

Preface

Generalized estimating equations (GEE) were introduced by Liang and Zeger in a series of papers (see, e.g., Liang and Zeger, 1986; Zeger et al., 1985; Zeger and Liang, 1986) about 25 years ago. They have become increasingly popular in biometrical, econometrical, and psychometrical applications because they overcome the classical assumptions of statistics, i.e., independence and normality, which are too restrictive for many problems. The assumption of normality is, for example, violated if dichotomous data, e.g., positive or negative outcome, are considered, while the assumption of independence is not fulfilled in family studies or studies with repeated measurements. The development of more complex statistical methods like GEE is closely related to the progress in computer technology, because many modern regression approaches are based on iterative algorithms.

Originally, GEE have been proposed and further developed without considering them as special cases of quite general statistical methods. The main goal of this monograph therefore is to give a systematic presentation of the original GEE and some of its further developments. Subsequently, the emphasis is put on the unification of various GEE approaches. This is done by the use of two different estimation techniques, the pseudo maximum likelihood (PML) method and the generalized method of moments (GMM).

The PML approach was proposed by Gourieroux et al. in 1984b and further explained in additional work (Gourieroux and Monfort, 1993; Gourieroux et al., 1984a). The theory has been widely recognized by econometricians (see, e.g., Laroque and Salanie, 1989; Foncel et al., 2004) but to a lower extent by biostatisticians. A concise treatment of the PML theory has been given in Gourieroux and Monfort (1995a).

GMM was introduced by Hansen in 1982. It is very popular among econometricians because it provides a unified framework for the analysis of many well-known estimators, including least squares, instrumental variables (IV), maximum likelihood (ML), and PML. Several excellent book chapters and textbooks have been published (Hall, 1993; Ogaki, 1993), where readers may find many various introductory examples. As already indicated, the theory

of GMM is very rich, see, e.g., the two special issues on GMM published in *J Bus Econ Stat* (1996, Vol. 14, Issue 3; 2002, Vol. 20, Issue 4), much richer than the part of the theory required for deriving GEE. For example, one important part of GMM theory is IV estimation (Baum et al., 2003; Stock et al., 2002), which is of no interest in GEE. As a result, for GEE only, “just identified” GMM models (Hall, 1993) are relevant. However, if only just identified models are of prime importance, other aspects of GMM theory, such as choice of the weight matrix, or 1-step, 2-step, or simultaneous GMM estimation do not play a role. For this short book, it was therefore difficult to decide whether GMM should be described comprehensively or whether the treatise should be concise and focus only on the aspects relevant for deriving GEE. The decision was to restrict the description of GMM to essential elements so that GEE remains the focus. For detailed descriptions of GMM, the reader may refer to the literature (Hall, 2005; Mátyás, 1999).

To increase readability, regularity conditions and technical details are not given, and many proofs are only sketched. Instead, references to the relevant literature discussing technical details are given for the interested reader. GEE have been proposed as methods for large samples. Therefore, only asymptotic properties will be considered throughout this book for both PML and GMM estimation.

The main aim of this monograph is the statistical foundation of the GEE approach using more general estimation techniques. This book could therefore be used as a basis for a course for graduate students in statistics, biostatistics, or econometrics. Knowledge of ML estimation is required at a level as usually imparted in undergraduate courses.

Organization of the Book

Several estimation techniques provide a quite general framework and include the GEE as a special case. An appealing approach is to embed the GEE into the framework of PML estimation (Gourieroux et al., 1984b; Gourieroux and Monfort, 1993). If the GEE are embedded into the PML approach, they can be interpreted as score equations derived from a specific likelihood model. The major advantage of this approach is that ML estimation is familiar to almost every statistician.

The PML approach is based on the exponential family. Chaps. 1 and 2 therefore deal with the linear and quadratic exponential family, respectively. The GEE method has been derived by Liang and Zeger (1986) as a generalization of the generalized linear model (GLM; McCullagh and Nelder, 1989), and Chapter 3 therefore deals with both univariate and multivariate GLM.

