UC Berkeley
CUDARE Working Papers

Title
Maximum Empirical Likelihood: Empty Set Problem

Permalink
https://escholarship.org/uc/item/71v338mh

Authors
Grendar, Marian
Judge, George G

Publication Date
2009-09-10

Department of Agricultural & Resource Economics, UCB

Maximum Empirical Likelihood: Empty Set Problem

Marian Grendar [*]   George G. Judge [†]

[*]Institute of Measurement Sciences SAS, Bratislava, Slovakia
[†]University of California, Berkeley and Giannini Foundation

Maximum Empirical Likelihood: Empty Set Problem

Abstract

In the Empirical Estimating Equations (E3) approach to estimation and inference estimating equations are replaced by their data-dependent empirical counterparts. It is odd but with E3 there are models where the E3-based estimator does not exist for some data set, and does exist for others. This depends on whether or not a set of data-supported probability mass functions that satisfy the empirical estimating equations is empty for the data set. In a finite sample context, this unnoted feature invalidates methods of estimation and inference, such as the Maximum Empirical Likelihood, that operate within E3. The empty set problem of E3 is illustrated by several examples and possible remedies are discussed.

Maximum Empirical Likelihood: Empty Set Problem

M. Grendár[*] and G. Judge[†]

Abstract

In the Empirical Estimating Equations (E^3) approach to estimation and inference estimating equations are replaced by their data-dependent empirical counterparts. It is odd but with E^3 there are models where the E^3-based estimator does not exist for some data set, and does exist for others. This depends on whether or not a set of data-supported probability mass functions that satisfy the empirical estimating equations is empty for the data set. In a finite sample context, this unnoted feature invalidates methods of estimation and inference, such as the Maximum Empirical Likelihood, that operate within E^3. The empty set problem of E^3 is illustrated by several examples and possible remedies are discussed.

1 Introduction

In statistics and other fields such as econometrics, it is rather common to formulate a probabilistic model for the random variable $X \in \mathrm{R}^d$ with a probability distribution $r_X(x; \theta)$ that is parametrized by $\theta \in \Theta \subseteq \mathrm{R}^K$, as a set $\Phi(\Theta)$ of probability distributions with certain moment properties. This is accomplished through estimating functions $u(X; \theta) \in \mathrm{R}^J$ of parameter θ. The estimating functions are employed to form the set $\Phi(\Theta) = \bigcup_{\theta \in \Theta} \Phi(\theta)$ of parametrized probability distributions $F(x; \theta)$, where $\Phi(\theta)$ is defined through the estimating equations [9] as

$$\Phi(\theta) = \left\{ F(x; \theta) : \int u(x; \theta) \, dF(x; \theta) = 0 \right\}.$$

In general, we assume that $r_X(x; \theta)$ need not belong to $\Phi(\Theta)$. Given a random sample $X_1^n = X_1, X_2, \ldots, X_n$ from $r_X(x; \theta)$ the objective is to estimate θ and draw inferences. To this end, the Empirical Estimating Equations (E^3) approach replaces the model $\Phi(\Theta)$ by its empirical, data-based counterpart $\Phi_q(\Theta) = \bigcup_{\theta \in \Theta} \Phi_q(\theta)$ where

$$\Phi_q(\theta) = \left\{ q(x; \theta) : \mathrm{ch}\left(u\left(X_1^n; \theta \right) \right) \cap \{0\} \right\}$$

and in this way, connects it with the data. There, $\mathrm{ch}(A)$ denotes the convex hull of the set $A \subseteq \mathrm{R}^J$, and 0 is the J-dimensional zero vector. Note that in doing so,

the set $\Phi(\Theta)$ of probability distributions is replaced by the set $\Phi_q(\Theta)$ of probability mass functions (pmf's) $q(x; \theta)$ that are supported on the data X_1^n. An estimate $\hat{\theta}$ of θ is then obtained by means of a method that selects a pmf $\hat{q}(\cdot; \hat{\theta})$ from $\Phi_q(\Theta)$. One prominent method is the Maximum Empirical Likelihood Estimation (MELE) method (cf. [20], [19]) which selects

$$\hat{q}(\cdot; \hat{\theta})_{\mathrm{MELE}} = \arg \sup_{q(\cdot; \theta) \in \Phi_q(\Theta)} \frac{1}{n} \sum_{i=1}^n \log q(x_i; \theta). \tag{1}$$

