PROGRAMME AND ABSTRACTS

5th CSDA International Conference on

Computational and Financial Econometrics (CFE 2011)

http://www.cfe-csda.org/cfe11

and

4th International Conference of the
ERCIM (European Research Consortium for Informatics and Mathematics) Working Group on

Computing & Statistics (ERCIM 2011)

http://www.cfe-csda.org/ercim11

Senate House, University of London, UK
17-19 December 2011

ERCIM
European Research Consortium
for Informatics and Mathematics

Queen Mary University of London
http://www.qmul.ac.uk

Birkbeck UNIVERSITY OF LONDON
http://www.bbk.ac.uk

LSE THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE
http://www.lse.ac.uk

penalty. We considered a variety of scenarios including cases where $2^p \gg n$. Finally, we analyse some data, report our findings and discuss the key issues.

E443: **Optimal choice of reference subclass in categorical regression models**
Presenter: **Defen Peng**, University of Limerick, Ireland
Co-authors: Gilbert MacKenzie

Issues surrounding the choice of reference subclasses for categorical variables in parametric regression models are addressed. We focus on: (a) techniques which maximize the precision of the resulting estimators, (b) discrepancy from the ideal allocation and (c) multi-collinearity. First, we derive the optimal design allocation and provide a statistic, based on generalized variance and its distribution for measuring the discrepancy between the optimal allocation and the observed allocations occurring in observational studies. Next we focus on reference category subclass choice, obtaining general expressions for the variance-covariance matrix of the estimators in linear and other generalized linear models for a single categorical variable. Then we extend the investigation to multiple categorical variables by means of simulation studies, wherein we demonstrate the use of the techniques. Later we address the issue of multi-collinearity which is especially important when the regression model contains multiple categorical variables with sparse subclasses. In this case the model design matrix may become unstable or even less than full rank. To begin, we consider the case of linear model with a single categorical variable and rapidly extend our investigation into GLMs with multiple categorical variables and end by drawing some general conclusions about the various methods developed.

ES33 Room S261 ROBUST METHODS IN SMALL AREA ESTIMATION Chair: Isabel Molina

E116: **Controlling the bias of robust small area predictors**
Presenter: **David Haziza**, University of Montreal, Canada
Co-authors: Valery Dongmo Jiongo, Pierre Duchesne

The user demand for small area estimators has been growing in most countries. This led survey statisticians to develop theoretically sound and yet practical estimation procedures, providing reliable estimators for small areas. A popular estimation method is the so-called empirical best linear unbiased prediction (EBLUP). However, the EBLUP is sensitive to the presence of outliers. That is, including or excluding outlying units from its computation may have a large impact on its magnitude. In recent years, the problem of robust small area estimation has received considerable interest. We introduce two new robust small area estimators that are robust in the presence of outliers. Results of a simulation study that compares the performance of several robust estimators in terms of relative bias and relative efficiency will be presented.

E425: **On computing and tuning some simple and robust unit-level SAE estimators**
Presenter: **Tobias Schoch**, University of Applied Sciences Northwestern Switzerland, Switzerland
Co-authors: Beat Hulliger

The need for robust estimates of area-level characteristics is of increasing importance for production in Official Statistics (OS), since of EBLUP methods are vulnerable to outliers. Statisticians in OS often fit mixed linear models (MLM) and use EBLUP methods for Small Area Estimation (SAE). The robust EBLUP methods, e.g. M-estimator of an MLM's core parameters (called RML), as well as the robust prediction of random effects and area means, however, require more attention to the (non-trivial) issue of robustness-tuning and to computational aspects in general. In the context of an RML fitting exercise, users typically attempt to refer to the well-known robustness properties of simple location-scale models (such as the fact that the M-estimator of location with Huber psi-function obtains a relative asymptotic efficiency of approx. 95% w.r.t. maximum likelihood estimator at the true model if the tuning constant is 1.345). However, these properties do not always directly carry over to the RML method. As a consequence, the choice of how best to tune RML-estimators is much more involved. A simulation study of the computational aspects (incl. breakdown-point considerations) of robust estimators for unit-level SAE models is carried out and recommendations concerning the choice of robustness tuning constants are made.

E466: **Robustness analysis of unbalanced linear mixed modeling**
Presenter: **Roland Fried**, TU Dortmund University, Germany
Co-authors: Isabel Molina Peralta, Betsabe Perez Garrido, Anita Thieler

Robust approaches for fitting linear mixed models are investigated and compared. Starting from the simplest situation of fitting a constant mean μ in the presence of independent random group effects u_i and independent random errors e_{ij}, $Y_{ij} = \mu + u_i + e_{ij}$, with $i = 1,\ldots,I$ being the group index and $j = 1,\ldots,n_i$ denoting the individuals, we subsequently elaborate more complicated models with covariates. We start from the simple model given above, since methods which do not work there will hardly be useful in more complex situations. While balanced scenarios with equal group sizes $n_1 = \ldots = n_I$ have been analyzed frequently in the past, unbalanced designs have been given less attention so far. In small area estimation we are usually confronted with unbalanced designs and want to protect against the possibility of outlying individuals or groups. Our analysis aims at identifying which methods are useful under which scenarios and under which type of contamination.

E712: **Outlier robust domain estimation for business survey data**
Presenter: **Nikos Tzavidis**, University of Southampton, United Kingdom
Co-authors: Sabine Krieg, Marc Smeets, Chiara Bocci, Virginie Blaess

Outliers are a fact of life for any survey, and especially so for business surveys. If outliers are a concern for estimation of population quantities, it is safe to say that they are even more of a concern in small area estimation, where sample sizes are considerably smaller and model-dependent estimation is the norm. It is therefore of interest to see how outlier robust survey estimation can be adapted to this situation. The development of outlier robust small area methodologies has been the focus of recent small area literature. We review a range of outlier robust small area methodologies and apply these to business survey data from the Netherlands. We discuss both point and Mean Squared Error estimation (MSE) using analytic and bootstrap-type MSE estimators. Finally, we place some emphasis on providing practical guidelines to the survey practitioner for working with outlier robust small area methodologies.

ES36 Room Jessel PARAMETRIC AND SEMIPARAMETRIC HAZARDS MODELS AND ANALYSES Chair: M. Carmen Pardo

E069: **Inference for a semiparametric generalized logit-based proportional hazards model in survival analysis**
Presenter: **Martha Avendano**, Complutense University of Madrid, Spain
Co-authors: Maria del Carmen Pardo, Narayanaswamy Balakrishnan

A generalized logit-based proportional hazards model of survival analysis is introduced as an alternative to the popular Cox proportional-hazards model. Our model is based on a generalization of the logistic distribution and it is semiparametric, since it can be factored into a parametric part consisting of a regression parameter vector associated with the covariates and a non-parametric part that can be left completely unspecified. The partial loglikelihood approach can be applied to estimate the model parameters. The asymptotic properties of the parameter estimates are established. We compare our models with the Cox proportional hazards model by using one example.