Because PML estimation can be considered a generalization of the ML approach, ML estimation is discussed in some detail in Chapt. 4. A crucial assumption of the ML method is the correct specification of the likelihood

function. If it is misspecified, ML estimation may lead to invalid conclusions. The interpretation of ML estimators under misspecification and a test for detecting misspecifications are also considered in Chapt. 4.

Chapt. 5 deals with the PML method using the linear exponential family. It allows consistent estimation of the mean structure, even if the pre-specified covariance matrix is misspecified. Because the mean structure is consistently estimated, the approach is termed PML1. However, there is no free lunch, and a price has to be paid in terms of efficiency. Thus, a different covariance matrix, termed the robust covariance matrix, has to be used instead of the model-based covariance matrix, i.e., the Fisher information matrix. It can be shown that the robust covariance matrix always leads to an increased covariance matrix compared with the covariance matrix of the correctly specified model. Examples for the PML1 approach include the independence estimating equations (IEE) with covariance matrix equal to the identity matrix. Efficiency for estimating the mean structure may be improved if observations are weighted with fixed weights according to their degree of dependency. This approach results in the GEE for the mean structure (GEE1) with fixed covariance matrix.

Instead of using fixed weights, the weights might be estimated from the data. This results in increased power if the estimated covariance matrix is “closer” to the true covariance matrix than the pre-specified covariance matrix. This idea leads to the quasi generalized PML (QGPML) approach, which will be discussed in Chapt. 6. An important aspect is that under suitable regularity conditions, no adjustments have to be made for the extra variability, that is introduced by estimating the possibly misspecified covariance matrix. Even more, the QGPML estimator is efficient for the mean structure in the sense of Rao-Cramér if both the mean structure and the association structure, e.g., the covariance structure, are correctly specified. The likelihood function might thus be misspecified. Examples of QGPML estimation include the IEE with estimated variances, the GEE1 with estimated working covariance matrix, and the well-known GEE1 with estimated working correlation matrix. Examples for common weight matrices, i.e., working covariance and working correlation structures, are discussed as well as time dependent parameters and models for ordinal dependent variables.

Chapt. 7 deals with the consistent estimation of both the mean and the association structure using the PML approach. It is based on the quadratic exponential family and therefore termed the PML2 method. Special cases include the GEE2 for the mean and the correlation coefficient as the interpretable measure of association and, for dichotomous dependent variables, the GEE2 for the mean and the log odds ratio as the interpretable measure of association. In the first special case, the estimating equations are formulated in the second centered moments, while the second ordinary moments are used as the measure of association in the second special case.

The two GEE2 approaches considered in Chapt. 7 require the simultaneous solution of the estimating equations for the mean structure and the associa-

tion structure. This has three disadvantages. First, the computational effort is substantially larger when both estimating equations are solved jointly. Second, if the association structure is misspecified, the parameter estimates of the mean structure are still estimated consistently if the estimating equations are solved separately, i.e., in a two-stage approach. The simultaneous estimation of the estimating equations for the mean and the association structure may lead to biased parameter estimates if the mean structure is correctly specified but the association structure is misspecified. Therefore, one aim is to separate the estimating equations into a two-step approach. Third, GEE2 using the second standardized moments, i.e., the correlation coefficient as the measure of association, cannot be derived using the PML2 method.

All three disadvantages can be overcome by estimating equations that can be derived using GMM (Chapt. 8). Specifically, GMM allow the formulation of GEE2 using the correlation as the measure of association in two separate estimating equations. Similarly, the alternating logistic regression (ALR) that uses the log odds ratio through the ordinary second moments as the measure of association can be formulated as a special case of GMM. Therefore, GMM will be considered in the last chapter. The use of GMM is illustrated with the linear regression model as the introductory example. Second, the IEE are derived within the GMM framework. Third, the GEE2 using the correlation as the measure of association is considered. Finally, the ALR are derived as a special case of GMM. Again, we stress that only a small portion of the rich theory of GMM is required for deriving the GEE2 models, and we restrict the discussion of GMM to the needs in this monograph. Specifically, we do not consider IV estimation, and we focus on “just identified” models.



<http://www.springer.com/978-1-4614-0498-9>

Generalized Estimating Equations

Ziegler, A.

2011, XV, 144 p., Softcover

ISBN: 978-1-4614-0498-9