Since the interest is in $\hat{\theta}_{\mathrm{MELE}}$, it can be - by means of the convex duality [21] - obtained as

$$\hat{\theta}_{\mathrm{MELE}} = \arg \inf_{\theta \in \Theta} \sup_{\lambda \in \mathrm{R}^J} -\frac{1}{n} \sum_{i=1}^n \log q(x_i; \theta, \lambda), \tag{2}$$

where

$$q(\cdot; \theta, \lambda) = \left[n \left(1 - \sum_{j=1}^J \lambda_j u_j(\cdot; \theta) \right) \right]^{-1}.$$

The Maximum Empirical Likelihood Estimator (MELE) is a non-linear function of the data. It can be obtained by a numerical solution of the above optimization problem. Asymptotic distributional properties of MELE are known (cf. [20]), and provide a basis for inference.

Methods other than MELE can be used for estimation within E^3 approach. In particular, the objective function in (1) can be replaced by another measure of closeness of $q(\cdot)$ to the uniform distribution, supported on the sample; cf. [1], [3], [5], [15], [16], [18], [19], [22]. Most common are the measures of closeness (also known as divergences) encapsulated by the Cressie Read family [7], or, in general, the so-called convex statistical distances, which lead to the Generalized Minimum Contrast (GMC) estimators [2], [6], [17]. The GMC class contains MELE and Maximum Entropy Empirical Likelihood (also known as the Exponential Tilt, cf. [15], [16], [5], [18]) as special cases.

Finally, let us note that there is also in use a modification of E^3, where the nonnegativity constraints $q(x_i; \theta) \geq 0$ are dropped out, so that $\Phi_q(\theta)$ is replaced by

$$\Phi_q^m(\theta) = \left\{ q(x; \theta) : \mathrm{ah}\left(u\left(X_1^n; \theta \right) \right) \cap \{0\} \right\},$$

In this case $\mathrm{ah}(A)$ is the affine hull [21] of the set $A \subseteq \mathrm{R}^J$. A method such as the Euclidean Empirical Likelihood [3], or some other member of the CR class of estimators operate within the modified E^3 approach (abbreviated mE^3) to provide a basis for estimation and inference.

2 Empty set problem

The replacement of $\Phi(\Theta)$ by $\Phi_q(\Theta)$ may seem natural[1]. However, this substitution is not as innocent as it may appear. Because $\Phi_q(\Theta)$ is data-dependent, it can be empty for some data and therefore may be subject to the *empty set problem* (ESP). In other words, ESP introduces the possibility that for the sample size n, there

---

[1]In particular, when it is phrased in context of the Generalized Minimum Contrast estimation, cf. [17], [2].

may exist $X_1^n \sim r_X(x;\theta)$ such that $\Phi_q(\Theta) = \emptyset$. Consequently for such a model the very existence of an estimator and corresponding inferences that are obtained by methods that operate within E^3 are data-dependent.

In the section ahead, it will be demonstrated that models considered in Qin and Lawless' seminal paper [20] (abbreviated QL) are subject to the empty set problem. A few other models are used to illustrate the problem and provide a basis for further discussions.

There are models, such as QL, Example 1, where even the modified E^3 approach is subject to an empty set problem. In this case the empty set problem concerns the possibility that for the sample size n, there may exist $X_1^n \sim r_X(x;\theta)$ such that $\Phi_q^m(\Theta) = \emptyset$. This form of ESP will be referred to as the *affine Empty Set Problem* (aESP). Whenever there is the affine ESP for a data set, then there is also ESP; the opposite need not be true.

