and other elements are defined as in (3). Application of inverse power transformation functions to the exponential terms of the dogit model defines the IPTD subclass of models characterized as

$$p_i = \frac{[(\lambda_i e^{V_i} + 1)\lambda_i^{-1} + \mu_i] + \theta_i \Sigma_j [(\lambda_j e^{V_j} + 1)^{\lambda_j^{-1}} + \mu_j]}{(1 + \Sigma_j \theta_j) \Sigma_j [(\lambda_j e^{V_j} + 1)\lambda_j^{-1} + \mu_j]}, \tag{6}$$

where (4′) holds and $i, j = 1, \ldots, N$.

Technically speaking, IPTL models are special cases of IPTD models and arise when $\theta_i = 0$ for all alternatives. We will still describe IPTL and IPTD models separately and concentrate on special cases in order to emphasize points of particular significance and not to complicate the description of the complex tree defined by the actual array of parameters absent from a logit format:

$$\left.\begin{array}{l} \lambda = (\lambda_1, \ldots, \lambda_i, \ldots, \lambda_N), \\ \mu = (\mu_1, \ldots, \mu_i, \ldots, \mu_N), \\ \theta = (\theta_1, \ldots, \theta_i, \ldots, \theta_N). \end{array}\right\} \tag{7}$$

3. THE INVERSE POWER TRANSFORMATION LOGIT (IPTL) MODEL

(a) *Main branches and variants*

The IPTL model given in (4)–(4′) comprises a number of cases of special interest based on the IBT transformation. Setting $\lambda = 1$, for instance, yields the variable tail exponential (VTE) variant

$$p_i = \frac{e^{V_i} + 1 + \mu_i}{\Sigma_j (e^{V_j} + 1 + \mu_j)}, \quad \mu_i > -(e^{V_i} + 1), \tag{8}$$

and allowing $\lambda \to 0$ yields the variable tail reciprocal extreme value (VTREV) or log-Weibull† variant

$$p_i = \frac{e^{(e^{V_i})} + \mu_i}{\Sigma_j (e^{(e^{V_j})} + \mu_j)}, \quad \mu_i > -1, \tag{9}$$

in which the reciprocal Weibull elements set a limit of -1 for the μ_i parameters.

Special cases based on the BCT are obtained by setting the $\mu_i = 0$ in (8) and (9). The equal tail exponential (ETE) and equal tail reciprocal extreme value (ETREV) variants are respectively

$$p_i = \frac{(e^{V_i} + 1)}{\Sigma_j (e^{V_j} + 1)} = \frac{e^{V_i} + 1}{\Sigma_j e^{V_j} + N} \tag{10}$$

†Johnson and Kotz (1970) define the extreme value or log-Weibull distribution as $p(y) = \exp\{-\exp(x)\}$. We define the reciprocal distribution as $p(y) = \exp\{\exp(x)\}$.

and

$$p_i = \frac{e^{(e^{V_i})}}{\Sigma_j e^{(e^{V_j})}}. \tag{11}$$

(b) *Formal properties*

(i) *Tails of the response function.* It is easy to see that, when the representative utility (the value of V_i) of an alternative approaches $-\infty$, the choice probability for that alternative approaches some value greater than zero in the IPTL model and in its special cases. Moreover, in the VTE and VTREV variants, the value to which the probability approaches will be different for different alternatives. This means that the tails of the response function can in principle differ among themselves and from zero, in contrast with the logit model where they are equal to zero.

