Transportation Research Part B 70 (2014) 151–162

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

A simple procedure for the calculation of the covariances of any Generalized Extreme Value model

Vittorio Marzano *

Dipartimento di Ingegneria Civile, Edile ed Ambientale (DICEA) – Università di Napoli "Federico II", Via Claudio 21, 80125 Napoli, Italy

ARTICLE INFO

Article history:
Received 30 December 2013
Received in revised form 10 July 2014
Accepted 25 August 2014

Keywords:
Generalized Extreme Value covariances
Network Generalized Extreme Value model
Cross-Nested Logit

ABSTRACT

This paper illustrates a simple procedure for calculating the covariances underlying any Generalized Extreme Value (GEV) model, based on an appropriate generalization of a result already established in the literature for the Cross-Nested Logit model (i.e. a particular GEV model). Specifically, the paper proves that the covariances in any GEV model are always expressed by a one-dimensional integral, whose integrand function is available in closed form as a function of the generating function of the GEV model. This integral may be simulated very easily with a parsimonious computational burden. Two practical examples are also presented. The first is an application to the CNL model, so as to check the consistency of the proposed method with the results already established in the literature. The second deals with the calculation of the covariances of the Network GEV (NGEV) model: notably, the NGEV is the most general type of GEV model available so far, and its covariances have not yet been calculated. On this basis, insights on the domain of the covariances reproduced by the NGEV model are also presented.

1. Introduction and background

Generalized Extreme Value (GEV) models (McFadden, 1978) represent a very powerful and practical class of discrete choice models, commonly and extensively applied to a wide range of transport-related modelling problems (Ben-Akiva and Lerman, 1985; Cascetta, 2009; Train, 2009).

In general, calculating the covariance matrix of the joint multivariate distribution of the perceived utilities[1] underlying a discrete choice model is crucial for supporting the analyst's modelling effort. First of all, knowledge of the covariances between alternatives enables a proper and conscious interpretation of model estimation results. Furthermore, choice contexts exist – such as route choice and activity-based modelling – where the analyst may formulate prior expectations on the covariances between alternatives, and therefore it would be desirable to specify a GEV model consistent with these expectations.

For such reasons, theoretical analysis of the covariance structure of GEV models and the corresponding search for practical tools for the calculation of their covariances are topics of great interest in the literature. The simplest GEV model able to account for non-zero covariances between alternatives is the Nested Logit (NL) model (Williams, 1977; Daly and Zachary, 1978), capable of reproducing only covariance structures represented by a tree. Importantly, a closed-form expression of NL covariances as a function of model parameters (Daganzo and Kusnic, 1993) is available, making its estimation and application very straightforward.

* Tel.: +39 081 7683935; fax: +39 081 7683946.
E-mail address: vmarzano@unina.it

[1] In the following, the more concise statement of "covariance between alternatives" will be adopted for the sake of simplicity.

http://dx.doi.org/10.1016/j.trb.2014.08.011
0191-2615/

In order to overcome the inherent limitations of the tree-structure of the NL covariances, more complex GEV models characterized by a network-based choice structure have been proposed, that is the Cross-Nested Logit (CNL) model (Small, 1987; Vovsha, 1997; Wen and Koppelman, 2001; Papola, 2004; Bierlaire, 2006; Marzano and Papola, 2008) and the Network GEV (NGEV) model (Daly, 2001; Daly and Bierlaire, 2006; Newman, 2008; Pinjari, 2011; Papola and Marzano, 2013). Unfortunately, both CNL and NGEV models are characterized by a non closed-form relationship between model parameters and corresponding covariances. As a consequence, the calculation of covariances in network-based GEV models is a very complex task, based on time-consuming numerical techniques also for rather simple CNL/NGEV structures.

In this respect, Papola (2004) proposed a closed-form approximation of CNL covariances, which however systematically overestimates the true CNL covariances, as proved by Abbe et al. (2007) and Marzano and Papola (2008). A recent improvement towards a more effective calculation of CNL covariances was provided by Marzano et al. (2013), who expressed the covariance between perceived utilities in a CNL model as a function of a one-dimensional integral. Even if this integral still does not have a closed-form primitive, its numerical calculation is very easy and not time-consuming. Indeed, the main proposition of the present paper, illustrated in Section 3, stems from proper generalization of the approach by Marzano et al. (2013). By contrast, covariances underlying the NGEV model have not yet been calculated, leading to a considerable missing step in the literature of GEV models.

