Addressing homoskedasticity in invariant random utility models

Vittorio Marzano [a,*], Andrew Daly [b]

[a] *Dipartimento di Ingegneria Civile, Edile ed Ambientale - Università di Napoli "Federico II" Via Claudio 21, 80125 Napoli, Italy*

[b] *Institute for Transport Studies, University of Leeds and RAND Europe*

* Corresponding author, Tel.: +39 081 7683935; fax: +39 081 7683946.

E-mail addresses: vmarzano@unina.it (V. Marzano), daly@rand.org (A. Daly)

Abstract

This paper addresses an important issue in the structure of invariant choice models, showing that the apparently important difference between homoskedastic and heteroskedastic models is not fundamental. The key result in the paper is the Homoskedasticity Theorem, showing that for an IRUM with normally distributed utilities, i.e. a probit model, there exists a homoskedastic model with the same choice probabilities and elasticities. Additionally we show that the homoskedastic model can be free of negative correlations between the utilities and we explain how to find, among the homoskedastic models with the required choice probabilities, the model with minimum correlations. For models with disturbances of other forms, this result should hold approximately. Relevant consequences of the Homoskedasticity Theorem are also discussed. First, the variance can be increased without limit, but there is a strict lower limit, which can be calculated based on constructing a matrix with minimal variance. Then, using the same mathematical approach, models are derived that have the same utility difference covariance matrix as any given IRUM, but are homoskedastic and have all their covariances positive, so that it is possible that closed-form GEV models could be found to approximate them.

Keywords

Invariant random utility models; homoskedasticity; generalised extreme value models

1. Introduction

This paper addresses an important issue in the structure of choice models. Many of these models are based on the random utility paradigm and within that paradigm an important characteristic of the models is whether they can be classified as homoskedastic or heteroskedastic, i.e. whether the random utilities on which the model is based have equal or unequal variances.

The skedasticity of models is particularly important in considering the function of GEV models, which since their formulation by McFadden (1978, 1981) have played a central role in choice modelling. GEV models are homoskedastic, because of the way they are defined, whereas many of the models developed more recently are heteroskedastic. In this paper we focus on the large group of invariant random utility models (IRUM), including GEV models, those for which the variances are not dependent on the values of utility.

The primary contribution of this paper is to show that this apparently important difference between homoskedastic and heteroskedastic models is not fundamental. The key result in the paper is a theorem, the Homoskedasticity Theorem, showing that for an IRUM with normally distributed utilities there exists a homoskedastic model with the same choice probabilities and elasticities.

Additionally we show that the homoskedastic model can be free of negative correlations between the utilities and we explain how to find, among the homoskedastic models with the required choice probabilities, the model with minimum correlations. For models with disturbances of other forms, this result should hold approximately.

Apart from uncovering an important property of choice models in general, the paper shows that the scope for modelling using homoskedastic models (e.g. GEV models) is less limited than may have been thought. GEV models are homoskedastic, with non-negative correlations between the utilities, so that being able to transform all IRUMs to have these properties opens the possibility for substitution or approximation of more complex models by GEV models which have computational advantages for both estimation and application. For example, the papers by Marzano and Papola (2008) and Marzano et al. (2013) make considerable progress in defining cross-nested logit models, which are in the GEV family, to fit most feasible correlation structures.

Formally, following Marschak (1960), we write the choice probabilities for a RUM with a finite choice set C_n for choice situation n as

$$Pr\{choice\ in\ situation\ n = c|C_n\} = Pr\{U_{cn} \geq U_{jn} \forall\ j \in C_n\}$$

where U_{jn} gives the consumer's utility of alternative j in choice situation n and

$$U_{jn} = V_{jn} + \epsilon_{jn}$$

where V_{jn} is the analyst's approximation of the utility of alternative j in situation n and
ϵ_{jn} is an error which includes that approximation as well as other decision-maker-related sources of randomness (Ben-Akiva and Lerman, 1985; Cascetta, 2009).

The covariance matrix of ϵ_{jn} across the alternatives will be denoted by Σ_n. Then we can define 'invariance' precisely by deeming a model to be invariant if Σ_n does not depend on V_{jn} for any j (or any component of V_{jn}), i.e. using the 'strong' definition of invariance as set out by Daly (2007) and Ibáñez (2007). This notation will be used throughout the paper.

Section 2 of the paper discusses the ways in which heteroskedasticity affects different choice situations and how this can be eliminated. This is followed in Section 3 by a review of the relevant literature on the structure of choice models and on the relationship between IRUM and GEV models. Section 4 presents the main mathematical lemmata and theorems, while Section 5 summarises the findings.

2. Heteroskedasticity across choice situations

In a choice model, heteroskedasticity may arise because of heterogeneity in the variances of random residuals within a given choice situation and/or across choice situations. In addition, heteroskedasticity may also arise as a function of the systematic utility V_j of each alternative j, or some of its components: such heteroskedasticity with respect to the systematic component of utility is not covered in the present paper, which deals with models that are invariant in this respect, i.e. IRUM.