Two notes are worth making in advance:

1) The empty set problem is a substantive extension of the convex hull problem, that is known in the literature on the Empirical Likelihood; cf. [19], see also Sect. 2.3.1. Indeed, the convex hull problem can be seen as the empty set problem, in the case when $\Theta = \{\theta_0\}$, i.e., when the parametric space comprises a single point. It is well-known that in such case the set $\Phi_q(\theta_0)$ may be empty, for some data set, and some θ_0. Our objective is to show, for some models there are data, for which $\Phi_q(\Theta)$ may be empty, for the entire Θ .

2) Another substantive point concerns the affine form $\Phi_q^m(\Theta)$ of the model. We show that even in this case there are models that are subject to the affine empty set problem. It means, that the modifications of MELE, which are designed to mitigate the convex hull problem by expanding it into the affine hull, cannot, in general, serve as a rescue from the empty $\Phi_q^m(\Theta)$ set problem.

2.1   QL, Example 1

QL, Example 1, p. 301 and pp. 309-311, consider the following setup: there is a random variable X, with pdf $r(X;\theta)$, where $\theta \in \Theta = \mathrm{R}$, and a random sample $X_1^n = X_1,\ldots,X_n$, drawn from $r_X(x;\theta)$. A researcher specifies a pair of estimating functions

$$u_1(X;\theta) = X - \theta,$$
$$u_2(X;\theta) = X^2 - (2\theta^2 + 1).$$

The estimating functions, through unbiased estimating equations, define a set of pdf's (i.e., the model) $\Phi(\Theta) = \bigcup_{\theta \in \Theta} \Phi(\theta)$ where $\Theta = \mathrm{R}$ and

$$\Phi(\theta) = \left\{ f_X(x;\theta) : \mathrm{E}u_1(X;\theta) = 0; \mathrm{E}u_2(X;\theta) = 0 \right\},$$

into which, in the researcher's view, the 'true' sampling distribution should belong. In this case $J > K$ and the model is over-identified.

As noted in the Introduction, the Empirical Estimation Approach (E^3) proceeds

by replacing the set $\Phi(\theta)$ by its empirical analogue

$$\Phi_q(\theta) = \left\{ q(x;\theta) : \sum_{i=1}^{n} q(x_i;\theta)u_1(x_i;\theta) = 0; \sum_{i=1}^{n} q(x_i;\theta)u_2(x_i;\theta) = 0; \right.$$
$$\left. \sum_{i=1}^{n} q(x_i;\theta) = 1; q(x_i;\theta) \geq 0, 1 \leq i \leq n \right\}.$$

First, a condition on data X_1^n under which the set $\Phi_q(\Theta)$ is empty, will be derived. To this end, note that the two empirical estimating equations $\sum_{i=1}^{n} q(x_i;\theta)$ $(x_i - \theta) = 0$, $\sum_{i=1}^{n} q(x_i;\theta)(x_i^2 - (2\theta^2 + 1)) = 0$ can be lumped into a single equation

$$\sum_{i=1}^{n} q(x_i;\cdot)x_i^2 - 2\left(\sum_{i=1}^{n} q(x_i;\cdot)x_i\right)^2 = 1. \tag{3}$$

The question is whether for an observed sample $X_1^n = x_1^n$ the value on the Right-Hand Side of (3) can be attained, for some $q(\cdot,\cdot)$. If not, then for the data x_1^n the set $\Phi_q(\Theta)$ is empty. It can be seen that the expression on the Left-Hand Side of (3) can attain its maximal value for such a $q(x;\cdot)$ that the only non-zero elements of $q(x;\cdot)$ are $q_{(1)} \triangleq q(x_{(1)};\cdot)$ and $q_{(n)} \triangleq q(x_{(n)};\cdot)$; there $x_{(1)}$ denotes the lowest, $x_{(n)}$ the largest value in x_1^n. The value $\hat{q}_{(1)}$ of $q_{(1)}$ for which the LHS of (3) attains its maximum is given as

$$\hat{q}_{(1)} = \arg\max_{q_{(1)} \in [0,1]} L(q_{(1)}), \tag{4a}$$

where

$$L(q_{(1)}) \triangleq q_{(1)}x_{(1)}^2 + (1 - q_{(1)})x_{(n)}^2 - 2(q_{(1)}x_{(1)} + (1 - q_{(1)})x_{(n)})^2. \tag{4b}$$

