INSTITUT DE
STATISTIQUE

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

DISCUSSION
PAPER

0920

REJOINDER ON: A REVIEW ON EMPIRICAL LIKELIHOOD METHODS FOR REGRESSION

CHEN S.X. and I. VAN KEILEGOM

Noname manuscript No.
(will be inserted by the editor)

Rejoinder on: A Review on Empirical Likelihood Methods for Regression

Song Xi Chen · Ingrid Van Keilegom

Received: date / Accepted: date

1 Introduction

We would first of all like to thank the two editors Ricardo Cao and Domingo Morales for giving us the opportunity to write this review paper. Furthermore, we would also like to take this opportunity to express our thanks to all discussants for their valuable comments, helpful suggestions, insightful ideas and stimulating discussions. We hope and believe that their input will lead to more interest and research in the area.

This rejoinder is organized as follows. In the next section we discuss some computational issues of the empirical likelihood method. The important problem of bandwidth selection, linked to the use of bootstrap procedures and Edgeworth expansions is discussed in Section 3. An often neglected problem is the choice of the criterion function in the construction of the empirical likelihood. This problem is discussed in Section 4. Finally, Sections 5, 6 and 7 are devoted to three types of complexities in the data : dependent data, censored data and high-dimensional data.

2 Computational cost

Professor Sperlich has raised some important practical issues encountered with empirical likelihood methods, which can be summarized as follows : computational aspects, smoothing parameter selection in semiparametric and nonparametric setting, reliability of the asymptotic approximations, handling of misspecified models and extensions

S.X. Chen
Department of Statistics, Iowa State University, Ames, Iowa 50011-1210, USA and
Guanghua School of Management, Peking University, China
Tel.: 1-515-2942729
Fax: 1-515-2944040
E-mail: songchen@iastate.edu; csx@gsm.pku.edu.cn

I. Van Keilegom
Institute of Statistics, Université catholique de Louvain, Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium
Tel.: +32 10 47 43 30
Fax : +32 10 47 30 32
E-mail: ingrid.vankeilegom@uclouvain.be

to different types of covariates. These are all important issues in applications and some of them in theoretical developments of the empirical likelihood as well.

The intensive computational character is a natural property of the empirical likelihood, a feature that is given by birth from its definition. It can be viewed as a price paid to attain some of the attractive properties of the empirical likelihood, for instance the confidence regions having natural shape and orientation and respecting the range of the parameter, and the elegant Bartlett correction. The latter draws empirical likelihood much closer to the conventional likelihood. However, if these properties are not regarded as important, for instance in goodness-of-fit tests, less computational versions of the empirical likelihood may be used. These include the "least squares empirical likelihood" (LSEL) (Owen, 1991 and Brown and Chen 1998), which is called "Euclidean likelihood" in Owen (1991). The LSEL is much easier to compute as it replaces $-\sum_{i=1}^{n}\log(np_i)$ with $\sum_{i=1}^{n}(p_i-1/n)^2$ as the objective function when evaluating the empirical likelihood ratio. See Baggerly (1998) for other alternative versions of empirical likelihood within the Cressie-Read discrepency measures, which share Wilks' theorem with the empirical likelihood but not the Bartlett properties. Indeed, in some implementations of empirical likelihood applications (Chen, Gao and Tang, 2008), computationally less expensive versions of the empirical likelihood have been employed. This discussion is also connected to the comments by Professors Peng and Zhang on nuisance parameters, as typically more computational power is required when dealing with nuisance parameters.

3 Bandwidth selection, bootstrap and Edgeworth expansions

We agree with Professor Sperlich that for complex inference problems like goodness-of-fit tests, which involve either semiparametric or nonparametric smoothing, choosing the smoothing bandwidth is an important and yet challenging task for empirical likelihood. We would think the same is true for statistics formulated based on measures other than empirical likelihood. Of course the issue may be less in magnitude if that measure is easier to compute than empirical likelihood. The challenge is partly due to the fact that the role of the bandwidths on the inference statistics (empirical likelihood or other) is largely in the second order asymptotically. Being the second order brings little comfort for us as we work on finite samples. However, it does imply that if we want to select the bandwidth objectively, we have to develop expressions for the second order terms. For simplier inference problems, like confidence intervals and hypothesis testing for nonparametric density and regression functions, the second order terms that describe the coverage errors or the second order power can be quantified by developing Edgeworth expansions, which can be used for choosing the bandwidth. There have been some works in the literature for the empirical likelihood as cited in the main paper. If such a second order quantification is available we can choose the bandwidth directly based on the second order expression or use the bootstrap method as outlined in Hall (1992). For more complex inference problems like goodness-of-fit tests, the bootstrap method has already been used to provide first order approximations to distributions of test statistics formulated by the empirical likelihood or its less expensive cousines (Chen, Gao and Tang, 2008; Chen and Van Keilegom, 2009). To justify the use of the bootstrap method for the second order bandwidth selection, more research is needed to establish second order expansions to the inference statistics. We note here that both the bootstrap and the empirical likelihood are computationally intensive, and