Initially, we will focus attention instead on heteroskedasticity (not arising from V) across choice situations, i.e. n: this type of variation can relate to the individual consumer and/or to the set of alternatives that are available and causes Σ_n to be different for different n. The main procedure that can be used to eliminate variation in Σ_n is scaling. For example, suppose we have a model in which the variance σ_n^2 of ϵ_{jn} depends on n, but not on j. Then if we write

$$U'_{jn} = {}^{U_{jn}}/_{\sigma_n} = {}^{V_{jn}}/_{\sigma_n} + {}^{\epsilon_{jn}}/_{\sigma_n}$$

we obtain a model which has the same ranking of preference, i.e. choice is unaffected, but which is homoskedastic. This procedure was applied by Daly and Zachary (1975) to achieve homoskedasticity in a binary choice model.

The context of this paper is not that of model estimation, where σ_n would be unknown, but of the theoretical properties of the model and its application once the model had been fully estimated. It is reasonable then to assume that σ_n is known and that we can make the scaling transformation as discussed. In addition, as pointed out by Marzano et al. (2013), there are choice contexts, such as route choice and activity-based modelling, in which the analyst may have prior expectations on covariances amongst alternatives before model estimation.

In models with multiple alternatives, where the variance of ϵ_{jn} depends on both j and n, the application of scaling is less obvious. However, scaling can be applied to achieve uniformity in the variance of (say) alternative 1, or in the determinant of Σ_n, or some other appropriate measure. Because of this, we focus in the remainder of the paper on the issue of heteroskedasticity across the alternatives and we drop the subscript n accordingly.

Once heteroskedasticity across choice situations has been removed through scaling, we will focus attention on heteroskedasticity with respect to the alternatives, i.e. affecting the rows of Σ: this is the main topic of the paper and is covered in the following sections.

3. Review of key literature

This section is devoted to a brief literature review, aiming at identifying the most relevant contributions related to the concepts of IRUM and of homoskedasticity.

For the purposes of the paper, two noteworthy classifications of IRUM are of interest, namely depending on:

- the hypotheses underlying the cumulative distribution function of the $U_j \forall j \in C$. For instance, GEV models assume that the random part ϵ of the utilities follows a Multivariate Extreme Value (MEV) distribution derived from a generating function complying with the requirements of the GEV theorem (McFadden, 1978), whilst the Probit model is based on a multivariate normal distribution and the Mixed Logit model is based on mixtures of normal (or lognormal etc.) and MEV distributions;
- the mathematical structure of the probability statement. IRUM may be closed-form or non-closed-form models, depending on the availability of a closed-form probability statement, i.e. not requiring explicit integration (e.g. by simulation) for the calculation of choice probabilities. Whether or not a closed form exists depends on the form of the distribution of ϵ.

Such classifications are mutually integrated in well-known common beliefs in the literature: for instance, GEV models are expected to be homoskedastic, invariant and closed-form, while heteroskedastic models are expected to be non-GEV and usually not closed form. In fact, in partial contrast with the above, some papers provide evidence that the class of GEV models is wider than expected with respect to the class of IRUM. Therefore, for the purposes of the paper, it is worth reviewing first the mathematical characterisations of GEV models, and then the relationship between GEV and IRUM.

GEV models represent so far the simplest and most effective class of IRUMs, thanks to some useful and intriguing properties. A set of sufficient conditions for the practical implementation of GEV models is given by McFadden's (1978) theorem, which allows us to obtain a GEV model from a homogeneous nonnegative function $G(.)$ satisfying specific requirements, leading to an (apparently) closed form probability statement and to a Multivariate Extreme Value (MEV) distribution of random

residuals with homoskedastic[1] Gumbel marginals. Therefore, any GEV-genuine generating function $G(.)$ leads to a homoskedastic model.

Smith (1984) proves that McFadden's conditions are also necessary to define the class of GEV generating functions. Ibáñez (2007), in the context of an exhaustive systematization of the consistency of probabilistic choice systems with random utility maximisation, reformulates the McFadden-Smith conditions by introducing a weaker condition on the mixed partial derivatives of the generating function.

Joe (2001) provided an exhaustive review of the statistical properties of the multivariate extreme value (MEV) distributions, based on the so-called Pickands' representation theorem (Galambos, 1987, theorem 5.4.3, p. 265). A remarkable outcome is that non-closed-form MEV distributions can be specified, therefore leading potentially to non-closed-form GEV models[2]. In that respect, it is worth underlining that neither the G function nor the F distribution of the McFadden (1978) theorem are required to exhibit a closed-form expression, i.e. allowing for non-closed-form GEV models is at least in principle consistent with McFadden's theorem requirements. Karlström (2003) showed the Pickands' representation of MEV distribution functions with Gumbel marginals to be consistent with the McFadden-Smith characterization of GEV models, and proposed a new GEV model (termed the PNL-L model) with a potentially non-closed generating function.

