Preface

This special proceedings volume contains eight selected papers that were presented in the International Symposium in Statistics (ISS) 2015 on Advances in Parametric and Semi-parametric Analysis of Multivariate, Time Series, Spatial-Temporal, and Familial-Longitudinal Data, held in St. John's, Canada, from July 6 to 8, 2015. The main objective of the ISS-2015 was the discussion on advances and challenges in parametric and semi-parametric analysis for correlated data in both continuous and discrete setups. Thus, as a reflection of the theme of the symposium, the eight papers of this proceedings volume are presented in four parts: Part I—Elliptical *t* Distribution Theory; Part II—Spatial and/or Time Series Volatility Models with Applications; Part III—Longitudinal Multinomial Models in Parametric and Semi-parametric Setups; Part IV—An Extension of the GQL Estimation Approach for Longitudinal Data Analysis. The ISS-2015 was the continuation of ISS-2009 and ISS-2012 held in Memorial University. More specifically, the ISS-2009 was organized focussing *on inferences in generalized linear longitudinal mixed models (GLLMMs)*, and a special issue of the *Canadian Journal of Statistics* (2010, Vol. 38, June issue, John Wiley) was published with seven selected papers from this symposium. These seven papers from ISS-2009 dealt with progress and challenges in the areas of longitudinal and/or time series data analysis. As compared to ISS-2009, the papers in the ISS-2012 proceedings volume dealt with inferences for *longitudinal data with additional practical issues such as measurement errors, missing values, and/or outliers.* This proceedings volume was published as the Lecture Notes in Statistics (2013, Vol. 211, Springer) with nine selected papers from the symposium.

It is understood that the elliptical distributions have densities with equiprobable surfaces constant on homothetic ellipsoids, a property possessed in particular by the well-known and widely used multivariate normal distribution. Multivariate *t* distribution also belongs to this elliptic class of distributions, which however has symmetric but fatter tails as compared to the multivariate normal distributions. This additional tail characteristic makes the multivariate *t* distribution useful to analyze the heavy tailed such as stock return data one encounters under volatile financial markets, for example. However, unlike the multivariate normal sampling

theory, the elliptical t theories are not adequately discussed in the literature. In fact, unlike the normal sampling theory, the sampling theories for multivariate t responses can be quite different depending on whether the responses in a sample are independent or uncorrelated but dependent. The first paper in Part I, by B.C. Sutradhar, provides an insight on the advances and challenges in inferences for the heavy-tailed data that follow either independent or uncorrelated multivariate t distributions. The paper also proposes a clustered regression model where the multivariate t responses in a cluster are uncorrelated, but such clustered responses are collected from a large number of independent individuals. In the second paper in Part I, R. Prabhakar Rao, B.C. Sutradhar, and V.N. Pandit deal with correlated count and binary data in mixed model setup, where invisible random effects of the individuals are considered to follow independent t distributions. A part of this second paper was presented in the symposium by B.C. Sutradhar in his keynote address because of the common use of t distributions, but the materials of the paper were not included in the first paper in order to show the difference between the analysis of continuous t responses and discrete such as count and binary responses but influenced by continuous t random effects. The main challenge in the second paper is the estimation of the regression parameters when it is known that the t random effects with unknown degrees of freedom parameter cannot be integrated out from the model for any marginal estimation of such regression effects. In fact this problem is also encountered in binary and count data analysis with Gaussian random effects where suitable simulation techniques are used by generating standard Gaussian random effects for the estimation of the parameters including the variance of the normal Gaussian random effects. In the current paper, a similar simulation technique is used where t random effects are generated with unit scale and four degrees of freedom, and a transformation is proposed for the purpose of estimation of the unknown degrees of freedom parameter along with regression and other such as correlation parameters.

The Part II of the volume contains two papers on spatial and temporal data analysis. The first paper by L.M. Ainsworth, C.B. Dean, and R. Joy is an application paper analyzing spatial counts with a significant portion of responses being zero. This type of zero-inflated count data can provide important clues to physical characteristics associated with, for example, habitat suitability or resistance to disease or pest infestations. However, the probability modeling for this type of spatial bimodal data especially after accommodating spatial correlations is not easy. The authors have considered various existing models and used their expert knowledge to examine what and how these models are doing in order to understand, for example, the white pine weevil infestations data, where many trees did not exhibit any weevil attack or infestation. These models can be grouped into two categories. Some models are appropriate for independent spatial counts with over-dispersion generated by inflated zeros. The rest of the zero-inflated probability models accommodate spatial correlations among counts through correlated random effects. The second paper in Part II contributed by V. Tagore, N. Zheng, and B.C. Sutradhar deals with a special temporal data where the variance of the response data appears to change over time. These heteroscedastic variances are then explained

through a suitable dynamic model, and it is of interest to obtain consistent estimates of the parameters involved in such a dynamic model along with consistent estimates for any regression parameters relating time-dependent covariates and the responses. This type of time series models produces larger kurtosis for the data as compared to the usual Gaussian time series with constant variance over time. Consequently these models are found to be suitable to explain the volatility in the data which is, quite often, exhibited in financial markets such as stock return data. The main contribution of the paper is the development of a simpler method of moments technique for consistent estimation of the parameters of the model, as compared to the existing QML (quasi-maximum likelihood) and the so-called popular but very lengthy and complex GMM (generalized method of moments) approaches. The authors have demonstrated the advantage of their estimation technique by reanalyzing the well-known US dollar and Swiss franc exchange rate data to understand the presence of any possible volatility.

