Chin. Ann. Math.
33B(6), 2012, 941–948
DOI: 10.1007/s11401-012-0740-1

Chinese Annals of Mathematics, Series B

Empirical Likelihood Approach for Treatment Effect in Pretest-Posttest Trial*

Qixiang HE[1]

Abstract The empirical likelihood approach is suggested to the pretest-posttest trial based on the constrains, which we construct to summarize all the given information. The author obtains a log-empirical likelihood ratio test statistic that has a standard chi-squared limiting distribution. Thus, in making inferences, there is no need to estimate variance explicitly, and inferential procedures are easier to implement. Simulation results show that the approach of this paper is more efficient compared with ANCOVA II due to the sufficient and appropriate use of information.

Keywords Empirical likelihood, Pretest-posttest trial, Treatment effect
2000 MR Subject Classification 62G05

1 Introduction

The pretest-posttest trial is an important and popular method to evaluate treatment effects in medicine, public health and numerous other fields. In a pretest-posttest trial, the responses of interest are measured both at the baseline and at follow-ups. The baseline responses are served as a basis for comparison with the follow-up responses. Commonly, subjects are randomly assigned to the two groups: the treatment group and the control group. The objective is to evaluate whether the treatment affects the follow-up responses.

There are several methods that were proposed to estimate the treatment difference or to test that there is no difference. These methods include the two-sample t-test, the paired t-test, the analysis of covariance I (ANCOVA I), the analysis of covariance II (ANCOVA II), the generalized estimating equations (GEE), etc. Yang and Tsiatis [9] discussed the asymptotic properties of the estimators and their relative efficiencies. They showed that all these estimators are consistent and asymptotically normal, and the GEE estimator is asymptotically equivalent to that by ANCOVA II. They are the most efficient estimator. Later, Leon, Tsiatis and Davidian [3] took a semiparametric perspective without assumptions about the distributions of baseline and posttest responses. They derived the class of all consistent treatment effect estimators, and gave the form of influence function for this class of estimators based on the observed data. Looking into the form of the influence function, we find that the form is actually the combination of some estimating equations. This motivates us to research on how to use all the given useful

Manuscript received February 17, 2011. Revised August 7, 2012.
[1]Department of Applied Mathematics, Shanghai University of Finance and Economics, Shanghai 200433, China. E-mail: qxhe@mail.shufe.edu.cn
*Project supported by the National Natural Science Foundation of China (No. 10971033).

information instead of only some of it, to transfer the information into the estimating equations, and then to combine the estimating equations efficiently.

In this paper, the empirical likelihood approach is used in the pretest-posttest trial based on the constrains, which we construct to summarize all the given useful information instead of only some of it. The empirical likelihood was proposed by Owen [4–5] for constructing generalized likelihood ratio test statistics and the corresponding confidence region. It was motivated by an earlier work of Thomas and Grunkmeier [8], which provided an approach to construct the confidence interval for the survival probability through the constrained likelihood ratio. Owen generalized the idea and showed that Wilks theorem of chi-squared limiting distribution for the log-likelihood ratio statistic still holds for the empirical likelihood. The approach is well recognized to possess several advantages including the range preserving, the transformation respecting, the data decided shape for the confidence region and the implicit studentizing without the need to estimate the variance explicitly. Thus, the approach was extended to many areas in statistics. For example, Owen [6] and Chen [1] discussed the empirical likelihood for linear regression analysis. Chen and Hall [2] studied the method on quantile estimation. Qing and Lawless [7] explored the empirical likelihood, and the optimal linear combination of estimating equations. They linked estimating equations and the empirical likelihood, and developed methods of combining information about parameters.

In the pretest-posttest trial, in order to try to use any given information, over-constraints for the parameter of treatment effects are inevitable. So, combining information efficiently is important. Based on the equations which we will construct, we obtain a log-empirical likelihood ratio test statistic that has a standard chi-squared limiting distribution. Thus, in making inferences, there is no need to estimate variance explicitly, and inferential procedures are easier to implement. Moreover, more importantly, simulation results show that our approach is more efficient compared with ANCOVA II due to the sufficient and appropriate use of information.

This paper is organized as follows. In Section 2, we set up the notion, describe the model and introduce the method. In Section 3, our proposed empirical likelihood ratio test statistic is defined, and we discuss the inference about the treatment effect. In Section 4, we present some Monte Carlo simulation results and compare them with ANCOVA II.