These suggestions and indications of the existence of non-closed-form GEV models raise a natural and quite relevant theoretical question: how general is the GEV class if compared to the IRUM class? Notably, since GEV models are inherently homoskedastic and since also heteroskedastic IRUM exist, answering such a question has implications also for the relationship between homoskedastic and heteroskedastic IRUM. A general characterization of the IRUM class is reported by Mattsson et al. (2014) and by Lindberg et al. (1995).

More specifically, Dagsvik (1995) proved that any IRUM probabilities[3] can be approximated through a GEV model. Specifically, given a vector of perceived utilities $\boldsymbol{U} = (U_1 \dots U_J)$ with a joint probability density function $f(y_1 \dots y_J)$, Dagsvik proves that the following joint cumulative distribution function:

$$F(x_1 \dots x_J) = \exp\left\{ - E\left[\left(\sum_k e^{\frac{U_k - x_k}{a}} \right)^{a^2} \right] \right\} =$$

$$= \exp\left\{ - \int_{y_1=-\infty}^{+\infty} \dots \int_{y_n=-\infty}^{+\infty} \left(\sum_k e^{\frac{y_k - x_k}{a}} \right)^{a^2} f(y_1 \dots y_J) dy_1 \dots dy_J \right\}$$

is an MEV distribution approximating, as $a \to 0$, the choice probabilities provided by the IRUM based on $\boldsymbol{U}$. Therefore, since the GEV class is dense in a mathematical sense within the IRUM class, any heteroskedastic IRUM might be approximated through a homoskedastic model. Anyway, practical consequences of Dagsvik's generating functions (e.g. what sorts of model are actually

[1] Notably, McFadden (1978) does not state explicitly the homoskedasticity of the GEV models. However, thanks to the homogeneity of the generating function $G(.)$, the marginal distribution of a single random residual ϵ_j may be obtained from the joint MEV distribution $F(\epsilon) = \exp(-G[\exp(-\epsilon)])$ as $F_j = \exp\left(-a_j \exp\left(-\epsilon_j\right)\right) = \exp\left(-\exp\left(-\epsilon_j + \log a_j\right)\right)$. In this expression, $a_j = G(0, \dots, 1, \dots)$ is a location parameter enabling different mean values among random residuals, but (see page 578 of McFadden (1978)) not affecting the variance of the marginal random residuals.

[2] See the generating function before the statement of the theorem on p. 181 in Joe (2001).

[3] Notably, the characterisation of invariant RUM is not explicitly stated in Dagsvik (1995), even though it might be considered necessary in the light of his assumption A.1 (p. 91).

obtained from such approximations) have not been explored yet in the literature. More recently, Fosgerau et al. (2013), in the context of establishing a general framework for the choice probability generating functions, review the role of RUM with multivariate extreme value distributed utilities and provide a slight generalisation of the Dagsvik result.

Daly (2007) shows that additive IRUMs defined by utility distributions meet McFadden's GEV theorem requirements and therefore 'are' (homoskedastic) GEV models. This finding is based on a function H which is a multi-dimensional integral over the utility space, which is shown to exist, and which can be exponentiated to get a McFadden GEV function. In particular, it allows proof that the homogeneity condition on the generating function G is equivalent to the assumption of model additivity. Therefore, a non-closed H function, respecting the assumption of additivity, can be chosen to meet McFadden's assumptions. This result has a twofold implication: firstly, it implies again that all IRUM are equivalent to homoskedastic models (e.g. an invariant Heteroskedastic Extreme Value model would be in fact equivalent to a homoskedastic GEV model, though not, of course, of closed form); secondly, it is possible to have non-closed-form GEV models.

It is therefore worth looking for a different viewpoint able to support these tenets and to find in practice the homoskedastic equivalent of a given heteroskedastic IRUM. This point will be made more explicit in the following section, where it will be explored from a new and different standpoint.

4. The Homoskedasticity Theorem

In this section we prove the Homoskedasticity Theorem which is the central result of the paper and follow up with results on the minimum variance of utilities and the possibility of making all the covariances positive. First, however, we establish a lemma concerning IRUM.

4.1 Potential for equivalent models

In sections 1 and 2 above we discussed models defined by the fundamental IRUM formula for each alternative j

$$U_j = V_j + \epsilon_j$$

This model may be transformed by adding further components

$$U_j = V_j + V_0 + \epsilon_j + \epsilon_0$$

where V_0 is a fixed value and ϵ_0 is a random value. Provided V_0 and ϵ_0 do not vary across the alternatives, the model is essentially unchanged by the addition of these components, in the sense that the order of preference of the alternatives and hence the choice probabilities are unchanged.

The fact that the addition of a constant value V_0 does not change the model is well known and can be considered to be part of the specification of RUM, following Marschak (1960). However, the addition of a random term is less familiar but also has no impact on the choices. In this paper we exploit this possibility extensively, choosing ϵ_0 to have specific patterns of correlation with ϵ_j to change the variance and covariance of U but without changing the choice probabilities or elasticities.