(ii) *Asymmetry of cross-elasticities of demand.* The own and cross elasticities of market shares of IPTL models are respectively

$$\eta_{ii} = \lambda_i\beta_k X_{ik}\left\{\frac{e^{V_i}\left[\frac{1}{\lambda_i}(\lambda_i\, e^{V_i}+1)^{(1/\lambda_i)-1}\right]}{[(\lambda_i\, e^{V_i}+1)^{1/\lambda_i}+\mu_i]} - \frac{e^{V_i}\left[\frac{1}{\lambda_i}(\lambda_i\, e^{V_i}+1)^{(1/\lambda_i)-1}\right]}{\Sigma_j\,[(\lambda_j\, e^{V_j}+1)^{1/\lambda_j}+\mu_j]}\right\}, \tag{12}$$

and

$$\eta_{ij} = \lambda_j\beta_k X_{jk}\left\{-\frac{e^{V_j}\left[\frac{1}{\lambda_j}(\lambda_j\, e^{V_j}+1)^{(1/\lambda_j)-1}\right]}{\Sigma_j[\lambda_j\, e^{V_j}+1)^{1/\lambda_j}+\mu_j]}\right\}. \tag{13}$$

The corresponding own and cross elasticities in the logit are special cases of (12) and (13) obtained by setting the $\lambda_i = 1$ and the $\mu_i = -1$:

$$\eta_{ii} = \beta_k X_{ik}\left[1 - \frac{e^{V_i}}{\Sigma_j\, e^{V_j}}\right], \tag{14}$$

and

$$\eta_{ij} = \beta_k X_{jk}\left[-\frac{e^{V_j}}{\Sigma_j\, e^{V_j}}\right]. \tag{15}$$

There is no simple way to compare (12)–(14) and (13)–(15): their differences will depend both on the μ_i and on the λ_i. This is not surprising when the role of the λ_i is examined. In eqn (15) cross-elasticities of demand are symmetric in the sense that they do not depend on the specific i, j pair considered but depend solely on the attribute weight β_k, on the attribute level X_{jk} and on $e^{V_j}/\Sigma_j\, e^{V_j}$, the market share of the jth alternative. If, for alternatives j and l, $X_{jk} = X_{lk}$ and market shares of alternatives j and l are equal, cross elasticities η_{ij} and η_{il} will be equal because the β_k do not very across alternatives. By contrast cross-elasticity expression (13) introduces asymmetry, or alternative-specificity, through the λ_j. Symmetry only occurs if all λ_j are equal. Asymmetry means that, as in Fig. 2, the minimum market share of each alternative will not be approached in the same way for all alternatives. It means that the cross-elasticity of the ith market share will depend on *which* particular jth alternative is modified. Introducing alternative specificity in this way is economical because the λ_j are associated with each alternative considered as a whole rather than being associated with each variable in each alternative: this implies N additional parameters instead of the very large number of possible parameters involved in a mode-specific models. Asymmetry is a useful property because there is no particular reason to assume that symmetry should occur.

(c) *Interpretation of those properties*

(i) *The role of the μ_i: modeler ignorance.* One possible interpretation of non zero and variable tails is to say that the constants in (4) represent the choice probabilities levels which remain *unexplained* by the modeler. If the modeler has no variables which can be used in the representative utility function, the V_i will equal zero for all alternatives and the μ_i will "explain" choices; as the modeler incorporates information into the model, this level of the "initial ignorance" is modified gradually. In VTE variant (8), the modeler's explanation may be of such a nature that the μ_i equal -1, a situation where basic choice probability levels have been explained entirely by the information used in the representative utility functions: in that sense, the logit model obtained as a special case of model VTE represents a situation of complete knowledge of the determinants of choice.

It may also be possible to interpret the basic levels represented by the constants as the "carryovers" from a previous time period or from an underlying choice process in which habit has a role but we will not develop this interpretation in this paper.

(ii) *The role of the* λ_i: *mode specificity.* We pointed out above the λ_i (and the μ_i) make the cross-elasticities of demand asymmetric in the sense that cross-elasticities η_{ij} and η_{il} will differ even if alternatives j and l have the same representative utilities. This mode-specificity of response is also true of the own elasticity shown in (12). Mode-specificity can occur in mode choice problems because of errors of specification of the representative utility functions: time and comfort, for instance, are very difficult to separate because comfort colours all of the characteristics of the mode; similarly, incomplete descriptions of consumer characteristics can lead to mode-specific responses.