Without a $[0,1]$ constraint on the range of the values that $q_{(1)}$ can take on, the optimal value $\hat{q}_{(1)}^m$ is

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

and $\hat{q}_{(n)}^m = 1 - \hat{q}_{(1)}^m$. Without the $[0,1]$-range constraint the maximal value of the LHS of (3) is thus given by plugging $\hat{q}_{(1)}^m$ into (4b); let us denote it by v. Under the range constraint the maximal value of the LHS of (3) can be only smaller or equal to v. If the data $X_1^n \sim r_X(x;\theta)$ are such that v is smaller than 1 (i.e., the RHS of (3)) then $\Phi_q^m(\Theta)$ as well as $\Phi_q(\Theta)$ is empty for such data. Consequently, for such data there is no mE^3-based or E^3-based estimator.

For a given $r(X;\theta)$ and n, we are interested in the probability $\Pr(v < 1)$ that the LHS maximal value is smaller than 1, the RHS value of (3); i.e., the probability that $\Phi_q^m(\Theta)$ (as well as $\Phi_q(\Theta)$) is empty. The probability can be estimated by means of a Monte Carlo simulation from $r(X;\theta)$. For $r(X;\theta)$ being $n(0,1)$ and $n = 15$, as in QL, the Monte Carlo estimate of the probability is 0.0173, based on $M = 10000$ samples. Thus, 17 of 1000 samples of size $n = 15$ drawn from $n(0,1)$ are such that $\Phi_q^m(\Theta)$ (as well as $\Phi_q(\Theta)$) is empty, and hence it is meaningless in these samples to look for EL, or any other E^3-based (or mE^3-based) estimate.

QL performed an MC study of small-sample properties of MELE, with the aim of comparison with other estimators (sample mean and ML). The above results

indicate that the numbers in Table 1 of QL are meaningless. Since MELE does not exist if $\Phi_q(\Theta)$ is empty, this puts into question the comparison of MELE with other competitive estimators. Indeed, the entire Empirical Estimating Equations approach is questionable, except of in asymptotics[2], since the feasible set $\Phi_q(\Theta)$ becomes 'always' non-empty as for $n \to \infty$. The same holds for mE^3.

The setup considered by QL in their Example 1 is a simple one. Its simplicity permits the empty set problem of E^3 to be illustrated analytically. E^3 approach, and its most common instance - the Maximum Empirical Likelihood Estimation, - are commonly studied with more complicated models, where the question whether $\Phi_q(\Theta)$ is non-empty for a particular data, may be much harder to answer. Anyway, the very possibility that an estimator based on E^3 exists for some data and does not exist for a different data, puts this estimation method in question. The same can be said about mE^3 approach.

2.2   QL, Example 2

In the second QL example, there is a random sample $(X, Y)_1^n$ of bivariate observations, such that $\mathrm{E}(X) = \mathrm{E}(Y) = \theta$, $\theta \in \Theta = \mathrm{R}$. The authors suggest that we use the bivariate estimating function $u(x, y; \theta) = (X - \theta, Y - \theta)$ and estimate θ by MELE. However, due to the problem of the empty set $\Phi_q(\Theta)$ this is not always possible. For instance, let $X \sim n(0, \sigma_1^2)$, $Y \sim n(0, \sigma_2^2)$. Then $\Phi_q(\Theta)$ will be empty for every sample, such that $X_i - Y_i > 0$, or $X_i - Y_i < 0$, for all $i = 1, \ldots, n$, as Qin and Lawless [20] note. For $n = 10$ the probability is $2(0.5)^{10} = 0.002$. The probability rises, as the model becomes misspecified. For instance, assume that $X \sim n(-0.3, 0.1)$ and $Y \sim n(0, 0.1)$, then $X - Y \sim n(-0.3, 0.2)$ and the probability $\Pr(X - Y < 0) = 0.749$, so that $\Pr(X - Y > 0) = 0.251$. Consequently, for $n = 10$, the probability that Φ_q is empty is $0.749^{10} + 0.251^{10} = 0.056$.

