

# UC Berkeley

## CUDARE Working Papers

### Title

Revised empirical likelihood

### Permalink

<https://escholarship.org/uc/item/6gs579r0>

### Authors

Grendar, Marian  
Judge, George G.

### Publication Date

2010-07-01

University of California, Berkeley  
Department of Agricultural &  
Resource Economics

## *CUDARE Working Papers*

---

*Year 2010*

*Paper 1106*

---

### Revised empirical likelihood

Marian Grendar and George Judge

# Revised empirical likelihood

Marian Grendár\*

*Department of Mathematics, Faculty of Natural Sciences, Bel University, Tajovského 40,  
SK-974 01 Banská Bystrica, Slovakia,  
Institute of Mathematics and CS, Slovak Academy of Sciences (SAS),  
Institute of Measurement Sciences of SAS, Bratislava, Slovakia  
e-mail: [marian.grendar@savba.sk](mailto:marian.grendar@savba.sk)*

George Judge

*Professor in the Graduate School, 207 Giannini Hall, University of California, Berkeley, CA,  
94720  
e-mail: [gjudge@berkeley.edu](mailto:gjudge@berkeley.edu)*

**Abstract:** Empirical Likelihood (EL) and other methods that operate within the Empirical Estimating Equations ( $E^3$ ) approach to estimation and inference are challenged by the Empty Set Problem (ESP). ESP concerns the possibility that a model set, which is data-dependent, may be empty for some data sets. To avoid ESP we return from  $E^3$  back to the Estimating Equations, and explore the Bayesian infinite-dimensional Maximum A-posteriori Probability (MAP) method. The Bayesian MAP with Dirichlet prior motivates a Revised EL (ReEL) method. ReEL *i*) avoids ESP as well as the convex hull restriction, *ii*) attains the same basic asymptotic properties as EL, and *iii*) its computation complexity is comparable to that of EL.

**AMS 2000 subject classifications:** Primary 62F10.

**Keywords and phrases:** empirical estimating equations, generalized minimum contrast, empirical likelihood, generalized empirical likelihood, empty set problem, convex hull restriction, estimating equations, maximum a-posteriori probability.

Date: July 8, 2010.

## Contents

1	Introduction: Estimating Equations, Empirical Estimating Equations . . . . .	2
2	Existence problems of $E^3$ . . . . .	4
3	Maximum A-posteriori Probability with Estimating Equations . . . . .	6

---

\*Supported by VEGA 1/0077/09 grant and by the Agency of the Slovak Ministry of Education for the Structural Funds of the EU, under project ITMS:26220120007.

3.1	Discrete case: Dirichlet prior MAP-EE . . . . .	6
3.2	Continuous case . . . . .	8
4	ReEL: Revised Empirical Likelihood . . . . .	8
5	Conclusions and implications . . . . .	14
	References . . . . .	14

## 1. Introduction: Estimating Equations, Empirical Estimating Equations

In Econometrics and other fields, it is rather common to formulate a model in terms of the *Estimating Equations* (EE); cf. [Godambe and Kale \(1991\)](#). The stochastic part of the model is specified by a random variable  $X \in \mathcal{X} \subseteq \mathbb{R}^d$ , with the cumulative distribution function (cdf)  $Q_r \in \mathcal{Q}(\mathcal{X})$ , where  $\mathcal{Q}(\mathcal{X})$  is the set of all cdf's on  $\mathcal{X}$ . Estimating equations are formulated in terms of estimating functions  $u(X; \theta) : \mathcal{X} \times \Theta \rightarrow \mathbb{R}^J$ , where  $\theta \in \Theta \subseteq \mathbb{R}^K$ . There,  $K$  can be, in general, different than  $J$ . Estimating equations

$$\Phi(\theta) = \{Q \in \mathcal{Q}(\mathcal{X}) : E_Q u(X; \theta) = 0\}$$

serve to form the model  $\Phi(\Theta) = \bigcup_{\theta \in \Theta} \Phi(\theta)$ .

Examples of simple models:

Model 1:  $\mathcal{X} = \mathbb{R}$ ,  $\Theta = [0, \infty)$ ,  $u(X; \theta) = X - \theta$ .

Model 2:  $\mathcal{X} = \mathbb{R}$ ,  $\Theta = \mathbb{R}$ ,  $u(X; \theta) = \{X - \theta, \text{sgn}(X - \theta)\}$ ; cf. [Brown and Chen \(1998\)](#).

Model 3:  $\mathcal{X} = \mathbb{R}$ ,  $\Theta = \mathbb{R}$ ,  $u(X; \theta) = \{X - \theta, X^2 - (2\theta^2 + 1)\}$ ; cf. [Qin and Lawless \(1994\)](#), Example 1.

Given a random sample  $X_1^n = X_1, \dots, X_n$  from  $Q_r$ , the objective is to select a  $\hat{Q}$  from  $\Phi(\Theta)$ , and this way make a point estimate  $\hat{\theta}$  of the 'true' value  $\theta_r$ . In the case of correctly specified model (i.e.,  $Q_r \in \Phi(\Theta)$ ),  $\theta_r$  solves  $E_{Q_r} u(X; \theta) = 0$ . If the model is not correctly specified, i.e.,  $Q_r \notin \Phi(\Theta)$ , then  $\theta_r$  can be defined as the solution of  $E_{\hat{Q}(Q_r)} u(X; \theta) = 0$ , where  $\hat{Q}(Q_r)$  is the  $L$ -projection of  $Q_r$  on  $\Phi(\Theta)$ , cf. [Grendár and Judge \(2009a\)](#).