marrying the two together would make the computation even more expensive. This puts more incentive to use some of the less expensive versions of the empirical likelihood that have the property of internal studentization of the full empirical likelihood before more computational capacity is available. The use of splines as suggested by Professor Sperlich is another option one may use to reduce the computation.

On the other points raised by Professor Sperlich, we note that the accuracy of the asymptotic approximation is generally less for a more complex inference problem like the goodness-of-fit test than for a simpler problem; see Chen, Härdle and Li (2003) and Chen and Van Keilegom (2009) for details. This leads to the need for bootstrap calibration. By the way, when a parametric model is misspecified, the empirical likelihood will be reacting with a big value which indicates the data are not comfortable. This is the rational for using it for goodness-of-fit statistics.

The first order theory for the empirical likelihood, in the context of regression or any other context, only assumes the existence of certain moments of the random variables. So, it is generally applicable to continuous and categorical variables. For the second order theory, we need to assume continuous random variables to make the Edgeworth expansions manageable, although it can be done for lattice valued random variables with some extra adjustments. However, for fixed design regression, the nature of the covariates does not matter, regardless the first or second order. What matters is the type of the distribution for the residuals. We agree that empirical likelihood for mixed effect models with possible endogeneity is a topic that requires further research. There is no doubt that heteroscedasticity has an impact on empirical likelihood for regression. However, it has no impact on the limiting distribution of the empirical likelihood, but rather on the length/area of the confidence region. This is due to the ability of the empirical likelihood to self-studentize. Direct modeling of the heteroscedasticity, as suggested by Professor Sperlich will bring more condensed confidence regions just like such effort would produce more efficient estimators (Carroll, 1982).

4 Choice of the criterion function

Professors Peng and Zhang discuss three ways to deal with nonlinear constraints, i.e. with situations where the criterion function in the numerator of the EL ratio is nonlinear. We agree with each of these three solutions, and believe that they are valuable methods, all having their merits and drawbacks depending on the particular situation at hand. In addition to these three proposals, we would like to add a fourth method, which works as follows. It is clear that the choice of the criterion function has a major effect on the limiting distribution of the EL ratio. For example, as Peng and Zhang indicate, the standard EL method for the variance leads to a weighted sum of chi-square variables, whereas a profile EL method gives rise to an unweighted chi-square limit. Another, more sophisticated example can be found in Section 3.6 in Hjort, McKeague and Van Keilegom (2009), which deals with EL inference for the survival function from current status data. Because of the complicated data structure, it does not seem possible to come up with a criterion function that does not depend on unknown nuisance functions. Hence, it seems unavoidable to work with an EL ratio that contains estimated nuisance functions. However, by choosing the criterion function in a 'clever' way, Hjort, McKeague and Van Keilegom (2009) showed that the EL-statistic has an unweighted chi-square limit, i.e. Wilks' theorem holds. See Zhu and Xue (2006) for another ex-

ample where estimated nuisance functions are involved in the criterion function, and where Wilks' theorem is nevertheless valid.

More generally, one can wonder how the criterion function should be chosen in order to obtain standard chi-square limits, even if unknown nuisance parameters or functions are floating around. This is, as far as we know, not studied in a general framework in the literature on empirical likelihood, and seems to us a challenging but very important problem.