2 Notations and Model

Let n be the total number of subjects in the trial, and each subject be randomized to "control" or "treatment", with known probabilities $(1-\delta)$ and δ, respectively. Accordingly, define $A_i = 0$ or 1 for subject i. Let Y_{1i} and Y_{2i} be $i's$ observed baseline and follow-up responses, respectively, leading to observed data for $i(Y_{1i}, Y_{2i}, A_i)$, where the subscript i is suppressed when no ambiguity will result. The variables Y_1 and A represent phenomena prior to the treatment action, while Y_2 is a post-treatment characteristic. Thus, let $Y_2^{(0)}$ and $Y_2^{(1)}$ be the follow-up responses, whose subjects would potentially exhibit if they are assigned to control and treat, respectively. We assume that subject i is assigned to the treatment group with probability δ independent of their baseline response Y_{1i}. We place no restrictions on the joint distribution of (Y_2, Y_1). Given A_i, we allow the conditional distribution of (Y_{2i}, Y_{1i}) to be arbitrary and subject to the existence of the first two moments. Specifically, let $Y_i = (Y_{2i}, Y_{1i})^{\mathrm{T}}$ be the

outcome vector, and its conditional expectation be

$$\mathrm{E}(Y_i \mid A_i) = \begin{pmatrix} \mu_2 + \beta A_i \\ \mu_1 \end{pmatrix}, \quad i = 1, 2, \cdots, n. \tag{2.1}$$

The conditional expectation of the response at the baseline does not depend on the treatment indicators A_i, because of the randomization assumption. The parameter β represents the treatment effect, which is the difference between the mean of the follow-up response of the treatment group and that of the control group, and is the parameter of primary interest.

The outcome vector is assumed to have finite second moments. The second moments are parameterized as

$$\mathrm{Var}(Y_i \mid A_i = 1) = \begin{pmatrix} \sigma_{22}^{(1)} & \sigma_{12}^{(1)} \\ \sigma_{12}^{(1)} & \sigma_{11} \end{pmatrix} \tag{2.2}$$

and

$$\mathrm{Var}(Y_i \mid A_i = 0) = \begin{pmatrix} \sigma_{22}^{(0)} & \sigma_{12}^{(0)} \\ \sigma_{12}^{(0)} & \sigma_{11} \end{pmatrix}. \tag{2.3}$$

We suppose that $X_i = (Y_{1i}, Y_{2i}, A_i)^{\mathrm{T}}$ are independent and identically distributed across i. Yang and Tsiatis [9] discussed the properties of five methods and therefore five estimators for the treatment effect β. Leon, Tsiatis and Davidian [3] found that all RAL estimators for β based on the observed data have an influence function of the form

$$\frac{A(Y_2 - \mu_2 - \beta)}{\delta} - \frac{(1 - A)(Y_2 - \mu_2)}{1 - \delta} + (A - \delta)h(Y_1), \tag{2.4}$$

where h is arbitrary with $\mathrm{Var}\{h(Y_1)\} < \infty$. All the estimators in [9] have influence functions of the form

$$\frac{A(Y_2 - \mu_2 - \beta)}{\delta} - \frac{(1 - A)(Y_2 - \mu_2)}{1 - \delta} + (A - \delta)(Y_1 - \mu_1)\eta \tag{2.5}$$

with the different η respectively. By (2.5), these estimators are all in class (2.4) with $h(Y_1) = \eta(Y_1 - \mu_1)$.

Both (2.4) and (2.5) are the combinations of three terms. It is easy to show that each term has zero mean and reflects different given information about the parameter vector $\theta = (\mu_1, \mu_2, \beta)^{\mathrm{T}}$. The first two terms correspond to the information of (2.1). The third term corresponds to the information of the independence between A and Y_1.

Now, we define

$$g^{(1)}(X, \theta) = Y_1 - \mu_1, \tag{2.6}$$
$$g^{(2)}(X, \theta) = \tfrac{(1-A)Y_2}{1-\delta} - \mu_2, \tag{2.7}$$
$$g^{(3)}(X, \theta) = \tfrac{AY_2}{\delta} - (\mu_2 + \beta), \tag{2.8}$$
$$g^{(4)}(X, \theta) = A(Y_1 - \mu_1), \tag{2.9}$$

where $X = (Y_1, Y_2, A)^{\mathrm{T}}$. By (2.1), we have

$$\mathrm{E}g^{(i)}(X, \theta) = 0, \quad i = 1, 2, 3. \tag{2.10}$$

We also notice that

$$\mathrm{E}g^{(4)}(X, \theta) = 0 \tag{2.11}$$

if and only if A and Y_1 are independent, since A is an indicator variable. (2.6)–(2.9) give us four estimating equations about three parameters μ_1, μ_2 and β. Apparently, it is an over-constrained problem.