In order to connect the model  $\Phi(\Theta)$  with the data  $X_1^n$ , the *Empirical Estimating Equations* ( $E^3$ ) approach to estimation and inference replaces the model  $\Phi(\Theta)$  by its empirical, data-based analogue  $\Phi_n(\Theta) = \bigcup_{\theta \in \Theta} \Phi_n(\theta)$ , where

$$\Phi_n(\theta) = \{Q_n \in \mathcal{Q}(X_1^n) : E_{Q_n} u(X; \theta) = 0\}$$

are the empirical estimating equations.  $E^3$  replaces the set  $\Phi(\Theta)$  of cdf's supported on  $\mathcal{X}$  by the set  $\Phi_n(\Theta)$  of cdf's that are supported on the data  $X_1^n$ . An

estimate  $\hat{\theta}$  of  $\theta_r$  is then obtained by means of a rule that selects  $\hat{Q}_n(x; \hat{\theta})$  from  $\Phi_n(\Theta)$ .

A broad class of methods for selecting a sample-based cdf from  $\Phi_n(\Theta)$  is provided by the Generalized Minimum Contrast (GMC) rule, cf. [Corcoran \(1998\)](#), [Bickel, Klaassen, Ritov et al \(1993\)](#), [Kitamura \(2007\)](#), that suggests to select

$$\hat{Q}_n(x; \hat{\theta}) = \arg \inf_{Q_n(x) \in \Phi_n(\Theta)} D_\phi(Q_n \| \tilde{Q}_r), \quad (1)$$

where  $\tilde{Q}_r$  is a non-parametric estimate of  $Q_r$ ; the divergence (or contrast)

$$D_\phi(Q_n \| \tilde{Q}_r) = E_{\tilde{Q}_r} \phi \left( \frac{dQ_n}{d\tilde{Q}_r} \right),$$

and  $\phi(\cdot)$  is a convex function, with minimum at 1. Typical choices of  $\phi(\cdot)$  are  $\phi(v) = -\log v$  (leads to the Maximum Empirical Likelihood, [Qin and Lawless \(1994\)](#)),  $v \log v$  (leads to the Exponential Empirical Likelihood, [Kitamura and Stutzer \(1997\)](#), [Imbens, Spady and Johnson \(1998\)](#), [Mittelhammer, Judge and Miller \(2000\)](#)),  $(v^2 - 1)/2$  (leads to the Euclidean Empirical Likelihood, [Brown and Chen \(1998\)](#)). Another popular option is to use in (1) the [Cressie and Read \(1984\)](#) class of discrepancies; the resulting estimators are known as the Generalized Empirical Likelihood (cf. [Smith \(1997\)](#)) estimators.

The  $\theta$  part of the optimization problem (1) can be expressed as

$$\hat{\theta}_{\text{GMC}} = \arg \inf_{\theta \in \Theta} \inf_{Q_n(x) \in \Phi_n(\theta)} E_{\tilde{Q}_r} \phi \left( \frac{dQ_n}{d\tilde{Q}_r} \right), \quad (2)$$

The convex dual form of (2) is

$$(\hat{\theta}, \hat{\lambda})_{\text{GMC}} = \arg \min_{\theta \in \Theta} \max_{\mu \in \mathbb{R}, \lambda \in \mathbb{R}^J} [\mu - E_{\tilde{Q}_r} \phi^*(\mu + \lambda' u(x; \theta))], \quad (3)$$

where  $\phi^*(w) = \sup_v vw - \phi(v)$  is the Legendre Fenchel transformation of  $\phi(v)$ . Asymptotic properties of GMC estimators are known, and provide a basis for inference. Hypothesis tests and confidence intervals can be also constructed by means of the GMC analogue of the Wilks' Theorem. A GMC test of the null hypothesis  $H_0 : \theta = \theta_0$  against the alternative  $H_1 : \theta = \theta_1$  is based on the GMC statistic

$$\lambda(\theta; X_1^n) = k_n(\phi) \left[ \inf_{Q \in \Phi_n(\theta_0)} D_\phi(Q \| \tilde{Q}_r) - \inf_{Q \in \Phi_n(\theta_1)} D_\phi(Q \| \tilde{Q}_r) \right], \quad (4)$$

that is asymptotically  $\chi_K^2$  distributed, under regularity conditions (cf. [Corcoran \(1998\)](#), [Owen \(2001\)](#)), and for an appropriate choice of  $k_n(\phi)$ .

The most prominent member of the GMC class is the Maximum Empirical Likelihood (MEL) estimator  $\hat{\theta}_{\text{MEL}}$ , that can be obtained by choosing  $\phi(v) = -\log v$  (hence  $\phi^*(w) = -1 - \log(-w)$ ) and choosing as the non-parametric estimate  $\tilde{Q}_r$  of  $Q_r$  the empirical cdf  $\hat{Q}_r(x) = \frac{\sum_{i=1}^n I(X_i \leq x)}{n}$ , where  $I(\cdot)$  denotes the indicator function:

$$(\hat{\theta}, \hat{\lambda})_{\text{MEL}} = \arg \min_{\theta \in \Theta} \max_{\lambda \in \mathbb{R}^J} E_{\hat{Q}_r} \log(1 + \lambda' u(x; \theta)). \quad (5)$$

It is worth noting that MEL is the Maximum Likelihood estimator in the class of data-based cdf's from  $\Phi_n(\Theta)$ . The GMC statistic is in this case usually known as the (log) Empirical Likelihood Ratio (ELR) statistic; cf. [Owen \(2001\)](#).

## 2. Existence problems of $E^3$

The  $E^3$ -based methods are subject to existence problems that arise from the fact that the model  $\Phi_n(\Theta)$  is data dependent. There are two classes of the existence problems. Although they are closely related, one pertains to hypothesis testing and confidence interval construction, the other concerns non-existence of  $E^3$ -based estimators.