5 Dependent data

Professor Velasco gives a very nice overview on how to deal with time series data in the context of empirical likelihood. He explains three methods that have been proposed in the literature : EL applied to data blocks, EL based on kernel smoothed moment conditions, and a frequency domain EL method. As he mentions it is possible that these methods can be used in other non- and semiparametric contexts involving dependent data. One (among many) of these contexts is the situation where data are time dependent and subject to censoring. This type of data is encountered e.g. in studies on duration of unemployment, where the data may be right censored and are typically correlated. El Ghouch, Van Keilegom and McKeague (2009) develop three types of confidence intervals for a general class of functionals of a survival distribution based on censored dependent data. The confidence intervals are constructed via asymptotic normality, the classical EL method and the blockwise EL method. It is worth noting that in this case the limiting distribution of the EL statistic is that of a weighted chi-square variable, even when blocking is used. This is caused by the fact that the data are subject to censoring. Hence, one may wonder whether blocking is really helpful in this case. Simulations show that the blockwise EL procedure (slightly) outperforms the classical EL method, provided the block length is chosen in an appropriate way. So, blocking does improve the performance, but the method requires more computational power than the classical EL procedure in order to choose the block length in a data-driven, optimal way.

6 Censored data

In this section we like to come back to the remarks that have been made in the context of censored data. First of all, when we refer to parametric regression (respectively semiparametric or nonparametric regression) we mean that the regression function is modeled in a parametric way (respectively semiparametric or nonparametric), even when the error distribution is modeled completely nonparametrically.

Professors Li and Lu discuss the estimation in parametric regression, when the response is subject to random right censoring. They compare the estimators proposed by Koul, Susarla and van Ryzin (1981) (KSV hereafter) and Buckley and James (1979) (BJ hereafter) through simulations. They do not incorporate these two methods in an EL context, but compare instead the performance of the original estimators. Their main remark is that the KSV method requires the independence between the censoring variable C and the covariate X, whereas this is not the case for the BJ method. The simulations show that the violation of this assumption can lead to serious bias and variance of the KSV estimator. Although we completely agree with this observation

and with the conclusions of the simulation study, we would like to point out that the assumption of independence between C and X can be avoided by replacing the marginal distribition $G(\cdot)$ by the conditional distribution $G(\cdot|X_i)$ of C given $X = X_i$ in the definition of the synthetic variable Y_{iG}. The estimation of this conditional distribution will require smoothing techniques. See e.g. the local Kaplan-Meier (1958) estimator of Beran (1981), which requires the choice of an appropriate bandwidth. The effect of this bandwidth on the behavior of the so-obtained adapted KSV estimator will however be of second order, as the local Beran estimators are averaged out over X in the formula of the least squares estimator of β.

Professors Li and Lu also give an interesting discussion on the pros and contras of a censored data EL procedure compared to a complete data EL. We like to add to this discussion that the latter procedure usually does not lead to an unweighted chi-square limiting distribution of the EL statistic (not only in the current context of parametric censored regression, but more generally). Although, as Liang Peng and Rongmao Zhang nicely point out, there exist ways out in this case by estimating e.g. the weights, our experience shows that in many situations the EL looses somewhat its main attractive features in practice when the limiting distribution is not an unweighted chi-square distribution.

In his discussion, Professor Sánchez-Sellero also comes back to the choice between a complete data and a censored data EL procedure. He proposes a very interesting idea, which consists in applying a censored data EL procedure to the defining equation of the estimator of the regression coefficients in censored regression studied in Stute (1999). His idea, which as far as we know has not been proposed in the literature, seems very promising, since it combines the nice properties of a censored data likelihood with the stability and good practical performance of the estimator of Stute (1999), provided the distributional assumptions mentioned by Sánchez-Sellero are met.

As Professor Sánchez-Sellero points out, many papers dealing with EL for regression with incomplete data restrict attention to missing and/or right censored data. In fact, not much research has been done in the context of EL for regression with biased data (like length biased or truncated data) or with more complicated censoring schemes. Recently, Bertail (2006) developed a quite general EL procedure for semiparametric models, and applied his procedure to EL for general functionals in biased sampling models, while Ren (2008) studied a weighted EL procedure for some two-sample regression models when the data are subject to various types of censoring (such as right censoring, double censoring or interval censoring) or when they are generated from various biased sampling schemes. It is clear however that this research area is still relatively under-explored and that more research is needed to handle the many types of incomplete data that are encountered in practice.

Regarding a query by Professor Sánchez-Sellero on confidence bands for regression curves, we would like to mention the paper by Hall and Owen (1993) in the context of density functions. It can be readily extended to regression.