2.3   QL, Example 3

The third QL example concerns the selection of the 'representative member' from

$$\Phi(\Theta = \{a\}) = \left\{ f_X(x; \theta) : \mathrm{E}_f u(X; \theta) = 0 \right\},$$

where $a \in \mathrm{R}$ is known, and there is a random sample X_1^n from unknown $r_X(x)$. In the E^3 approach, the model $\Phi(a)$ is replaced by its empirical analogue

$$\Phi_q(a) = \left\{ q(x; a) : \sum_{i=1}^n q(x_i; a) u(x_i; a) = 0, \sum_{i=1}^n q(x_i; a) = 1, q(x_i; a) \geq 0, \forall i \right\},$$

and a pmf is selected from $\Phi_q(a)$ by some method, such as MELE, or Minimum Discriminant Information [12]. Even in this simple setting the empty set problem appears. For the sake of illustration, let us assume that $r_X(x) = [0.025, \ 0.025, \ 0.15, \ 0.8]$ is a pmf on $\mathscr{X} = \{1, 2, 3, 4\}$. Let $u(X) = X - a$, $a = 2.0$ and let the sample size $n = 40$. The probability that $\Phi_q(2)$ is empty is the probability that the sample contains only values greater than 2, or only values smaller than 2, which is 0.129.

---

[2] The point that MELE is an asymptotic method has already been made in [10], albeit from a different point of view.

2.3.1 EL confidence intervals and tests

The empty set problem in the setting of this Example also has bearing for constructing confidence intervals and tests by the method of Empirical Likelihood (EL). Although we are concerned here with MELE estimation and inference, it is worth noting that EL inference is also undermined by the empty set problem. For instance, assume - in the context of the above Example - that the null hypothesis $H_0 : \mathrm{E}X = \mu = 2.0$ is to be tested. EL test and confidence interval are based on the Non-parametric Likelihood Ratio statistic (cf. [19]), which involves computation of the value of the Non-parametric Likelihood at $\mu = 2.0$. If the data are such that $\Phi_q(2.0)$ is empty then no confidence interval or test can be performed. In the above setting this happens with probability 0.129, for the data of size $n = 40$.

Existence of the empty set problem in the construction of confidence intervals and statistical tests by EL is not new, and in [19] (cf. Sect. 3.14, Sect 10.4, Chap. 12, among others), it is referred to as the convex hull condition. Recently, Chen, Variyath and Bovas [4] suggested an adjustment of EL, with the aim of mitigating the problem. The authors suggest adding to the data an additional observation which is the negative multiple of the sample average. The multiplication constant serves to rescue EL from the empty set problem. However, the value of the constant which achieves the goal is also data dependent.

2.4 Estimation of location parameter from noisy data

Motivated by [11], consider the following data-generating process: $Y = X + \epsilon$, where $X \sim \mathrm{Exp}(1)$, $\epsilon \sim n(0, \sigma^2)$ and $X \perp \epsilon$. A researcher observes a random sample Y_1^n and would like to estimate the location parameter θ of the distribution $f_X(x; \theta)$ of X. Since neither $f_X(x; \theta)$ nor the distribution of ϵ are known, the researcher chooses an over-determined model based on the first two moments of Y:

$$\Phi(\theta) = \{f_Y(y; \theta) : \mathrm{E}(Y - \theta) = 0; \mathrm{E}(Y^2 - (2\theta^2 + \delta^2) = 0\},$$

and $\theta \in \Theta = (0, \infty)$. To make it operative, $\Phi(\theta)$ is replaced by its empirical analogue $\Phi_q(\theta)$, in the usual way. Observe that the model is just a modification of Example 1 of QL, so the argument of Sect. 2.1 can be employed to find probability that in this setting $\Phi_q(\Theta)$ is empty. For $\sigma = 3$ and $n = 100$ the probability is 0.215. This is another illustration that the ESP is not only a problem that applies to small samples.