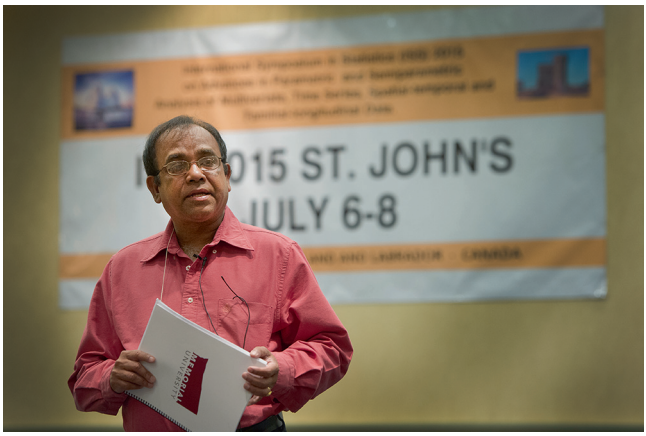
Binary dynamic mixed logit (BDML) models are used to fit longitudinal binary data collected from the members of a large number of independent families. In the first paper of Part III, B.C. Sutradhar, R. Viveros-Aguilera, and T. Mallick have provided a generalization of the BDML model to the categorical data setup, binary case being a special case with two categories. Similar to the BDML model, this MDML (multinomial dynamic mixed logit) model uses parametric correlation structure for repeated multinomial data. More specifically, the authors have used dynamic dependence of the current multinomial response on a past response to model the correlations. The regression and dynamic dependence parameters of the model have been estimated by using the likelihood estimation approach. The variance component of the random effects is also estimated by using the likelihood approach. For the binary case, these three parameters, namely, the regression effects, dynamic dependence, and variance component parameters, may be estimated conveniently by using the GQL (generalized quasi-likelihood) approach. The authors have conducted an empirical study to examine the relative performance of the GQL and likelihood estimates for the parameters of a BDML model. The MDL (multinomial dynamic logit) model has been applied as an illustration to analyze a real-life longitudinal categorical data with three categories. The second paper in Part III deals with ordinal categorical data, whereas the first paper was confined to the nominal categorical/multinomial data. This paper by B.C. Sutradhar and N. Dasgupta discusses two correlation models for the ordinal multinomial data. The first model is constructed by cumulating the probabilities for the nominal repeated multinomial responses. The well-known likelihood estimation approach is used to compute the estimates of the parameters of the cumulative model. The second model is constructed in a completely different way than the first model. This model is developed by assuming that the responses at a given time point are available in a cumulative form so that they follow a binary distribution with so-called binary logistic probabilities. To accommodate the correlations for these repeated cumulative responses, a BDL (a binary dynamic logit) model is written involving dynamic dependence between two binary responses. Next, the parameters of this BDL model are estimated by forming a pseudo-likelihood

for all possible lag 1 transitional binary probabilities. The pseudo-likelihood estimating equations are provided for all regression and binary dynamic dependence parameters. In the third paper, B.C. Sutradhar considers nominal multinomial variables which is similar to that of the first paper. However, unlike the first paper, this paper developed a semi-parametric probability model for longitudinal multinomial responses. To be specific, to construct such a semi-parametric model, first it is assumed that the traditional specified parametric regression function is not enough to explain the multinomial probabilities. Consequently a non-parametric function is added to the specified regression function which yields a semi-parametric regression function for the construction of the desired multinomial probabilities. The dynamic dependence part remains the same as in the first paper by Sutradhar, Viveros-Aguilera, and Mallick. This type of longitudinal semi-parametric model for multinomial responses is not adequately addressed in the literature. To make this semi-parametric model easily understandable, the author presents a sequence of longitudinal semi-parametric models for repeated linear and count data which are already discussed in the literature to a reasonably good extent. For all these models including the proposed semi-parametric models for the repeated multinomial data, the paper developed suitable estimating equations for the non-parametric function and all other parameters, namely, the regression and the dynamic dependence parameters. These estimation formulas should be useful for any empirical study to analyze repeated multinomial data in a semi-parametric setup.