The inference-related existence problem is well-known, and concerns the impossibility of constructing a test based on GMC statistic, for a data set  $X_1^n$ , for which  $\Phi_n(\theta_0) = \emptyset$ . In the literature on the Empirical Likelihood this is known as the convex hull restriction, cf. [Owen \(2001\)](#). Several ways of mitigating the problem were proposed; cf. [Chen, Variyath and Brown \(2008\)](#), [Emerson and Owen \(2009\)](#). Among them is a modification of  $E^3$  that lifts up the restriction that  $Q_n \in \mathcal{Q}(X_1^n)$ , in the sense that if  $x_1 < x_2$ ,  $x_1, x_2 \in \mathcal{X}$ , then  $F(x_1)$  may be greater than  $F(x_2)$ . Permitting the negative weights on data points allows to escape the convex hull restriction. GMC with divergences that allow negative 'weights', such as the euclidean one, which is given by  $\phi(v) = (v^2 - 1)/2$ , can be used within the modified  $E^3$  approach ( $mE^3$ ).

Recently, it was demonstrated that  $E^3$  faces an even deeper existence problem, the Empty Set Problem (ESP); cf. [Grendar and Judge \(2009b\)](#). ESP concerns the possibility that  $\Phi_n(\Theta) = \emptyset$ , for some models and data. If for a model and a data set  $\Phi_n(\Theta)$  is empty, then no  $E^3$ -based estimate can be obtained. Consequently, also none  $E^3$ -based test can be constructed.

The setting of Model 1 provides a simple instance of ESP. If  $X_1^n$  is such that the largest value  $X_{(n)}$  in the sample is smaller than 0, then  $\Phi_n(\Theta) = \emptyset$ . Probability of obtaining such data set depends on the sample size  $n$  and on the sampling distribution  $Q_r(x)$ .

In the Model 3,  $\Phi_n(\Theta)$  is empty, for any data set  $X_1^n$  for which  $E_{Q_n} X^2 - 2(E_{Q_n} X)^2 - 1 < 0$ , for any  $Q_n \in \mathcal{Q}(X_1^n)$ . As it was demonstrated in [Grendar and Judge \(2009b\)](#),  $\Phi_n(\Theta)$  is empty for any data set  $X_1^n = x_1^n$  such that

$$q_{(1)}^m x_{(1)}^2 + (1 - q_{(1)}^m) x_{(n)}^2 - 2 \left( q_{(1)}^m x_{(1)} + (1 - q_{(1)}^m) x_{(n)} \right)^2 < 1, \quad (6)$$

where

$$\hat{q}_{(1)}^m = \frac{\frac{x_{(1)}^2 - x_{(n)}^2}{4(x_{(1)} - x_{(n)})} - x_{(n)}}{x_{(1)} - x_{(n)}}, \quad (7)$$

and  $x_{(1)}$  ( $x_{(n)}$ ) is the smallest (the largest) observed value in the sample. Consequently, neither MEL nor any other  $E^3$ -based estimator exists for such a data set. As it was also noted by the authors, even the methods that operate within the modified  $E^3$  approach (i.e., the methods that allow negative weights) face an empty set problem for such data sets.

Not every  $E^3$  model is subject to ESP. For instance, Model 2 is free of ESP. Whether a model is subject to ESP or not is, in general, difficult to assess.

ESP, in our view, depreciates methods that operate within  $E^3$  approach. In [Grendar and Judge \(2009b\)](#) few ways out were identified. One of them is to return back to EE and use the Maximum Likelihood or Bayesian methods. Maximum Likelihood with Estimating Equations (ML-EE) method has been explored in [Grendar and Judge \(2010\)](#). In the discrete case, ML-EE is not difficult to find, numerically. In the continuous case, it is an open problem how to make ML-EE operative. In [Grendar and Judge \(2010\)](#), a discretized version of ML-EE was proposed. The discretized ML-EE avoids ESP, enjoys the same asymptotic properties as MEL, but its construction is difficult and model-dependent.

In this work we continue the quest for a method that *i*) operates within EE, consequently *ii*) it avoids ESP (and also the convex hull problem), *iii*) enjoys the same asymptotic properties as EL, *iv*) is comparable to MEL also from the computational point of view. To this end we explore the non-parametric Bayesian Maximum A-posteriori Probability (MAP) method within EE approach. MAP-EE helps us to find a non-bayesian method that achieves all the four goals.

### 3. Maximum A-posteriori Probability with Estimating Equations

The Empty Set Problem (ESP) of  $E^3$  arises as a consequence of the restriction to the probability distributions that are supported by data. Due to this restriction,  $E^3$  methods ignore the information about the support  $\mathcal{X}$  of  $X$ .

One option to avoid ESP is to return back to the model,  $\Phi(\Theta)$ , that is specified by Estimating Equations (EE). To place EE into Bayesian setting, let a strictly positive prior  $\Pi$  be put over  $\Phi(\Theta)$ . The prior combines with data  $X_1^n = X_1, X_2, \dots, X_n$  to define the posterior distribution

$$\Pi_n(O | X_1^n) = \frac{\int_O e^{-l_n(q)} \Pi(dQ)}{\int_{\Phi(\Theta)} e^{-l_n(q)} \Pi(dQ)},$$

where  $O \subseteq \Phi(\Theta)$ ,  $l_n(q) = -\sum_{l=1}^n \log q(x_l)$ ,  $q = \frac{dQ}{d\mu}$ , and  $\mu$  is assumed to be either the counting or the Lebesgue measure. The Maximum A-posteriori Probability with Estimating Equations (MAP-EE)  $\hat{\theta}_{\text{MAP-EE}}$  estimator of  $\theta$  is defined as the parametric component of

$$\hat{Q}(x; \hat{\theta})_{\text{MAP-EE}} = \arg \sup_{Q \in \Phi(\Theta)} \Pi_n(\cdot | X_1^n). \quad (8)$$