4. THE INVERSE POWER TRANSFORMATION DOGIT (IPTD) MODEL

All IPTL models described above share with the logit model the property of consistency with the independence from irrelevant alternatives axiom (IIA) of choice theory which requires that the relative probabilities of choice of any two alternatives i and j depend solely on the attributes of these alternatives. This can be readily verified by noticing that relative probabilities in (4) involve only the utility functions V_i and V_j, the ignorance levels μ_i and μ_j and the asymmetry parameters λ_i and λ_j.

The dogit model, however, is not generally consistent with the independence from irrelevant alternatives axiom because, in (5), all system parameters are involved in determining the relative probabilities of any pair of alternatives. Formally, McFadden (1978) has recently shown that the dogit model is a special case of mixed probability random utility models. A simple interpretation of the dogit model can be obtained by rewriting (5) as

$$p_i = \frac{e^{V_i}}{(1+\Sigma_j \theta_j)\,\Sigma_j e^{V_j}} + \frac{\theta_i}{(1+\Sigma_j \theta_j)}, \qquad \theta_i \geq 0, \tag{16}$$

where it is easy to see that the second term on the right-hand side represents the particular positive level reached by p_i as $e^{V_i} \to 0$. As each alternative can have a distinct θ_i, the dogit is by structure a variable tail model. It has recently been suggested (Gaudry, 1978; Gaudry and Wills, 1979) that the dogit can be more useful than the logit for situations in which a part of the market appears to be captive to an alternative.

By using inverse power transformations in a dogit framework, it is in principle possible to distinguish between ignorance and captivity as both elements have distinct roles in determining the size of tails. Indeed, as $e^{V_i} \to 0$, choice probabilities of IPTD models defined in (6) tend to

$$p_i = \frac{1+\mu_i+\theta_i\,\Sigma_j\,[(\lambda_j\,e^{V_j}+1)^{\lambda_j^{-1}}+\mu_j]}{(1+\Sigma_j \theta_j)\,\Sigma_j\,[(\lambda_j\,e^{V_j}+1)^{\lambda_j^{-1}}+\mu_j]}. \tag{17}$$

By contrast, choice probabilities of IPTL models tend to

$$p_i = \frac{1+\mu_i}{\sum_j [\lambda_j\,e^{V_j}+1)^{\lambda_j^{-1}}+\mu_j]}. \tag{18}$$

As both μ_i and θ_i parameters determine the thickness of tails, it might not be possible to distinguish between them easily in particular applications. So only empirical work can determine whether the conceptual difference between them is of practical import.

It would be straightforward to show that the market share elasticities of the IPTD model (not shown here) are also asymmetric in the sense that η_{ij} and η_{il} will generally differ when $V_j = V_l$. When we further take into account the fact that the dogit format generally implies inconsistency with the IIA axiom, or $\eta_{hj} \neq \eta_{ij}$, we see that, when all representative utilities are equal, the non-diagonal elements of both rows (asymmetry) and columns (IIA) of a matrix of IPTD cross-elasticities will generally differ among themselves; by contrast, because of consistency with IIA, only the non-diagonal elements of the rows of the matrix would differ in the IPTL cross-elasticity matrix.

The dogit-based analogs of the VTE, VTREV, ETE and ETREV models (8)–(11) are straightforward. The VTED variant (adding suffix D to indicate the dogit-based format), for instance, is

$$p_i = \frac{(e^{V_i} + 1 + \mu_i) + \theta_i \Sigma_j (e^{V_j} + 1 + \mu_j)}{(1 + \Sigma_j \theta_j) \Sigma_j (e^{V_j} + 1 + \mu_j)}. \tag{19}$$