Part IV of the volume contains one paper by T. Nadarajah, A.M. Variyath, and J.C. Loredo-Osti on the inferences for longitudinal data subject to a challenge of important covariates selection from a set of large number of covariates available for the individuals in the study. The inference technique uses an idea of penalization and estimates the regression parameters corresponding to all covariates involved in a related penalized generalized estimating function, where this later function is constructed by modifying a generalized quasi-likelihood (GQL) estimating function. Any regression parameter estimate close to zero obtained by solving the penalized GQL (PGQL) estimating equation automatically removes the unimportant covariates from the model, leading to a reduced model with important covariates for interpretation. The authors of the paper have verified the performance of this PGQL inference technique through an intensive simulation study. A data analysis is also provided justifying the technique.

St. John's, NL, Canada Brajendra C. Sutradhar

ISS-2015 delegates



ISS-2015 welcome address by Brajendra Sutradhar (Organizer)

Further to the welcome by Professor Charmaine Dean [former president of the SSC (Statistical Society of Canada) and the current dean of science of the University of Western Ontario] and Dr. Alwell Oyet [deputy head of the Department of Mathematics and Statistics at Memorial University], once again I welcome all of you with the name of the Lord to this International Symposium in Statistics 2015 (ISS-2015) on Advances in Parametric and Semi-parametric Analysis of Multivariate, Time Series, Spatial-Temporal, and Familial-Longitudinal Data. As noted in the symposium web site (http://www.iss-2015-stjohns.ca./), the ISS-2015 is the continuation of ISS-2009 and ISS-2012. Both of the last two symposiums were held in Memorial University, Canada, and they were devoted to the discussion of progresses and challenges in the analysis of longitudinal data subject to measurement errors, missing values, and/or outliers. These two symposiums were highly successful with two high-quality proceedings volumes, one in the form of a special issue of the *Canadian Journal of Statistics* in 2010 and the other as a Springer Lecture Notes in Statistics in 2013. It is my pleasure to note that we have been able to keep up the spirit of the last two symposiums in organizing the discussion topics of the present symposium covering the progress and advances in correlated data analysis in a variety of setups, such as spatial and/or temporal setup, semi-parametric setup for discrete longitudinal data, and multivariate setups for discrete familial-longitudinal and continuous non-Gaussian elliptical data.

I am very grateful to Bhagawan Sri Sathya Sai Baba, my guru, the universal spiritual master, for his blessings and inspirations in organizing these international community services. I am also thankful to all of you for your interest and response to this 2015 symposium that has attempted to attract the noble group of researchers including graduate students. I hope that you will find the symposium stimulating and will derive spirits for doing more quality research in these challenging areas as a service to the society and mankind at large.

As far as the presentation structure of this symposium is concerned, four keynote speeches are organized in four different areas to be delivered by three

speakers. Professor Anthony C. Davison from EPFL, Switzerland, will give his keynote address on max-stable processes on river networks, under the theme of spatial-temporal data analysis. Professor Brajendra C. Sutradhar from Memorial University, Canada, will deliver part 1 of his keynote presentation on advances and challenges in correlated data analysis in non-Gaussian multivariate setup and part 2 of the presentation on advances and challenges in analyzing ordinal categorical data in semi-parametric setup. Part 3 of the keynote address will be given by Professor Andrew Harvey from Cambridge University, UK, on new developments in modeling dynamic volatility. Nine special invited talks over 3 days of the symposium will be given by Professors Paul D. Sampson, University of Washington; Grace Y. Yi, University of Waterloo; Nairanjana Dasgupta, Washington State University; Roman Viveros-Aguilera, McMaster University; Julio M. Singer, Universidade de Sao Paulo; David E. Tyler, Rutgers—the State University of New Jersey; Refiq Soyer, George Washington University, Charmaine Dean; University of Western Ontario; and Richard J. Cook, University of Waterloo. The symposium has another two invited speakers, Dr. Alwell Oyet from Memorial University and Dr. Ashis SenGupta from Indian Statistical Institute. Also, contributed papers will be presented by seven speakers including four graduate students. Furthermore, it is planned that a selected number of papers presented in the symposium will be published in the near future as lecture notes in the Springer's Lecture Note Series.

It is also a pleasure to note that we have 46 delegates in this specialized symposium from many countries such as Brazil, France, India, Switzerland, the USA, and Canada, covering a large part of the globe. The organizing committees would like to extend a hearty welcome to all of you including all graduate students.

We also welcome you to St. John's, the oldest city of North America, known as the City of Legends, where you can view icebergs, watch whales, and experience Newfoundland and Labrador's unique culture. It is a progressive city and is the site of many world-class facilities including an international center in marine science and technology. A mosaic of fishing villages, cultural festivals, and wildlife tours bring variety to the city. Also, Cape Spear, the most easterly point of North America, is not far from the city, where one can experience the unique beauty of sunrise. We hope that you have planned for an extended stay in St. John's following the symposium to enjoy these and other endless options!

http://www.springer.com/978-3-319-31258-3

Advances and Challenges in Parametric and Semi-parametric Analysis for Correlated Data
Proceedings of the 2015 International Symposium in Statistics
Sutradhar, B.C. (Ed.)
2016, XIX, 256 p. 12 illus., 6 illus. in color., Softcover
ISBN: 978-3-319-31258-3