7 High-dimensional data

Professor McKeague is correct in observing that the current review has been confined to data (covariates) of fixed dimensions. Indeed, this was based on a consideration that the field of empirical likelihood with high-dimensional data is still being developed, and a belief that we should wait for a while for the field to mature. There have

been works on empirical likelihood for high-dimensional data. Hjort, McKeague and Van Keilegom (2009) have a section on the empirical likelihood for a high-dimensional mean parameter $\mu \in R^p$, where μ is the common mean of independent and identically distributed $X_1, \cdots, X_n$, and the dimensionality $p \to \infty$ as $n \to \infty$. They showed that the empirical likelihood ratio statistic is asymptotically normal with mean p and variance $2p$ if $p = o(n^{1/3})$ together with some other conditions. The asymptotic normality of the empirical likelihood mirrors Wilks' theorem when p is finite. Chen, Peng and Qin (2009) provided a more refined analysis and showed that the growth rate for p can reach $o(n^{1/2})$, which is likely to be the best rate for the asymptotic normality of empirical likelihood for means. The machineries used in Hjort, McKeague and Van Keilegom (2009) and Chen, Peng and Qin (2009) can be used for regression with growing number of covariates to establish asymptotic normality. Using the empirical likelihood to replace the parametric likelihood or the least squares goodness-of-fit in the LASSO formulation is a real possibility, and some of the advantages of the empirical likelihood could be utilized in this context. Shahidi (2009) considered empirical likelihood with a penalty for a fixed dimensional problem. Tang and Len (2009) considered an empirical likelihood with a LASSO penalty for growing dimension. This together with the direction pointed out by Professor McKeague on "targeted maximum likelihood learning" represent some of the new directions on empirical likelihood that seem promising and that are worth exploring in the near future.

References

Baggerly, K.A. (1998). Empirical likelihood as a goodness-of-fit measure, *Biometrika*, **85**, 535-547.

Beran, R. (1981). Nonparametric regression with randomly censored survival data. Technical Report, Univ. California, Berkeley.

Bertail, P. (2006). Empirical likelihood in some semiparametric models. *Bernoulli*, **12**, 299–331.

Brown, B.M. and Chen, S.X. (1998). Combined empirical likelihood. *Ann. Inst. Statist. Math.* **50**, 697–714.

Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika*, **66**, 429–436.

Carroll, R.J. (1982). Adapting to heteroscedasticity in linear model. *Ann. Statist.*, **10**, 1224–1233.

Chen, S.X., Gao, J. and Tang, C.Y. (2008) A test for model specification of diffusion processes. *Ann. Statist.*, **36**, 167–198.

Chen, S.X., Härdle, W. and Li, M. (2003). An empirical likelihood goodness-of-fit test for time series. *J. R. Statist. Soc. - Series B*, **65**, 663–678.

Chen, S.X., Peng, L. and Qin, Y.-L. (2009). Effects of data dimension on empirical likelihood. *Biometrika*, **96**, 711–722.

Chen, S.X. and Van Keilegom, I. (2009). A goodness-of-fit test for parametric and semiparametric models in multiresponse regression. *Bernoulli* (to appear).

El Ghouch, A., Van Keilegom, I. and McKeague, I. (2009). Empirical likelihood confidence intervals for dependent duration data. *Econometric Theory* (to appear).

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer-Verlag.

Hall, P. and Owen, A. B. (1993) Empirical likelihood confidence bands in density estimation. *J. Comput. Graph. Statist.* **2**, 273–289.

Hjort, N.L., McKeague, I.W. and Van Keilegom, I. (2009). Extending the scope of empirical likelihood. *Ann. Statist.*, **37**, 1079–1115.

Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.*, **53**, 457–481.

Koul, H., Susarla, V. and van Ryzin, J. (1981). Regression analysis with randomly right-censored data. *Ann. Statist.*, **9**, 1276–1288.

Owen, A.B. (1991). Empirical likelihood for linear models. *Ann. Statist.*, **19**, 1725–1747.

Ren, J.-J. (2008). Weighted empirical likelihood in some two-sample semiparametric models with various types of censored data. *Ann. Statist.*, **36**, 147–166.

Shahidi, A.R. (2009). *Model selection for moment condition models using the penalized empirical likelihood procedure*. Technical report (available at `http://www.pitt.edu/~ahs19/JMP.pdf`).

Stute, W. (1999). Nonlinear censored regression. *Statist. Sinica*, **9**, 1089–1102.

Tang, C.Y. and Len, C. (2009). *Penalized high dimensional empirical likelihood*. Technical report.

Zhu, L. and Xue, L. (2006). Empirical likelihood confidence regions in a partially linear single-index model. *J. Roy. Statist. Soc., Series B*, **68**, 549–570.