#### 3.1. Discrete case: Dirichlet prior MAP-EE

Let  $X$  be a univariate, discrete random variable with finite support  $\mathcal{X} = \{x_1, x_2, \dots, x_m\}$ . Let the prior  $\Pi$  over  $\Phi(\Theta)$  be given as a normalization of the prior over  $\mathcal{Q}(\mathcal{X})$ , i.e.,  $\Pi(O) = \int_O \Pi(dQ) / \int_{\Phi(\Theta)} \Pi(dQ)$ , where  $O \subseteq \Phi(\Theta)$ . We assume the Dirichlet prior which leads to the Dirichlet posterior

$$\Pi_n(q | X_1^n) \propto e^{-l_n^D(q)},$$

where

$$l_n^D(q) = -\sum_{i=1}^m (n_0 p_{0,i} + n_i - 1) \log q(x_i; \theta),$$

$n_0 \in \mathbb{N}_+$  is the precision parameter, and  $p_{0,i} > 0$ ,  $1 \leq i \leq m$ ,  $\sum_{i=1}^m p_{0,i} = 1$ , is the base measure, with cdf  $P_0$ ; cf. [Ghosh and Ramamoorthi \(2003\)](#). Also, there  $n_i = \sum_{l=1}^n I(X_l = x_i)$  and  $I(\cdot)$  is the indicator function. Then, MAP-EE is given as

$$\hat{Q}(x; \hat{\theta})_{\text{MAP-EE}} = \arg \inf_{Q \in \Phi(\Theta)} l_n^D(\cdot)/n. \quad (9)$$

Density (i.e., pmf), corresponding to the optimal cdf  $\hat{Q}(x; \hat{\theta})_{\text{MAP-EE}}$  can be expressed as

$$\hat{q}(x_i; \hat{\theta}, \hat{\lambda}(\hat{\theta}))_{\text{MAP-EE}} = \frac{v_i^n + (n_0 p_{0,i} - 1)/n}{1 + (n_0 - m)/n + \hat{\lambda}' u(x_i; \hat{\theta})}, \quad 1 \leq i \leq m,$$

where  $\lambda \in \mathbb{R}^J$ , and  $v_i^n = n_i/n$  is pmf corresponding to the empirical cdf  $\hat{Q}_r$ . The optimal values of  $\theta$ ,  $\lambda$  can be found as the solution of the convex dual problem

$$(\hat{\theta}, \hat{\lambda})_{\text{MAP-EE}} = \arg \inf_{\theta \in \Theta} \sup_{\lambda \in \mathbb{R}^J} E_{\hat{Q}_r + (n_0 p_0 - 1)/n} \log \left( 1 + \frac{n_0 - m}{n} + \lambda' u(x_i; \theta) \right). \quad (10)$$

The convex dual for MEL (5) and MAP-EE (10) are worth comparing. Since MAP-EE operates within the EE approach, it is free of ESP. More precisely, if  $n_0$  and  $P_0$  are such that,  $n_0 p_{0,i} \neq 1$  for all  $i = 1, \dots, m$ , then  $\hat{q}_{\text{MAP-EE}}(x_i; \cdot)$  is non-zero, even for data sets for which  $v_i^n$  may be zero. Under this condition the Dirichlet prior MAP-EE avoids ESP

*Example 1.* Let  $\mathcal{X} = (0, 1, 2, 3)$ ,  $\Theta = [0.52, 0.96]$ , so that  $\Phi(\Theta) \neq \emptyset$ . Let the estimating functions be the same as in Model 3. Let the sample of size  $n = 10$  be such that it induces type (i.e., the vector of relative frequencies)  $v^{10} = [2, 3, 5, 0]/10$ .

$E^3$ -based methods *a priori* put zero probability to those outcomes that do not appear in the sample. Since  $\Phi_n(\Theta)$  is data-dependent, it may be empty for some models and data sets, as it is the case in this setting, for this sample (type); cf. Grendar and Judge (2009b), Sect. 2.1 and Grendar and Judge (2010). Thus, neither MEL nor any other  $E^3$ -based estimate exists for the sample.

However, MAP-EE exists. Assume, for instance, the uniform base measure and  $n_0 = 5$ . By means of the standard numeric methods, it can be found that  $\hat{\lambda}_{\text{MAP-EE}} = (20.02761, -6.69760)$  and  $\hat{\theta}_{\text{MAP-EE}} = 0.74757$ . Though the fourth outcome from  $\mathcal{X}$  has not appeared in the sample, the pmf recovered by MAP-EE has fourth element non-zero,  $\hat{q}(x_4; \hat{\lambda}, \hat{\theta})_{\text{MAP-EE}} = 0.21547$ .  $\diamond$

The main difference between the Dirichlet MAP-EE (10) and MEL (5) lays in the  $(n_0 p_0 - 1)/n$  term, that permits to avoid ESP. Asymptotic frequentist properties of the estimator in (10) are the same as those of MEL: under the regularity conditions of Theorem 3.6 of Owen (2001), the MAP-EE estimator

of  $\theta$  is consistent and asymptotically normally distributed, with the covariance matrix

$$\Sigma = \left[ E \frac{\partial u'}{\partial \theta} (E u u')^{-1} E \frac{\partial u}{\partial \theta} \right]^{-1}. \quad (11)$$

The log-posterior ratio  $R_{\text{MAP-EE}}(\Theta_0)$  for the null hypothesis that  $\theta = \theta_0$ ,

$$R_{\text{MAP-EE}}(\Theta_0) = \log \frac{\Pi_n(\hat{q}(\cdot; \theta_0)_{\text{MAP-EE}} | X_1^n)}{\Pi_n(\hat{q}(\cdot; \hat{\theta})_{\text{MAP-EE}} | X_1^n)}$$

is such that  $-2R_{\text{MAP-EE}}(\theta_0)$  is asymptotically  $\chi_K^2$  distributed. Proofs follow the lines of [Qin and Lawless \(1994\)](#). The log-posterior ratio is free of the convex hull problem, that hampers the Empirical Likelihood Ratio (cf. Sect. 2).

### 3.2. Continuous case

In the continuous case, it is in general difficult to make MAP-EE (8) operational. Part of the difficulty arises from the fact, that even for the simplest prior – the Dirichlet process prior, it is necessary to resort to Monte Carlo simulations. [Florens and Rolin \(1994\)](#) explored this route, for the exactly identified EE model (i.e.,  $J = K$ ). The authors note that this approach is not viable for over-identified models (i.e.,  $J > K$ ).