The choice probabilities of the model are essentially defined by specifying a matrix Ω of utility *differences*. It may be noted that the covariance matrix Σ of the utilities has $\frac{1}{2}J(J+1)$ degrees of freedom, if there are J alternatives, while the matrix of utility differences has $\frac{1}{2}J(J-1)$ degrees of freedom, i.e. J fewer. Within these J degrees of freedom, which do not affect the choice probabilities, there is considerable scope for changing the specification.

For example, a common procedure, following Daganzo (1979), is to choose one alternative (e.g. alternative 1) as the base and to set $V_0 = -V_1,\ \epsilon_0 = -\epsilon_1$ (so that $U_1 = 0$). This transformation reduces the dimensionality of the problem for processing in a number of contexts. However, this freedom can be exploited in other ways to achieve other objectives. Let $S(\Sigma)$ be the set of matrices that yield the same choice probabilities as Σ.

Starting from these premises, Daly (2001) explored the possibility of finding within $S(\Sigma)$ a covariance matrix Σ_h that was homoskedastic. Let $H(\Sigma) \subseteq S(\Sigma)$ be the set of such matrices, if they exist. The following **Homoskedasticity Lemma** characterises any such matrices.

> A homoskedastic utility covariance matrix giving the same covariance matrix of utility differences as Σ has the form
>
> $$\Sigma_h = \begin{bmatrix} k & k - \frac{\delta_{12}}{2} & \dots & k - \frac{\delta_{1J}}{2} \\ & k & \dots & k - \frac{\delta_{2J}}{2} \\ & & \dots & \dots \\ & & & k \end{bmatrix}$$
>
> where $\delta_{jk} = \sigma_j^2 + \sigma_k^2 - 2\sigma_{jk}$
> σ_j^2 gives the variance of ϵ_j and σ_{jk} the covariance of ϵ_j and ϵ_k, i.e. these are the elements of Σ and
> k is the homoskedasticity parameter.
>
> Proof of the lemma follows from simply writing down the conditions that such a matrix must satisfy.

It will be useful in the remainder of the paper to refer to the property that the covariance matrix of a random vector is positive definite, in the absence of linear dependence between the components of the vector. Of course, this property still applies if the random vector is constructed by adding a single random number to a random vector.

For example, it is essential that Σ_h should be positive semi-definite to maintain a real-valued model and indeed it should be positive definite to maintain a non-degenerate model. Because of the condition for covariance matrices, this will be the case if it is created by adding a real-valued ϵ_0 to the utilities, but it is not immediately obvious from its form that Σ_h is positive (semi-)definite. Daly (2001) gave the necessary condition $k > \max(\delta_{jk})/4$, but also gave a simple example showing that this condition is not sufficient.

The number of degrees of freedom in Σ_h is $\frac{1}{2}J(J-1)+1$, i.e. this matrix covers the space of utility differences (as is clear from the construction of δ) plus the additional degree of freedom given by k, which does not affect the utility difference distribution.

In the following section we establish that, for models with normally-distributed utilities, positive definite matrices of this form can be created by appropriate specification of ϵ_0. It is clear that if such a matrix can be found, further matrices in $H(\Sigma)$ can be developed with larger values of k by increasing the variance of ϵ_0, e.g. by adding a further independent random component. However, reduction of the variance of ϵ_0 is possible only down to the limit of zero, beyond which point we no longer have a real-valued model, which suggests that there might be a minimum value of k. This issue is discussed in section 4.3 below.

4.2 The Homoskedasticity Theorem

We prove the following theorem.

> *For a non-degenerate IRUM with normally-distributed utilities (i.e. a Probit model), there exists a homoskedastic IRUM that gives the same choice probabilities and elasticities.*

The proof consists of constructing a real random variable ϵ_0 that is added to the utilities of the original IRUM to make their covariance heteroskedastic. Because we are adding the same value to all the utilities the choice probabilities remain constant, as do their derivatives, i.e. the elasticities are also constant.

Let Σ be the covariance matrix of ϵ, i.e. giving the covariance of ϵ_j over the alternatives j for a non-degenerate IRUM. Let Λ be the Cholesky decomposition of Σ, i.e. the unique lower triangular matrix such that $\Lambda\Lambda^T = \Sigma$. We know that Λ exists because Σ is positive definite, being the covariance matrix of the utilities of a non-degenerate model. We may then define normally distributed variables η to have the same covariance matrix as ϵ by

$$\eta_j = \sum_{k=1}^{j} \lambda_{jk} \xi_k$$

where λ_{jk} are the elements of Λ and
ξ_k are independent standard normal variables.