3 Empirical Likelihood Method

Since each subject datum $X_i = (Y_{1i}, Y_{2i}, A_i)^{\mathrm{T}}$ can be viewed as an observation from unknown distribution F, we have that there is a parameter θ associated with F, and the datum is a random sample of n individual. The empirical likelihood function can be written as

$$\prod_{i=1}^{n} P_i$$

with suitable constraints, because

$$\mathrm{E}g(X, \theta) = 0, \tag{3.1}$$

where $g(X, \theta) = (g^{(1)}(X, \theta), g^{(2)}(X, \theta), g^{(3)}(X, \theta), g^{(4)}(X, \theta))^{\mathrm{T}}$. We introduce the constraint

$$\sum_{i=1}^{n} P_i g(X_i, \theta) = 0, \tag{3.2}$$

as well as the standard unit total probability constraint $\sum\limits_{i=1}^{n} P_i = 1$. To obtain the confidence regions for θ, we define the empirical likelihood ratio function

$$R(\theta) = \sup\Big\{ \prod_{i=1}^{n} nP_i \,\Big|\, \sum_{i=1}^{n} P_i g(X_i, \theta) = 0,\ \sum_{i=1}^{n} P_i = 1, P_i \geq 0 \Big\}. \tag{3.3}$$

As noted by Owen [4–5], the unique value $\widehat{\theta} = \big(\widehat{\mu}_1, \widehat{\mu}_2, \widehat{\beta}\big)^{\mathrm{T}}$ for the right-hand side of (3.3) exists. By a Langrage multiplier argument, an explicit expression for $R(\theta)$ can be derived (see [7]) as

$$R(\theta) = \prod_{i=1}^{n} \frac{1}{1 + t^{\mathrm{T}}(\theta) g(X, \theta)},$$

where $t(\theta)$ is a 4-dimensional vector given as the solution to

$$\sum_{i=1}^{n} (1 + t^{\mathrm{T}}(\theta) g(X_i, \theta))^{-1} g(X_i, \theta) = 0.$$

Define an empirical likelihood function for θ as

$$L(\theta) = \prod_{i=1}^{n} \Big\{ \frac{1}{n} \cdot \frac{1}{1 + t^{\mathrm{T}}(\theta) g(X_i, \theta)} \Big\}.$$

So the empirical log-likelihood ratio is

$$l(\theta) = \sum_{i=1}^{n} \log(1 + t^{\mathrm{T}}(\theta) g(X_i, \theta)).$$

Apparently, $\widehat{\theta}$ maximizes $L(\theta)$ and minimizes $l(\theta)$.

According to [7, Theorem 1], by direct calculation, we have the following result.

Lemma 3.1 *If θ_0 is the true value of θ, then*

$$\sqrt{n}(\widehat{\theta}-\theta) \xrightarrow{\mathcal{L}} N(0,V),$$

where

$$V=\Big[\mathrm{E}\Big(\frac{\partial g}{\partial\theta}\Big)(\mathrm{E}gg^{\mathrm{T}})^{-1}\Big(\frac{\mathrm{E}g}{\partial\theta}\Big)\Big]^{-1},$$

$$\frac{\partial g}{\partial\theta}=\begin{pmatrix}-1&0&0\\0&-1&0\\0&-1&-1\\-\delta&0&0\end{pmatrix},$$

and

$$\mathrm{E}gg^{\mathrm{T}}=\begin{pmatrix}\sigma_{11}^2&\sigma_{12}^{(0)}&\sigma_{12}^{(1)}&\delta\sigma_{11}^2\\ \sigma_{12}^{(0)}&\dfrac{\sigma_{22}^{(0)}}{1-\delta}+\dfrac{\delta}{1-\delta}\mu_2^2&-\mu_2(\mu_2+\beta)&0\\ \sigma_{12}^{(1)}&-\mu_2(\mu_2+\beta)&\dfrac{\sigma_{22}^{(1)}}{\delta}+\dfrac{1-\delta}{\delta}(\mu_2+\beta)^2&\sigma_{12}^{(1)}\\ \delta\sigma_{11}^2&0&\sigma_{12}^{(1)}&\delta\sigma_{11}^2\end{pmatrix}.$$

It is clear that, in order to use the given information sufficiently, we constructed more estimating equations than the number of parameters. As pointed by Qin and Lawless [7], the maximized empirical likelihood estimator (MELE) $\widehat{\theta}$ based on $g(X,\theta)$ is fully efficient, in the sense that it has the same asymptotic variance as the optiomal estimator obtained from the estimating equations that are linear combinations of $g^{(1)}(X,\theta)$, $g^{(2)}(X,\theta)$, $g^{(3)}(X,\theta)$ and $g^{(4)}(X,\theta)$.