Another option is to partition the support  $\mathcal{X}$ , and this way turn the setup into a discrete one. However, partitioning is arbitrary. Moreover, there is another problem with this option. Even in the univariate case, where the Dirichlet prior can be employed, the resulting posterior involves the range of support, which must be necessarily finite. Despite these limitations, discretized support together with the Dirichlet prior can be used to motivate a revision of MEL that can be used in discrete as well as continuous case, univariate or multivariate case, and what is the most important, avoids ESP. We now turn to the Revised Empirical Likelihood method.

## 4. ReEL: Revised Empirical Likelihood

Consider the following method for selecting a cdf from  $\Phi(\Theta)$

$$\hat{Q}(x; \hat{\theta}) = \arg \inf_{Q \in \Phi(\Theta)} L(Q || (1 - \alpha)Q_r + \alpha P_0), \quad (12)$$

where  $L(V || U) = -E_U \log \frac{dV}{dU}$  is the L-divergence [Grendar and Judge \(2009a\)](#) of  $V \in \mathcal{Q}(\mathcal{X})$  with respect to  $U \in \mathcal{Q}(\mathcal{X})$ ;  $\alpha \in (0, 1)$ , and  $P_0 \in \mathcal{Q}(\mathcal{X})$ . The convex dual form of (12) is

$$(\hat{\theta}, \hat{\lambda}) = \arg \inf_{\theta \in \Theta} \sup_{\lambda \in \mathbb{R}^J} E_{(1-\alpha)Q_r + \alpha P_0} \log (1 + \lambda' u(x; \theta)). \quad (13)$$

To tie the method with data, replace the unknown  $Q_r$  in (12), (13) by a data-based estimate  $\tilde{Q}_r$ . Resulting method will be named Revised Empirical Likelihood (ReEL<sup>1</sup>). Though ReEL is not a Bayesian method, the cdf  $P_0 \in \mathcal{Q}(\mathcal{X})$  will be called the base measure, and  $\alpha$  will be related to the precision parameter  $n_0 \in (0, n)$ , as  $\alpha = n_0/n$ . The main purpose of  $P_0$  is to bring into model the information about support and this way avoid ESP. Through  $P_0$  it is also possible to enter an extra-sample, extra-EE, prior information into the model.

In what follows, we will consider two instances of ReEL. In the first one,  $\tilde{Q}_r = \hat{Q}_r$  (i.e., the empirical cdf). It is worth stressing that this instance of ReEL turns into MEL, for  $n_0 \rightarrow 0$ . The other instance of ReEL is such that the pdf corresponding to  $Q_r$  is estimated by kernel density estimator. This instance of ReEL will be named kernel ReEL. With a little danger of confusion, we reserve ReEL to mean the empirical cdf ReEL.

Density corresponding to the sought cdf  $\hat{Q}_{\text{ReEL}}$  takes the following form

$$\hat{q}(x; \hat{\theta}, \hat{\lambda})_{\text{ReEL}} = \frac{(1 - \alpha)\tilde{q}_r(x) + \alpha p_0(x)}{1 + \hat{\lambda}' u(x; \hat{\theta})}, \quad (14)$$

where  $\tilde{q}_r = \frac{d\tilde{Q}_r}{d\mu}$  is the density corresponding to the estimator  $\tilde{Q}_r$  of the true data-sampling cdf  $Q_r$ ; and  $p_0 = \frac{dP_0}{d\mu}$ ,  $p_0 > 0$ . For  $\tilde{Q}_r = \hat{Q}_r$ ,  $\hat{q}_r(x) = \frac{1}{n} I_{x=x_l}$ ,  $l = 1, 2, \dots, n$ .

*Example 2.* As the first illustration of ReEL, consider the setting of Model 1:  $\mathcal{X} = \mathbb{R}$ ,  $\Theta = [0, \infty)$ ,  $u(X; \theta) = X - \theta$ . Recall that in this setting MEL does not exist for any data set  $X_1^n$ , for which the largest value  $X_{(n)} < 0$ . Since ReEL operates on  $\Phi(\Theta)$  rather than on  $\Phi_n(\Theta)$ , it is not affected by ESP. The ReEL estimator of  $\theta$  is

$$\hat{\theta}_{\text{ReEL}} = \begin{cases} 0 & \text{if } \hat{\theta}_u \leq 0, \\ \hat{\theta}_u & \text{otherwise,} \end{cases}$$

where

$$\hat{\theta}_u = E_{(1-\alpha)\tilde{Q}_r + \alpha P_0} X$$

---

<sup>1</sup>Re' may stand for revised, regularized, reconsidered, rescued, reformed, recovered, etc.

is the unrestricted ReEL estimator of  $\theta$ .

In order to use the uniform prior, we decided to work with cdf's on  $[-10, 10]$ . Note that then,  $E_{p_0}X = 0$ . Thus,  $\hat{\theta}_u = (1 - \alpha)E_{\hat{Q}_r(x)}X$ . We put  $n_0 = 1$  and estimate  $Q_r$  by the empirical cdf.

A random sample of size  $n = 10$  from  $n(0, 1)$  was generated, with the highest observed value  $-0.63617$ . For such a sample, MEL does not exist. ReEL estimator of  $\theta$  is 0, since  $\hat{\theta}_u < 0$ . Associated ReEL density  $\hat{q}_{\text{ReEL}}(x; 0, \hat{\lambda})$  is depicted at the upper left panel of Figure 1. Next, we shifted the data by 1 to the right. Since  $\hat{\theta}_u$  remained negative,  $\hat{\theta}_{\text{ReEL}} = 0$ . Corresponding ReEL density  $\hat{q}_{\text{ReEL}}(x; 0, \hat{\lambda})$  is at the upper middle panel of Figure 1. Finally, yet another shift of the data by 1 to the right, resulted in  $\hat{\theta}_u > 0$ , and hence  $\hat{\theta}_{\text{ReEL}} = \hat{\theta}_u$ . Corresponding ReEL density is at the upper right panel of Figure 1.