Let σ_j be the diagonal elements of Σ, i.e. the variance of η_j. Define coefficients β for each alternative recursively:

$$\beta_1 = {}^{\sigma_1}\!/_{2\lambda_{11}}$$

$$\beta_j = {}^{\left(\sigma_j - \sum_{k=1}^{j-1} \beta_k \lambda_{kj}\right)}\!/_{2\lambda_{jj}}, \text{ for } j > 1$$

These definitions imply that $\sigma_j = 2\sum_{k=1}^{j} \beta_k \lambda_{kj}$.

Consider $\epsilon_0 = -\sum_{j=1}^{J} \beta_j \xi_j$. We find that

$$var(\epsilon_0) = \sum_{j=1}^{J} \beta_j^2 = k, \text{ say}$$

$$covar(\epsilon_0, \eta_j) = covar\left(-\sum_{j=1}^{J} \beta_j \xi_j, \sum_{k=1}^{j} \lambda_{jk} \xi_k\right) = -\sum_{k=1}^{j} \beta_k \lambda_{kj} = -\frac{1}{2}\sigma_j$$

Define $\epsilon_j^* = \epsilon_j + \epsilon_0$. Then

$$var(\epsilon_j^*) = var(\epsilon_0) + var(\epsilon_j) + 2covar(\epsilon_0, \epsilon_j) = k + \sigma_j + 2covar(\epsilon_0, \epsilon_j)$$

Because ϵ is assumed to be normally distributed, $\eta = \epsilon$ is the unique vector with covariance matrix Σ, so that

$$var(\epsilon_j^*) = k + \sigma_j + 2covar(\epsilon_0, \eta_j) = k$$

This is constant and the model defined by ϵ^* is therefore homoskedastic. Because of its construction it has choice probabilities equal to the model defined by ϵ, proving the theorem.

It may be noted that the only reliance on the normality of ϵ is in calculating its covariance with ϵ_0. When the model is not based on normal disturbances, the representation by Cholesky components is only approximate and the covariance calculation and homoskedasticity finding are therefore also only approximate.

The value of k that is obtained, i.e. $\sum_{j=1}^{J} \beta_j^2$, is well-defined but is not a characteristic of Σ. In fact, if we apply the procedure to the transformed matrix, we obtain a different value of k. It is therefore worth enquiring whether there are specific values of k that are more useful than others and this is the subject of the Section 4.3.

To obtain the off-diagonal elements of the utility covariance matrix of the transformed model, we calculate

$$covar(\epsilon_j^*, \epsilon_k^*) = covar(\epsilon_j, \epsilon_k) + var(\epsilon_0) - covar(\epsilon_0, \epsilon_j) - covar(\epsilon_0, \epsilon_k)$$

$$= \sigma_{jk} + k - \tfrac{1}{2}\sigma_j - \tfrac{1}{2}\sigma_k = k - \tfrac{1}{2}\delta_{jk}$$

with δ_{jk} defined as in section 4.1. The matrix of utility variances of the transformed model is thus exactly of the form required in section 4.1. Additionally, because of the construction of ϵ^*, we know that the matrix is positive semi-definite and, in the absence of linear dependence, positive definite.

4.3 Minimum-variance models

The parameter k appears as a positive value in all the cells of the matrix in 4.1. Given that k can simply be increased without limit, by adding an independent random variable to all the utilities, it is interesting to ask the questions

- what is the minimum value of k that is consistent with a real or non-degenerate model?
- what is the minimum value of k that ensures that all of the off-diagonal elements in the matrix are non-negative?

The second question arises in the context of approximation of models by GEV near-equivalents, which require homoskedasticity and non-negative correlation of utilities.

In the Appendix, it is shown that, for matrices with the structure of Σ_h, we can write the determinants as

$$\det \Sigma_h = a.k + b$$

and that, in this equation, $a > 0$. There is then a minimum value of k for which one of the sub-determinants of Σ_h is zero, noting that the relevant sub-matrices have the same structure as Σ_h. Then, by the Jacobi-Sylvester criterion, the matrix Σ_h is positive definite for all values of k greater than this minimum.

Define the vector $\mathbf{1}$ to have all its elements 1 and to be of length J. Then for a random vector ϵ with non-degenerate covariance matrix Σ we can calculate a random variable

$$\zeta = (\Sigma^{-1}\mathbf{1})^T \epsilon = \sum_k \sum_h \tau_{kh} \epsilon_k$$

where τ are the elements of Σ^{-1}. This variable ζ has the following simple variance formula

$$var(\zeta) = (\Sigma^{-1}\mathbf{1})^T \Sigma (\Sigma^{-1}\mathbf{1}) = \mathbf{1}^T \Sigma^{-1} \mathbf{1} = z, \text{ say}$$

where z is the sum of the elements of Σ^{-1}. It must be the case that $z > 0$, because Σ^{-1} is positive definite. We can also calculate

$$covar(\zeta, \epsilon_j) = \sum_k \sum_h \tau_{kh} \sigma_{kj} = 1$$

where σ are the elements of Σ, the equality to 1 following because τ are simply the elements of the inverse matrix. Because of this uniform covariance, adding any multiple of ζ to the utilities preserves the skedasticity of the model. Thus if the model has been adjusted to be homoskedastic, using the procedures of the Theorem, we may investigate the maximum value of a scalar μ for which we can subtract $\mu\zeta$ from the utilities of the model while retaining definiteness. We calculate

$$var(\epsilon_j - \mu\zeta) = \sigma_{jj} + \mu^2 z - 2\mu$$

$$covar(\epsilon_j - \mu\zeta, \epsilon_k - \mu\zeta) = \sigma_{jk} + \mu^2 z - 2\mu$$