2.5 Mean and median

Brown and Chen [3] investigated the problem of estimating a location parameter of a sampling distribution by a data-based combination of the mean and the median. Their empirical model is

$$\Phi_q(\theta) = \left\{q(x; \theta) : \sum_{i=1}^{n} q(x_i; \theta)(x_i - \theta) = 0; \sum_{i=1}^{n} q(x_i; \theta)\mathrm{sgn}(x_i - \theta) = 0; \right.$$

$$\left. \sum_{i=1}^{n} q(x_i; \cdot) = 1; q(x_i; \cdot) \geq 0, \, 1 \leq i \leq n \right\},$$

and $\theta \in \Theta = \mathrm{R}$. In their model $J > K$ and thus is over-determined. Unless Θ is restricted to a subspace of R (cf. Sect. 2.8), there is no problem of empty $\Phi_q(\Theta)$ set, and hence MELE always exists in this case. However, EL confidence interval/test need not exist for every θ.

2.6   Score E³

Let $r_X(x;\theta)$ and Θ be such that the Maximum Likelihood (ML) estimator $\hat{\theta}_{\mathrm{ML}}$ of θ is identical with the solution of the score equations. If $\Phi(\Theta)$ is defined by estimating functions which are based on the score equations, then the corresponding $\Phi_q(\Theta)$ is free of the empty set problem, for any random sample X_1^n drawn from $r_X(x;\theta)$. Any of the GMC estimators is then identical with $\hat{\theta}_{\mathrm{ML}}$. A trivial example is given by the gaussian $n(0,1)$ sampling distribution, where the score equation for the location parameter θ is $\frac{1}{n}\sum_{i=1}^{n} X_i - \theta = 0$. Then for the estimating function $u(X;\theta) = X - \theta$ the set $\Phi_q(\Theta)$, where $\Theta = \mathrm{R}$, cannot be empty for any X_1^n from $n(0,1)$. Changing Θ into the halfline $[0,\infty)$ gives rise to ESP; cf. Sect. 2.8.

2.7   Discrete random variable

It is clear that the empty set problem is not a peculiarity of a continuous random variable. As an illustration, consider an over-identified model given as the set of pmfs which satisfy the estimating equations for the estimating functions $u_1(x;\theta) = X - \theta$, $u_2(x;\theta) = X^2 - \theta^2 - \theta$, $\theta \in \Theta = [0,\infty)$, and defined on the support $\mathscr{X} = \{0,\ldots,\infty\}$. Its empirical analogue $\Phi_q(\theta)$ is empty for any data X_1^n, for which

$$q_{(1)}x_{(1)}^2 + (1 - q_{(1)})x_{(n)}^2 - (q_{(1)}x_{(1)} + (1 - q_{(1)})x_{(n)})^2 - q_{(1)}x_{(1)} - (1 - q_{(1)})x_{(n)} < 0,$$

where $q_{(1)} = 0.5 + 0.5/(x_{(n)} - x_{(1)})$. This follows from the same reasoning as in Sect. 2.1. For the Poisson sampling distribution Poi(1), the probability that for a sample of size $n = 10$ there will be no empirical estimator is 7/1000 (estimated by 10000 runs of MC).

2.8   Restricted parameter space

Let the data-sampling distribution $r_X(x;\theta)$ be parametrized by a parameter θ which can take on any value in $\Theta = \mathrm{R}$. Assume, for the sake of simplicity, the exactly identified model specified by a single estimating function $u(X;\theta) = X - \theta$, and by $\theta \in \tilde{\Theta} = (a,b)$, a, b are finite. The model is misspecified in the sense that $\tilde{\Theta} \subset \Theta$. In E³ approach, the model $\Phi(\tilde{\Theta})$ is replaced by its empirical analogue $\Phi_q(\tilde{\Theta})$. The probability that $\Phi_q(\tilde{\Theta})$ is empty depends on $r_X(x;\theta)$, the values of a, b, and the sample size n. It is not difficult to construct an illustrative example. Let the sampling distribution be $n(0,1)$, $n = 100$, $\tilde{\Theta} = (2,3)$. The probability that the set $\Phi_q(\tilde{\Theta})$ is empty is essentially determined by the probability that all the observed values are below 2, which is 0.100. An extreme case of this setting was considered in Sect. 2.3, where $\tilde{\Theta}$ comprised a single element.