For the same data, we estimate  $q_r$  by the kernel density estimator with bandwidth selected by Silverman (1986) 'rule of thumb' (the default of R's [R Development Core Team](#) (2009) function `density`). Resulting kernel-ReEL pdf's

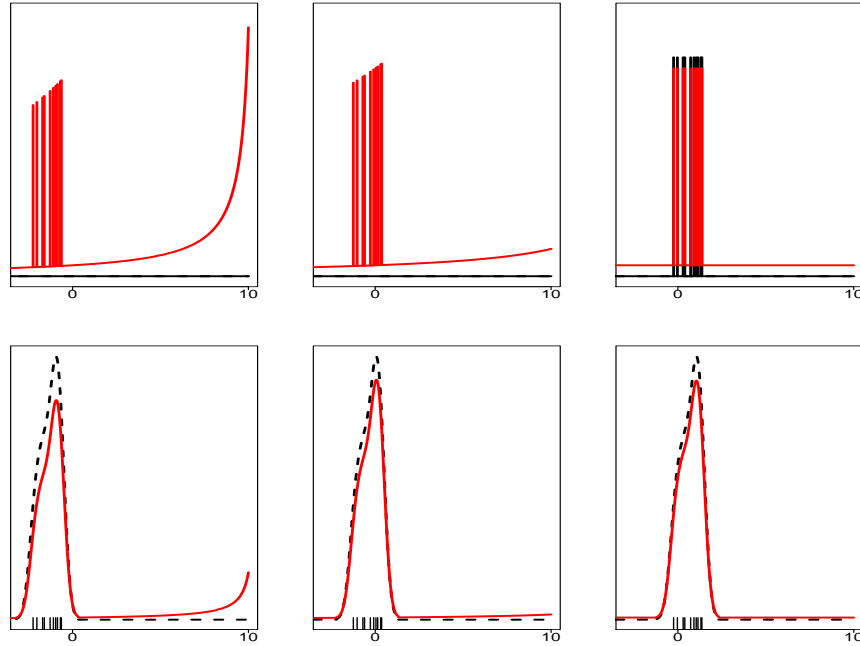


Fig 1. Upper row: ReEL density (solid red line), MEL weights (vertical black lines). Lower row: Kernel ReEL density (solid red line), kernel density (dashed line).

are exhibited at the lower part of Figure 1.  $\diamond$

*Example 3.* Next, consider the setting of Model 3 (i.e., [Qin and Lawless \(1994\)](#), Example 1):  $\mathcal{X} = \mathbb{R}$ ,  $\Theta = \mathbb{R}$ , and the feasible set of cdfs is  $\Phi(\Theta) = \{Q \in \mathcal{Q}(\mathcal{X}) : E_Q(X - \theta) = 0, E_Q(X^2 - (2\theta^2 + 1)) = 0; \theta \in \Theta\}$ . Recall, that in the  $E^3$  approach, the set  $\Phi(\Theta)$  of cdfs supported on  $\mathcal{X}$  is replaced by the set  $\Phi_n(\Theta) = \{Q_n \in \mathcal{Q}(X_1^n) : E_{Q_n}(X - \theta) = 0, E_{Q_n}(X^2 - (2\theta^2 + 1)) = 0; \theta \in \Theta\}$  of cdfs supported on the data  $X_1^n$ . Also, recall that  $\Phi_n(\Theta)$  is empty for any data set for which (6), (7) hold.

ReEL selects cdf from  $\Phi(\Theta)$ , hence existence of ReEL estimator is not data dependent. As in the above Example, we use the uniform base measure on  $[-10, 10]$ . The precision parameter  $n_0$  is set up to 0.1.

For a three samples of size  $n = 10$  from  $n(0, 1)$ , resulting ReEL pdf's are exhibited on Figure 2. On the left and middle panels, there are data for which MEL exists, since  $\Phi_n(\theta) \neq \emptyset$ . For the data on the right panel there is ESP, hence

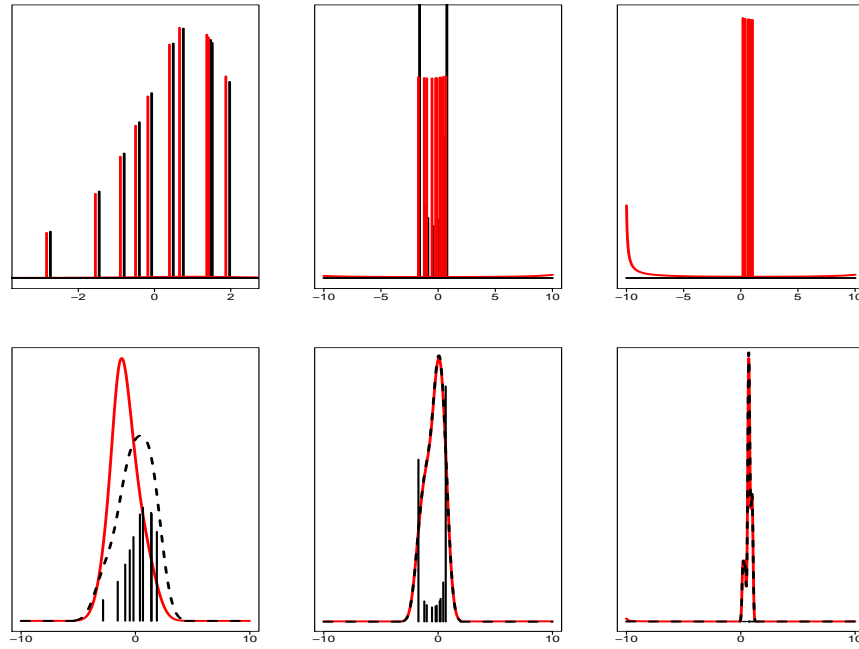


Fig 2. Upper row: ReEL  $\hat{q}(x; \cdot)_{\text{ReEL}}$  (solid red line), EL weights (vertical lines). Lower row: Kernel density estimator  $\hat{q}_r(x)$  (dashed), Kernel ReEL  $\hat{q}(x; \cdot)_{\text{ReEL}}$  (solid red line), EL weights (vertical lines)

MEL does not exist. At the bottom part of Figure 2, kernel ReEL densities are depicted, for the same data as in the upper part.