With respect to μ, these values are minimised when $\mu = 1/z$, giving

$$var(\epsilon_j - \mu\zeta) = \sigma_{jj} - {}^1/_z$$

$$covar(\epsilon_j - \mu\zeta, \epsilon_k - \mu\zeta) = \sigma_{jk} - {}^1/_z$$

In a homoskedastic model, σ_{jj} is constant and this value of μ therefore gives a limit to the variance in the model, below which it cannot be reduced. Choosing this specific value may not be suitable because it may give a singular matrix, but there is a smaller value of μ that will give a positive definite matrix, because there is a minimum value of k above which positive definiteness is guaranteed. In any case, we may wish to choose a smaller value, to ensure that the covariances are positive and therefore that the model can be approximated by a GEV model. Clearly, the covariances can be made non-negative by appropriate choice of μ, since μ^2 appears with a positive coefficient in the covariance formula above.

We conclude that it is possible to adjust a Probit model to be homoskedastic, non-degenerate and to have all of its covariances non-negative. Within those constraints, it may be possible to minimise the variance.

5. Summary of the main findings and research perspectives

This paper has primarily been motivated by the findings of various researchers that the GEV and IRUM paradigms overlap to a much greater extent than has previously been apparent in the literature. An important question suggested by these findings is how inherently homoskedastic and apparently closed-form GEV models can possibly mimic the wide range of potentially heteroskedastic and/or non-closed-form IRUM models.

The first conclusion of the paper is that heteroskedasticity across choice situations can often be eliminated simply by scaling the utilities. It then suffices to consider heteroskedasticity between alternatives. Moreover, the same simple scaling also brings some RUM models – where the only dependence of the random utility component on the non-random component is in adjusting the variance – into the IRUM class.

It is generally known that it is the matrix of utility differences (dimension $\frac{1}{2}J(J-1)$) rather than the matrix of utilities (dimension $\frac{1}{2}J(J+1)$) that is fundamental to RUM. Within the additional J dimensions of utility variation, effectively arbitrary, there may well be models that are homoskedastic, even if an initial model is heteroskedastic.

In the central result of the paper, the Homoskedasticity Theorem is proved: given an IRUM, it is always possible to construct a Probit IRUM that has a homoskedastic matrix of covariance of the alternative utilities and the same covariance matrix of utility differences as the original model. If the original model is Probit, the correspondence of the old and new models is exact, otherwise it may be approximate. The new model is constructed from the old model by manipulation of the Cholesky decomposition of the covariance matrix, the construction process ensuring that the new model is strictly real, eliminating the issue present in previous work by Daly (2001), but retaining the matrix structure of that work.

Relevant consequences of the Homoskedasticity Theorem are then discussed. First, the variance can be increased without limit, but there is a strict lower limit, which can be calculated based on constructing a matrix with minimal variance. Then, using the same mathematical approach, models are derived that have the same utility difference covariance matrix as any given IRUM, but are homoskedastic and have all their covariances positive, so that it is possible that closed-form GEV models could be found to approximate them. When the equivalence is exact, this means that homoskedastic models can be substituted for heteroskedastic models, without changing coefficient values, elasticities or willingness-to-pay estimates.

In summary, the importance of the Homoskedasticity Theorem is first to remove an important reservation concerning work such as Dagsvik (1995) and Daly (2007) who suggest GEV equivalence for many heteroskedastic models. This paper explains how such an equivalence can occur. Second, this paper gives further motivation to the search for explicit and preferably closed-form GEV models that are equivalent to or closely approximate IRUM, e.g. in the work of Marzano et al. (2013). Given the findings of the present paper, the target for that work can be extended from homoskedastic models to all or most IRUM. Should acceptable equivalent closed-form GEV models be found for IRUM, the gains to modelling practice would be substantial, because of the much greater speed associated with closed-form model calculations.

Finally, we hope to have advanced the understanding of choice modelling in a difficult area.

Acknowledgment

The authors would like to thank J. Nicolás Ibáñez for helpful comments and suggestions on earlier versions of the paper. The authors remain responsible for remaining errors and interpretations. The comments of three anonymous referees have also helped us to improve the clarity and precision of the paper.

This research was funded partially by a research grant from the University of Napoli "Federico II" (Italy) for a visit by the first author to the Institute for Transport Studies, University of Leeds (UK).