In addition, it is worth recalling that the class of GEV models is very wide (Marzano and Daly, 2007; Mattsson et al., 2014) and also contains models which appear far from the CNL/NGEV framework: this is the case, for instance, of the model proposed by Karlström (2003). Again, calculating the covariances underlying these models is no trivial task and there are no studies available in the literature.

Starting from these premises, this paper generalizes the result by Marzano et al. (2013), proving that the covariances underlying any GEV model may always be expressed by means of a one-dimensional integral, very easily solvable with very limited computational effort. Therefore, thanks to this result, calculation of covariances in GEV models becomes a much easier task, with positive impacts for both researchers and practitioners. A noteworthy side result is that the covariances of the NGEV model may also be calculated very easily, as shown practically in the paper through a simple example. In addition, this result also makes it easy to carry out some research developments indicated by Marzano and Papola (2008), with reference to the characteristics of the domain of the covariance matrices reproduced by the NGEV model in contrast with the CNL model.

The paper is structured as follows: Section 2 provides the mathematical background for the calculation of covariances for the GEV models. Section 3 introduces the main mathematical results of the paper. Section 4 checks the consistency of the proposed procedure with results already established in the literature for the CNL model. Section 5 illustrates a practical application to the calculation of the covariances of the NGEV model, and provides insights on the characteristics of the domain of the covariances reproduced by the NGEV model. Finally, Section 6 draws conclusions and outlines developments for future research.

2. Calculation of covariances in the GEV model

The covariance $Cov[\varepsilon_i\varepsilon_j]$ between a pair of random residuals ε_i and ε_j in a discrete choice model, with joint cumulative distribution function (cdf) $F_{\varepsilon_i\varepsilon_j}(t_it_j)$ and joint probability density function (pdf) $f_{\varepsilon_i\varepsilon_j}(t_it_j)$, and with marginal cdf $F_{\varepsilon_i}(t_i)$ and $F_{\varepsilon_j}(t_j)$ respectively, can be expressed directly from the formal definition of covariance of a bivariate random variable:

$$Cov[\varepsilon_i\varepsilon_j] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} t_it_jf_{\varepsilon_i\varepsilon_j}(t_it_j)dt_idt_j - E[\varepsilon_i]E[\varepsilon_j] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} t_it_j\frac{\partial F_{\varepsilon_i\varepsilon_j}(t_it_j)}{\partial t_i\partial t_j}dt_idt_j - E[\varepsilon_i]E[\varepsilon_j] \quad (1)$$

Equivalently, a result by Hoeffding (1940) states that:

$$Cov[\varepsilon_i\varepsilon_j] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} [F_{\varepsilon_i\varepsilon_j}(t_it_j) - F_{\varepsilon_i}(t_i)F_{\varepsilon_j}(t_j)]dt_idt_j \quad (2)$$

It is worth noting that the above expressions do not depend upon any specific assumptions on the distribution of random residuals, i.e. they are valid for any discrete choice model. In the case of the GEV models, McFadden's (1978) theorem provides a direct expression of the joint cdf $F_{\varepsilon_1...\varepsilon_n}(t_1 \ldots t_n)$ of the random residuals $\varepsilon_1\ldots\varepsilon_n$ of a GEV model with n alternatives, whose generic perceived utility U_j is given by $U_j = V_j + \varepsilon_j$ where $V_j = E[U_j]$, as a function of a μ-homogeneous generating function $G(y_1, \ldots, y_n)$:

$$F_{\varepsilon_1...\varepsilon_n}(t_1 \ldots t_n) = e^{-G(e^{-t_1},\ldots,e^{-t_n})} \quad (3)$$

which is a Multivariate Extreme Value (MEV) distribution. From expression (3), the joint and marginal cdf and pdf of any subset of random residuals may be obtained, thanks to straightforward calculations of proper limits and of first derivatives. In this respect, McFadden's (1978) theorem implies the marginal distributions of the random residuals $\varepsilon_1\ldots\varepsilon_n$ to be Gumbel variables with same variance $\pi^2\theta_0^2/6$, $\theta_0 = 1/\mu$ being the reciprocal of the homogeneous degree μ of the generating function $G(y_1, \ldots, y_n)$.

Unfortunately, this approach leads to integrals (1) and/or (2) not in closed-form, meaning that numerical integration (i.e. simulation) is required. The computational effort required for this double numerical integration is very high even for CNL structures of limited complexity and practically not applicable to more complex GEV models, such as the NGEV model. Indeed, this paper introduces a different and much simpler mathematical approach, described in Section 3.