There is also other possibility to get an empty $\Phi_q(\Theta)$, by restricting the parameter space Θ of the sampling distribution. As an illustration, assume the gaussian $n(\theta,1)$ sampling distribution, with $\theta \in \Theta = [0,\infty)$, and the model let be given by the estimating function $u(X;\theta) = X - \theta$. Whenever the sample X_1^n is such that all its elements are smaller than 0, the parametric space restriction cannot be satisfied;

cf. also [19], Sect. 10.5. Note, that for such a sample the parametric space restriction can be satisfied within the modified E^3, and hence, there is no affine ESP in this model.

The two possibilities are pertinent also to the over-identified (i.e., $J > K$) and under-identified (i.e., $J < K$) models. As an example of the under-determined model consider the one given by the estimating function $u(X; \theta, \sigma^2) = (X - \theta)^2 - \sigma^2$, where θ, σ^2 are parameters of the sampling distribution; $\theta \in \Theta = \mathrm{R}$, $\sigma^2 \in \Sigma = (0, \infty)$. Let the space of values which both θ and σ^2 can take on be restricted in the model, in such a way that the resulting set of distributions is non-empty. The corresponding empirical counterpart of the model can be empty for any data which, regardless of q, cannot satisfy restrictions of the parametric space.

3   Summary and discussion

As a way of summing up the statistical implications of the examples we make the following comments:

1) There are models $\Phi(\Theta)$ and data-sampling distributions $r_X(x; \theta)$ for which the corresponding empirical counterpart $\Phi_q(\Theta)$ may be empty for some data X_1^n from $r_X(x; \theta)$, and non-empty for others.

2) Whether a particular model is subject to the empty set problem (ESP) is, in general, not easy to discern[3]. For the simple models considered here, we were able to address the existence of the empty set problem analytically.

3) If an E^3 is subject to ESP, then the probability that the set $\Phi_q(\Theta)$ is empty depends on the set, the sample size and the underlying sampling distribution $r_X(x; \theta)$.

4) Examples 1,2,3 from Qin and Lawless [20] are subject to ESP.

5) There are models for which the E^3 approach is not affected by the empty set problem (cf. Sect. 2.5, 2.6 for examples).

6) The convex hull condition resulting from construction of Empirical Likelihood confidence intervals and moment restriction tests, is a special case of the empty set problem. This is true in the sense that it concerns calculation of MELE when Θ comprises (usually) a single point (cf. Sect. 2.3).

7) Once $\Phi_q(\Theta)$ is subject to ESP, any method which is used to select a $q(x; \theta)$ from it, breaks down, for any data for which $\Phi_q(\Theta)$ is empty.

8) Over-identified, exactly-identified as well as under-identified models can be subject to the empty $\Phi_q(\Theta)$ set problem.

9) There are models (cf. Sect. 2.1 for an example) where the modified E^3 approach is subject to the affine ESP. For such models, the existence of estimators that are obtained by methods, such as the Euclidean Empirical Likelihood, that operate within the modified E^3 (mE^3) approach, can be data-dependent.

In general, the Empirical Estimating Equations (E^3) approach (the modified Empirical Estimating Equations, mE^3) is undermined by the empty set problem (the affine empty set problem), as are also methods of estimation and inference – such as MELE (the Euclidean Empirical Likelihood), – that operate within it. Some possible responses to the empty set problem of the E^3/mE^3 approach are:

---

[3]For instance, in context of the Example 1 of QL (cf. Sect. 2.1) data for which the set $\Phi_q(\Theta)$ is empty can lead to a meaningfully looking MELE estimate. However, a review of the 'optimal' result of the inner optimization in (2) will reveal that it failed. On the other hand, difficulty with the inner optimization cannot be taken as a sign that the set $\Phi_q(\Theta)$ is empty, for the data in hand.

- Abandon E^3/mE^3. An option, then, is to return back to the Estimating Equations and use the Semi-parametric Bayesian approach, cf. [8], [10].