References

Ben-Akiva, M., Lerman, S.R. (1985) *Discrete choice analysis: theory and application to predict travel demand,* MIT Press, Cambridge, MA

Cascetta, E. (2009) *Transportation systems analysis: models and application*, 2nd Edition, Springer.

Daganzo, C. (1979) *Multinomial Probit - The Theory and its Application to Demand Forecasting*, New York: Academic Press.

Dagsvik, J.K. (1995) How large is the class of generalised extreme value models? **Journal of Mathematical Psychology**, 39, pp. 90-98.

Formatted: English (United Kingdom)

Daly, A. (2001) Independence, homoskedasticity and existence in random utility models. **ITS Working Paper** 557, Institute for Transport Studies, University of Leeds.

Daly, A. J. (2007) Properties of Random Utility Models of Consumer Choice, Ch. 14 of *Mathematics in Transport*, ed. B. Heydecker, Elsevier.

Daly, A.J. and Zachary, S. (1975) Commuters' Values of Time, **Report** T55, Local Government OR Unit, Reading, England.

Fosgerau M., McFadden D. and Bierlaire M. (2013) *Choice probability generating functions*, **The Journal of Choice Modelling**, vol. 8, 1-18.

Galambos, L. (1987) *Asymptotic theory of extreme order statistics,* Second edition, Wiley.

Ibáñez, J. N. (2007) On the compatibility of probabilistic choice systems with random utility maximisation. **ITS Working Paper** 592, Institute for Transport Studies, University of Leeds.

Joe, H. (2001) Multivariate extreme value distributions and coverage of ranking probabilities. **Journal of Mathematical Psychology**, 45, pp. 180-188.

Karlström, A. (2003) Developing Generalized Extreme Value Models Using the Pickands' Representation Theorem: theory and applications. **Working Paper**, Infrastructure and Planning, Royal Institute of Technology, Stockholm, Sweden.

Lindberg P.O., Eriksson E.A. and Mattsson L.-G. (1995) Invariance of achieved utility in random utility models. **Environment & Planning A,** 27(1), 121–142.

Marschak, J. (1960) Binary choice constraints and random utility indicators, in Marschak, J. (1974) *Economic Information, Decision and Prediction: Selected Essays* (Volume 1). D. Reidel, Dordrecht.

Marzano, V. and Papola, A. (2008) On the covariance structure of the Cross-Nested Logit Model. **Transportation Research B,** vol. 42, 83-98.

Marzano V., Papola A., Simonelli F. and Vitillo R. (2013) A practically tractable expression of the covariances of the Cross-Nested Logit model, **Transportation Research Part B-Methodological**, vol. 57, 1-11.

Mattsson L.-G., Weibull J.G., and Lindberg P.O. (2014) Extreme values, invariance and choice probabilities, **Transportation Research Part B-Methodological**, vol. 59, 81-95.

McFadden, D. (1978) Modelling the choice of residential location, in Karlqvist A. et al (eds.), *Spatial interaction theory and planning models*, North Holland, Amsterdam.

McFadden, D. (1981) Econometric models of probabilistic choice, in: Manski C., McFadden, D. (Eds.), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press Cambridge, MA, USA, 198–272 (Chapter 5).

Smith, T.E. (1984) A choice probability characterisation of generalised extreme value models, **Applied Mathematics and Computation**, 14, pp. 35-62.

Appendix

The linearity of the determinant of the matrix in k, stated in Section 4, may be derived by means of the Jacobi-Sylvester criterion. The criterion requires that, if $\boldsymbol{D}_i$ is the principal submatrix obtained from $\boldsymbol{\Sigma_h}$ taking into account the first i rows and columns, $\boldsymbol{\Sigma_h}$ is positive definite if and only if $\det(\boldsymbol{D}_i) > 0, \forall i \in 1 \dots J$. Note that, whatever i, the "structure" of each principal submatrix $\boldsymbol{D}_i$ is always equal to that of $\boldsymbol{\Sigma_h}$ given in section 4.1, i.e. with elements equal to k on the main diagonal and to $k - \delta_{ij}/2$ off the main diagonal. Moreover, given the utility difference variances δ, each $\det(\boldsymbol{D}_i)$ is a function of k only.

These findings allow us to state, as shown below, that for any $\boldsymbol{D}_i$ of this form, $\det(\boldsymbol{D}_i)$ can be expressed as $\det(\boldsymbol{D}_i) = a_i k_i + b_i$, where:

$$b_i = \det\begin{bmatrix} 0 & -\frac{\delta_{12}}{2} & \dots & -\frac{\delta_{1i}}{2} \\ -\frac{\delta_{12}}{2} & 0 & \dots & -\frac{\delta_{2i}}{2} \\ \dots & \dots & \dots & \dots \\ -\frac{\delta_{1i}}{2} & -\frac{\delta_{2i}}{2} & \dots & 0 \end{bmatrix} \quad \text{and} \quad a_i = \det\begin{bmatrix} 1 & 1-\frac{\delta_{12}}{2} & \dots & 1-\frac{\delta_{1i}}{2} \\ 1-\frac{\delta_{12}}{2} & 1 & \dots & 1-\frac{\delta_{2i}}{2} \\ \dots & \dots & \dots & \dots \\ 1-\frac{\delta_{1i}}{2} & 1-\frac{\delta_{2i}}{2} & \dots & 1 \end{bmatrix} - b_i$$