Theorem 3.1 *For the test $H_0:\beta=\beta_0$, the profile empirical likelihood ratio test statistic is*

$$W=2l(\widetilde{\mu}_1,\widetilde{\mu}_2\beta_0)-2l(\widehat{\mu}_1,\widehat{\mu}_2,\widehat{\beta}),$$

where $\widetilde{\mu}_1$ and $\widetilde{\mu}_2$ minimize $l(\mu_1,\mu_2,\beta_0)$ with respect to μ_1 and μ_2, respectively. Under H_0, $W\sim\chi_1^2$, as $n\to\infty$.

This theorem can be easily obtained by applying the corollary in [7]. We can use the empirical likelihood ratio statistics for testing or obtaining confidence limits for the treatment effect.

Actually, in practice, δ is often unknown. We may replace δ with $\widehat{\delta}=\frac{1}{n}\sum\limits_{i=1}^{n}A_i$. Then the estimating functions are asymptotically unbiased. Based on this fact, we define

$$\widetilde{g}(X,\theta)=\begin{pmatrix}\widetilde{g}^{(1)}(X,\theta)\\ \widetilde{g}^{(2)}(X,\theta)\\ \widetilde{g}^{(3)}(X,\theta)\\ \widetilde{g}^{(4)}(X,\theta)\end{pmatrix},$$

where

$$\widetilde{g}^{(2)}(X,\theta)=\frac{(1-A)Y_2}{1-\widehat{\delta}}-\mu_2,$$

$$\widetilde{g}^{(3)}(X,\theta)=\frac{AY_2}{\widehat{\delta}}-(\mu_2+\beta)$$

and

$$\widetilde{R}(\theta) = \sup\Big\{ \prod_{i=1}^{n} nP_i \Big| \sum_{i=1}^{n} P_i \widetilde{g}(X_i, \theta) = 0,\ \sum_{i=1}^{n} P_i = 1,\ P_i \geq 0 \Big\}.$$

Then

$$-2\log \widetilde{R}(\theta) = -2\sum_{i=1}^{n} (1 + \widetilde{t}^{\mathrm{T}}(\theta)\widetilde{g}(X_i, \theta)),$$

and $\widetilde{t}(\theta)$ satisfies

$$\sum_{i=1}^{n} \big(1 + \widetilde{t}^{\mathrm{T}}(\theta)\widetilde{g}(X_i, \theta)\big)^{-1} \widetilde{g}(X_i, \theta) = 0.$$

Similarly, by [7], we obtain

$$\begin{aligned} -2\log \widetilde{R}(\theta) = {} & \Big[\frac{1}{n}\sum_{i=1}^{n} \widetilde{g}(X_i, \theta_0)^{\mathrm{T}}\Big] \cdot \Big[\frac{1}{n}\sum_{i=1}^{n} \widetilde{g}(X_i, \theta_0)\widetilde{g}(X_i, \theta_0)^{\mathrm{T}}\Big]^{-1} \\ & \cdot \Big[\frac{1}{n}\sum_{i=1}^{n} \widetilde{g}(X_i, \theta_0)\Big] + o(1), \quad \text{a.s.} \end{aligned}$$

On one hand,

$$\frac{1}{n}\sum_{i=1}^{n} \widetilde{g}(X_i, \theta_0)\widetilde{g}(X_i, \theta_0)^{\mathrm{T}} \xrightarrow{\text{a.s.}} \mathrm{E}gg^{\mathrm{T}}.$$

On the other hand, although the asymptotic distribution of $\frac{1}{n}\sum\limits_{i=1}^{n} \widetilde{g}(X_i, \theta_0)^{\mathrm{T}}$ is normal distributed with mean zero, the asymptotic variance-covariance matrix is not $\mathrm{E}g\widetilde{g}^{\mathrm{T}}$. This yields that $-2\log \widetilde{R}(\theta_0)$ has a weighted chi-square limiting distribution instead of a standard chi-squared distribution. It does not follow the standard asymptotic result for the empirical likelihood method. In making inference about the treatment effect, one has to estimate the weights.