- Use the empirical estimating equations, and satisfy them approximately, only. The Generalized Method of Moments [13] follows this route. Recall that the GMM estimator θ_{GMM} of θ is given as a solution of

$$\hat{\theta}_{\text{GMM}}(W) = \arg\min_{\theta \in \Theta} \left[\frac{1}{n} \sum_{i=1}^{n} u(x_i; \theta) - 0 \right]' W \left[\frac{1}{n} \sum_{i=1}^{n} u(x_i; \theta) - 0 \right],$$

where W is a positive definite symmetric weight matrix. The objective of GMM is to find a θ that satisfies the equations $\frac{1}{n} \sum u(x_i; \theta) = 0$ approximately, as best as possible, where goodness of the approximation is measured by the weighted euclidean distance. Thus, GMM is not a method that operates within mE^3/E^3-approach. The same holds for the Continuous updating GMM [14].

- Stay with E^3 (mE^3), but use it only when $\Phi_q(\Theta)$ ($\Phi_q^m(\Theta)$) is non-empty.

4 Acknowledgements

References

[1] Back, K., and Brown, D. (1990). Estimating distributions from moment restrictions. Working paper, Graduate School of Business, Indiana University.

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

[3] Brown, B. M., and Chen, S. X. (1998). Combined and Least Squares Empirical Likelihood. *Ann. Inst. Statist. Math*. 90:443-450.

[4] Chen, J., Variyath, A. M., and Abraham, B. (2008). Adjusted Empirical Likelihood and its properties. *J. Comput. Graph. Stat.*. 17/2:426-443.

[5] Corcoran, S. A. (2000). Empirical exponential family likelihood using several moment conditions. *Stat. Sinica*. 10:545-557.

[6] Corcoran, S. A. (1998). Bartlett adjustment of empirical discrepancy statistics. *Biometrika*. 85:967-972.

[7] Cressie, N., and Read, T. (1984). Multinomial goodness of fit tests. *JRSS Ser. B*. 46:440-464.

[8] Florens, J.-P., and Rolin, J.-M. (1994): Bayes, Bootstrap, Moments. Discussion paper 94.13. Institute de Statistique, Université catholique de Louvain.

[9] 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.

[10] Grendár, M., and Judge, G. (2009). Asymptotic equivalence of Empirical Likelihood and Bayesian MAP. *Ann. Statist.* 37/5A:2445-2457.

[11] Gzyl, H., and Ter Horst, E. (2009). Recovering decay rates from noisy measurements with Maximum Entropy in the Mean. *J. Prob. Stat.* Article ID 563281, 13 pages, doi:10.1155/2009/563281.

[12] Haberman, S. (1984). Adjustment by Minimum Discriminant Information. *Ann. Statist.* 12/3:971-988.

[13] Hansen, L. P. (1982). Large sample properties of Generalized Method of Moments estimators. *Econometrica*. 50:1029-1054.

[14] Hansen, L.P., Heaton, J., and Yaron, A. (1996). Finite-sample properties of some alternative GMM estimators. *J. Bus. Econ. Stat.* 14:262-280.

[15] Imbens, G. W., Spady, R. H., and Johnson, P. (1998). Information theoretic approaches to inference in moment condition models. *Econometrica*. 66/2:333-357.

[16] Kitamura, Y., and Stutzer, M. (1997). An information-theoretic alternative to Generalized Method of Moments estimation. *Econometrica*. 65:861-874.

[17] Kitamura, Y. (2007). Nonparametric Likelihood: efficiency and robustness. *Jap. Econ. Rev.*. 58/1:26-45.

[18] Mittelhammer, R., Judge, G. and Miller, D. (2000). *Econometric Foundations*. Cambridge:Cambridge University Press.

[19] Owen, A. B. (2001). *Empirical Likelihood*. New York:Chapman-Hall/CRC.

[20] Qin, J., and Lawless, J. (1994). Empirical Likelihood and General Estimating Equations. *Ann. Statist.* 22:300-325.

[21] Rockafellar. R. T. (1970). *Convex Analysis*. New Jersey: Princeton University Press.

[22] Schennach, S. M. (2005). Bayesian exponentially tilted empirical likelihood. *Biometrika*. 92/1:31-46.