To compare small-sample performance of ReEL and kernel-ReEL with the gaussian Maximum Likelihood (ML) and the sample mean, we conduct an MC study for the gaussian  $n(0, 1)$  data. For  $n = 10$ , kernel-smoothed distribution of ReEL and kernel ReEL for  $n_0 = (0.1, 0.5, 1)$ , as well as for ML and the sample mean are depicted at Figure 3. Among the 1000 MC samples there were 91 for which  $\Phi_n(\Theta) = \emptyset$ , consequently, MEL was excluded from the comparison.

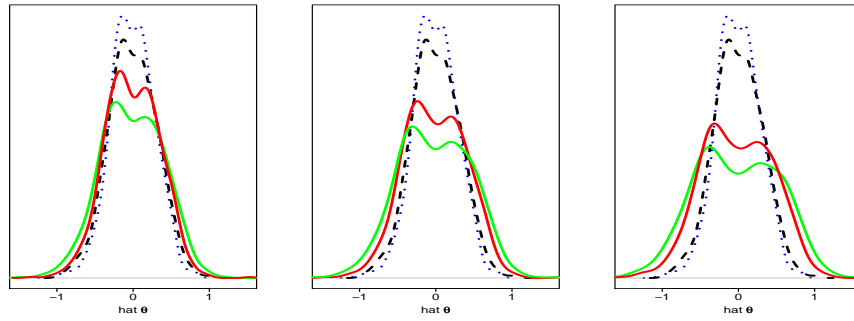


Fig 3. Distribution of estimators: ReEL (solid red line), kernel ReEL (solid green line), ML (dotted line), mean (dashed line). Left,  $n_0 = 0.1$ ; middle,  $n_0 = 0.5$ ; right,  $n_0 = 1.0$ . Based on 1000 MC samples of size  $n = 10$ .

For a more quantitative comparison, MC study of mean and variance of the estimators was conducted for  $n = (10, 50, 100)$ . Results are presented in Table 1.  $\diamond$

Table 1

Mean and variance, multiplied by  $10^3$ , of the sample mean, ML and ReEL (with  $n_0 = (0.1, 0.5, 1)$ ), from 1000 MC simulations. In the ReEL block, the upper line is for ReEL, the lower line for the kernel ReEL.

$n$	Mean		ML		ReEL					
					0.1		0.5		1.0	
10	-5	98	-6	80	-7	125	-9	163	-15	216
					-15	166	-17	219	-21	296
50	-7	19	-7	19	-8	22	-9	24	-9	27
					-9	28	-10	31	-11	36
100	-0.5	10	-0.8	9	-1.6	10	-1.6	11	-1.6	12
					-2	13	-2	14	-2	15

As far as asymptotic properties of ReEL are concerned, the estimator is asymptotically normally distributed with covariance matrix given by (11), under the regularity conditions of Theorem 3.6 of Owen (2001). It is straightforward to see that ReEL also satisfies the Bayesian consistency under misspecification requirement; cf. Grendár and Judge (2009a).

In analogy with the Empirical Likelihood Ratio (ELR) it is possible to define the (log) Revised Empirical Likelihood Ratio (ReELR)  $R_{\text{ReEL}}(\theta_0)$ , for the null hypothesis that  $\theta = \theta_0$ ,

$$R_{\text{ReEL}}(\theta_0) = \inf_{Q \in \Phi(\theta_0)} R - \inf_{Q \in \Phi(\Theta)} R. \quad (15)$$

Under the regularity conditions of Theorem 3.6 of Owen (2001), it can be shown that  $+2nR_{\text{ReEL}}(\theta_0) \rightarrow \chi_K^2$ . Since ReEL is not affected by empty set problem, ReELR is not affected by convex hull constraint.

*Example 4.* As an illustration of the  $\chi^2$  asymptotics of ReELR, consider the simple setting of testing the mean; i.e.,  $u(X; \theta) = X - \theta$ ,  $\mathcal{X} = \mathbb{R}$ ,  $\Theta = \mathbb{R}$ . The data are sampled from  $n(0, 1)$ . The null hypothesis is given by  $\Theta_0 = \{0\}$ . As before, the base measure is uniform on  $[-10, 10]$ . Then,

$$R_{\text{ReEL}}(\theta_0) = E_{(1-\alpha)\hat{Q}_r + \alpha P_0} \log(1 + \hat{\lambda}(0)x).$$

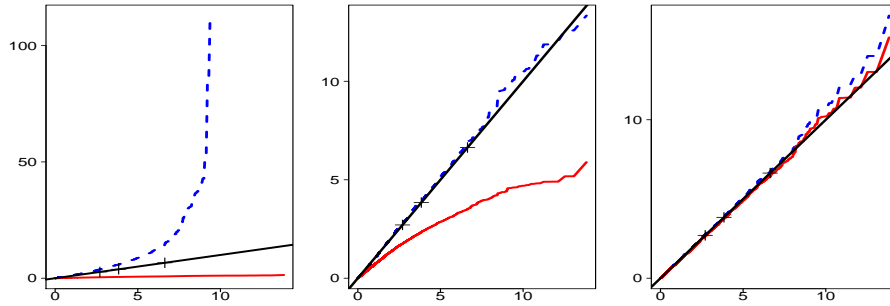


Fig 4. Quantile-quantile plots for ReELR, uniform base measure,  $n_0 = 1$  (solid), ELR (dashed) versus the reference  $\chi_1^2$  distribution. Left:  $n = 10$ , middle:  $n = 100$ , right:  $n = 1000$ . The crosses on the  $y = x$  line indicate the 90%, 95%, and 99% quantiles of the  $\chi_1^2$  distribution. MC sample size 5000.