5. MODEL ESTIMATION

An important question is that of the possibility of estimating the additional inverse power transformation parameters. Consider an example. Assume that one has observations on market shares of two modes, y_1 and $y_2 = 1 - y_1$, and that one is interested in the IPTD model. Rewriting (6) in the format of (16), rearranging terms and taking logarithms of the ratios of market shares yields

$$\ln\left[\frac{y_1 - \theta_1/1 + \theta_1 + \theta_2}{1 - y_1 - \theta_2/1 + \theta_i + \theta_2}\right] = \ln[(\lambda_1 e^{V_1} + 1)^{\lambda_1^{-1}} + \mu_1] - \ln[(\lambda_2 e^{V_2} + 1)^{\lambda_2^{-1}} + \mu_2] + u, \tag{20}$$

where an error term has been added and the admissible regions for θ_1 and θ_2 are respectively $[\theta_1/1 + \theta_1 + \theta_2 < \min(y_1)]$ and $[\theta_2/1 + \theta_1 + \theta_2 < \min(y_2)]$. If we assume that the error term is normally and identically distributed with a mean equal to zero and a constant variance σ^2, the likelihood of the original N observations on market share y_1 is

$$L(y_1; \beta, \sigma^2, \lambda, \mu, \theta) = (2\pi\sigma^2)^{-N/2} \exp(-u'u/2\sigma^2) \cdot |J(\theta; y_1)|, \tag{21}$$

where $|J(\theta; y_1)|$ is the absolute value of the Jacobian of the transformation from the dependent variable in (20) to the actually observed y_1. This function may be maximized by nonlinear methods or more simply by scanning. In terms of testing for the values of the parameters of the model, one approach is to obtain asymptotic variances and covariances for these parameters from the matrix of second partial derivatives of the likelihood functions and to perform asymptotic tests based on the normal distribution. Another approach is to use the likelihood ratio test and to compare the value of (21) at the maximum under more or less severe constraints. It should therefore be possible to estimate the parameters of both IPTL and IPTD models and to allow the data to decide whether the most general forms or the main variants explain the data better than the logit and dogit special cases.

6. CONCLUSION

The use of inverse power transformations on logit and dogit formats may be promising because it permits the incorporation of interesting parameters which can solve the "thin and equal tails" problem of the logit model and increase the realism of model structures at the same time: they allow for variable tails and asymmetry or mode-specificity of the response functions. These gains do not require the use of a large number of supplementary parameters (2 per alternative) over and above those of logit or dogit formats.

We note in passing that it is possible to use in IPT models the Box–Cox and Box–Tukey transformations on the *variables* of the V_i functions. These functions have been shown to be useful in improving both logit (Gaudry and Wills, 1978) and dogit models (Gaudry and Wills, 1979) and yield interesting special cases such as the product market share mode choice model.

There are reasons to hope that the simultaneous use of power transformations on the V_i functions and of inverse power transformations on the exponential functions of logit and dogit models will improve our understanding of mode choice processes.

Acknowledgements—This research was 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.* François Soumis provided a partial impetus for this research in the context of a discussion on the origins of the dogit model and Marcel Dagenais and a referee made very useful comments.

REFERENCES

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

Box G. E. P. and Cox D. R. (1964) An analysis of transformations. *J. R. Stat. Soc.* Series B, 211–243.

Gaudry M. (1978) A Dogit Model of Travel Mode Choice in Montreal. Publication No. 104, Centre de recherche sur les transports, Université de Montréal. *The Canadian Journal of Economics* (1980) (To be published).

Gaudry M. and Dagenais M. (1979) The dogit model. *Transpn Res.* **13**, 105–112.

Gaudry M. and Wills M. (1978) Estimating the functional form of travel demand models. *Transpn Res.* **12**, 257–289.

Gaudry M. and Wills M. (1979) Testing the dogit model with aggregate time-series and cross-sectional travel data. *Transpn Res.* **13B**, 155–166.

Johnson N. L. and Kotz S. (1970) *Continuous Unvariate Distributions—1,* Chapt. 21. Houghton Mifflin, Boston.

McFadden D. (1978) *Econometric Models for Probabilistic Choice,* Mimeographed, p. 19. Paper presented at the NBER-NSF Conference on Decisions Rules and Uncertainty, Carnegie-Mellon University.

Tukey J. W. (1957) On the comparative anatomy of transformations. *Anneals of Mathematical Statistics* **28**, 602–632.

Zarembka P. (1974) Transformation of variables in econometrics. *Frontiers in Econometrics* (Edited by P. Zarembka), Chap. 3.