In other words, $\det(\boldsymbol{D}_i)\,[k]$ is a linear function of the homoskedasticity parameter k, so if $a_i \neq 0$, one and only one k exists such that $\det(\boldsymbol{D}_i) = 0$. Let us define k_i to be the k value such that $\det(\boldsymbol{D}_i)[k] = 0$. Moreover, as noted above, since the matrix is positive definite for $k(\Sigma)$, it is also positive definite for $k > k(\Sigma)$, i.e. a_i cannot be negative. This implies that if all the a_i are non-zero, then a positive definite solution exists for $k > \max\{k_i\}$ and this k then forms the lower limit for real models.

To prove this result, it is sufficient to take into account the following properties about determinants:

(a) $$\det\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \dots & a_{2n}+b_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} = \det\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} + \det\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

(b) $$\det\begin{bmatrix} ka_{11} & ka_{12} & \dots & ka_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} = k\det\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

(c) if $\boldsymbol{D}$ contains two or more linearly dependent rows, then $\det(\boldsymbol{D}_i) = 0$

Applying recursively the property (a) to the matrix $\boldsymbol{D}_i$ (or equivalently to $\boldsymbol{\Sigma_h}$ since they have the same structure), and using the properties (b) and (c), it follows that:

$$\det\Sigma_h = k\left(\det\begin{bmatrix} 1 & 1 & \dots & 1 \\ -\frac{\delta_{12}}{2} & 0 & \dots & -\frac{\delta_{2n}}{2} \\ \dots & \dots & \dots & \dots \\ -\frac{\delta_{1n}}{2} & -\frac{\delta_{2n}}{2} & \dots & 0 \end{bmatrix} + \det\begin{bmatrix} 0 & -\frac{\delta_{12}}{2} & \dots & -\frac{\delta_{1n}}{2} \\ 1 & 1 & \dots & 1 \\ \dots & \dots & \dots & \dots \\ -\frac{\delta_{1n}}{2} & -\frac{\delta_{2n}}{2} & \dots & 0 \end{bmatrix} + \dots + \det\begin{bmatrix} 0 & -\frac{\delta_{12}}{2} & \dots & -\frac{\delta_{1n}}{2} \\ -\frac{\delta_{12}}{2} & 0 & \dots & -\frac{\delta_{2n}}{2} \\ \dots & \dots & \dots & \dots \\ 1 & 1 & \dots & 1 \end{bmatrix}\right) +$$

$$+ \det\begin{bmatrix} 0 & -\frac{\delta_{12}}{2} & \dots & -\frac{\delta_{1n}}{2} \\ -\frac{\delta_{12}}{2} & 0 & \dots & -\frac{\delta_{2n}}{2} \\ \dots & \dots & \dots & \dots \\ -\frac{\delta_{1n}}{2} & -\frac{\delta_{2n}}{2} & \dots & 0 \end{bmatrix}$$

which can also be written as:

$$\det\Sigma_h = k\left(\det\begin{bmatrix} 1 & 1-\frac{\delta_{12}}{2} & \dots & 1-\frac{\delta_{1n}}{2} \\ 1-\frac{\delta_{12}}{2} & 1 & \dots & 1-\frac{\delta_{2n}}{2} \\ \dots & \dots & \dots & \dots \\ 1-\frac{\delta_{1n}}{2} & 1-\frac{\delta_{2n}}{2} & \dots & 1 \end{bmatrix} - \det\begin{bmatrix} 0 & -\frac{\delta_{12}}{2} & \dots & -\frac{\delta_{1n}}{2} \\ -\frac{\delta_{12}}{2} & 0 & \dots & -\frac{\delta_{2n}}{2} \\ \dots & \dots & \dots & \dots \\ -\frac{\delta_{1n}}{2} & -\frac{\delta_{2n}}{2} & \dots & 0 \end{bmatrix}\right) + \det\begin{bmatrix} 0 & -\frac{\delta_{12}}{2} & \dots & -\frac{\delta_{1n}}{2} \\ -\frac{\delta_{12}}{2} & 0 & \dots & -\frac{\delta_{2n}}{2} \\ \dots & \dots & \dots & \dots \\ -\frac{\delta_{1n}}{2} & -\frac{\delta_{2n}}{2} & \dots & 0 \end{bmatrix}$$

The preceding formula, which can be written in a compact notation as $\det(\boldsymbol{D}_i) = a_i k_i + b_i$, allows immediate calculation of the required k_i. Note that a_i and b_i depend on the elements of the starting matrix $\boldsymbol{\Sigma}$.