ELR is affected by the convex hull restriction. If the data  $X_1^n = x_1^n$  are such that  $x_{(n)} < 0$ , ELR does not exist. Among 5000 MC samples of size  $n = 10$ , ELR

did not exist for 9 samples. ReELR is not affected by convex hull restriction. For the sake of comparison with ReELR, we plot the quantile-quantile curve also for ELR, excluding the data sets that are out of the convex hull. Figure 4 demonstrates the asymptotics, for empirical cdf ReELR. Kernel-based ReELR behaves similarly.

## 5. Conclusions and implications

Replacement of Estimating Equations (EE) by Empirical Estimating Equations ( $E^3$ ) seems to be a natural way to link the model with data. The resulting  $E^3$  model  $\Phi_q(\Theta)$  is data dependent and for some models there are data samples for which  $\Phi_q(\Theta)$  is empty. Consequently,  $E^3$  models has to be checked on case-by-case basis for Empty Set Problem (ESP) – a tedious task. The  $E^3$ -based methods can be applied only to those models that are free of ESP and this makes the  $E^3$  approach peculiar.

In [Grendar and Judge \(2009b\)](#), a few ways out of the ESP were identified. One of them is the Generalized Method of Moments; cf. [Hansen \(1982\)](#). Another option, proposed here, is to return back to EE and apply to it the Bayesian Maximum A-Posteriori Probability (MAP) method. In the univariate discrete case, the Dirichlet prior MAP-EE is a natural and at the same time ESP-free method. For continuous, multivariate case implementation of MAP-EE would require Monte Carlo sampling. However, Dirichlet MAP-EE motivates a Revised Empirical Likelihood (ReEL) method, that *i*) operates on EE, and hence avoids ESP as well as the convex hull restriction, *ii*) is applicable to multivariate as well as univariate, continuous as well as discrete cases, *iii*) enjoys the same basic asymptotic properties as EL, and finally, *iv*) its computational complexity is comparable to that of EL. ReEL combines an EE model, a data-based estimate of sampling distribution, and a 'prior' information about range of values (support). We explored two instances of ReEL: the empirical cdf ReEL, and kernel smooth ReEL, at few simple settings. Much of further explorations is warranted.

## References

- Bickel, P. J., Klaassen, C.A.J., Ritov, Y., and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore: Johns Hopkins University Press.

- Brown, B. M., and Chen, S. X. (1998). Combined and Least Squares Empirical Likelihood. *Ann. Inst. Statist. Math.* 90:443-450.
- Chen, J., Variyath, A. M., and Abraham, B. (2008). Adjusted Empirical Likelihood and its properties. *J. Comput. Graph. Statist.* 17/2:426-443.
- Corcoran, S. A. (1998). Bartlett adjustment of empirical discrepancy statistics. *Biometrika*. 85:967-972.
- Corcoran, S. A. (2000). Empirical exponential family likelihood using several moment conditions. *Stat. Sinica*. 10:545-557.
- Cressie, N., and Read, T. (1984). Multinomial goodness of fit tests. *J. Roy. Statist. Soc. Ser. B*. 46:440-464.
- Emerson, S. C. and Owen, A. B. (2009). Calibration of the empirical likelihood method for a vector mean. *Electron. J. Statist.* 3:1161-1192.
- Florens, J.-P and Rolin, J.-M. (1994). Bayes, bootstrap, moments. Discussion paper 94.13. Institute de Statistique, Université catholique de Louvain.
- Ghosh, J. K. and Ramamoorthi, R. V. (2003). *Bayesian Nonparametrics*. Springer, New York.
- Godambe, V. P., and Kale, B. K. (1991). Estimating functions: an overview. In *Estimating Functions*, V. P. Godambe (ed.), Oxford University Press, Oxford, UK, pp. 3-20.
- Grendár, M., and Judge, G. (2009a). Asymptotic equivalence of Empirical Likelihood and Bayesian MAP. *Ann. Statist.* 37/5A:2445-2457.
- Grendár, M., and Judge, G. (2009b). Maximum Empirical Likelihood: Empty Set Problem. CUDARE Working Paper 1090. Retrieved from: <http://escholarship.org/uc/item/71v338mh>
- Grendár, M., and Judge, G. (2009c). Empty Set Problem of Maximum Empirical Likelihood Methods. *Electron. J. Statist.* 3:1542-1555.
- Grendár, M., and Judge, G. (2010). Maximum Likelihood with Estimating Equations. CUDARE Working Paper 1094. Retrieved from: <http://escholarship.org/uc/item/1r45k876>
- Hansen, L. P (1982). Large sample properties of Generalized Method of Moments estimators. *Econometrica*. 50:1029-1054.
- Imbens, G. W., Spady, R. H., and Johnson, P (1998). Information theoretic approaches to inference in moment condition models. *Econometrica*. 66/2:333-357.
- Kitamura, Y., and Stutzer, M. (1997). An information-theoretic alternative to Generalized Method of Moments estimation. *Econometrica*. 65:861-874.
- Kitamura, Y. (2007). Nonparametric Likelihood: efficiency and robustness. *Jap. Econ. Rev.* 58/1:26-45.

- Mittelhammer, R., Judge, G. and Miller, D. (2000). *Econometric Foundations*. Cambridge:Cambridge University Press.
- Owen, A. B. (2001). *Empirical Likelihood*. New York:Chapman-Hall/CRC.
- Qin, J., and Lawless, J. (1994). Empirical Likelihood and General Estimating Equations. *Ann. Statist.* 22:300-325.
- R Development Core Team (2009). R: A Language and Environment for Statistical Computing. ISBN 3-900051-07-0. Vienna, Austria.
- Silverman, B. W. (1986) Density Estimation. London: Chapman and Hall.
- Smith, R. J. (1997). Alternative semi-parametric likelihood approaches to Generalized Method of Moments estimation. *Economic J.* 107:503-519.