Now we remodel the problem as follows. When δ is unknown, we define $g^{(5)}(X, \theta^*) = A - \delta$, where $\theta^* = (\mu_1, \mu_2, \beta, \delta)$. Define

$$g^*(X, \theta^*) = \begin{pmatrix} g^{(1)}(X, \theta^*) \\ g^{(2)}(X, \theta^*) \\ g^{(3)}(X, \theta^*) \\ g^{(4)}(X, \theta^*) \\ g^{(5)}(X, \theta^*) \end{pmatrix},$$

where $g^{(1)}$ $g^{(2)}$, $g^{(3)}$, $g^{(4)}$ are the same as (2.6)–(2.9), except that we treat δ as an unknown parameter. We define

$$l^*(\theta^*) = \sum_{i=1}^{n} \log(1 + t^{*T}(\theta^*) g^*(X_i, \theta^*)),$$

where $t^*(\theta^*)$ is a 5-dimensional vector satisfying

$$\sum_{i=1}^{n} (1 + t^{*T}(\theta^*) g^*(X_i, \theta^*))^{-1} g(X_i, \theta^*) = 0.$$

Similar to Theorem 3.1, we have the following result.

Theorem 3.2 *For the test $H_0 : \beta = \beta_0$, the profile empirical likelihood ratio test statistic is*

$$W^* = 2l^*(\widetilde{\mu}_1^*, \widetilde{\mu}_2^*, \beta_0, \widetilde{\delta}^*) - 2l^*(\widehat{\mu}_1^*, \widehat{\mu}_2^*, \widehat{\beta}^*, \widehat{\delta}^*)$$

for unknown δ, where $\widetilde{\mu}_1^*$, $\widetilde{\mu}_2^*$, $\widetilde{\delta}^*$ minimize $l^*(\mu_1, \mu_2, \beta_0, \delta)$. Then under H_0,

$$W^* \sim \chi_1^2, \quad \text{as } n \to \infty.$$

Theorems 3.1–3.2 allow us to use the empirical likelihood ratio statistic for testing or obtaining confidence limits for the treatment effect with both the known δ and the unknow δ. If we know more information, we can construct and add more constraints accordingly, and therefore improve the estimate accuracy of β.

4 Numerical Studies

To examine the finite sample performance of the proposed empirical likelihood approach, with a view comparing with ANCOVA II, we conduct a series of simulation. The following three cases are considered.

(C1) The baseline responses Y_{1i} are generated from the exponential distribution with mean 2. For the control group, we generate the follow-up responses $Y_{2i}^{(0)}$ from the exponential distribution with mean 3, while for the treatment group, we generate the follow-up responses $Y_{2i}^{(1)}$ from the exponential distribution with mean 4.

(C2) Y_{1i} are generated from the Weibull population, whose density functions are

$$p(x) = \begin{cases} \alpha\lambda x^{\lambda-1}\mathrm{e}^{-\lambda x^\alpha}, & x > 0, \\ 0, & x \le 0 \end{cases}$$

with $\alpha = 0.5$, $\lambda = \frac{1}{2}$. For the control group, the follow-up responses $Y_{2i}^{(0)}$ are generated from the Weibull population with $\alpha = 0.4$, $\lambda = \frac{1}{2}$, and the follow-up responses of the treatment group $Y_{2i}^{(1)}$ are generated from the Weibull population with $\alpha = 0.2$, $\lambda = \frac{1}{2}$.

(C3) Y_{1i}, $Y_{2i}^{(0)}$ and $Y_{2i}^{(1)}$ are generated from the Log-normal distribution, whose density functions are

$$p(x) = \begin{cases} \dfrac{1}{\sigma x\sqrt{2\pi}}\mathrm{e}^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}, & x > 0, \\ 0, & x \le 0. \end{cases}$$

For the baseline responses, the parameters are chosen as $\mu_1 = 1$, $\sigma_1^2 = 1$. For the control group, the parameters are chosen as $\mu_2 = 1.5$, $\sigma_2^2 = 1$, while for the treatment group, the parameters are chosen as $\mu_3 = 2$, $\sigma_3^2 = 1$.

In every case, the subjects in the trial are assigned randomly to the control group and the treatment group with probabilities $1 - \delta = 0.45$ and $\delta = 0.55$, respectively. The true value of the treatment effect β is 1. We vary the sample size n. For each case, 100 sets of data are generated to simulate the type I error, the nominal level of which is chosen to be 0.05. The simulation results are summarized in Tables 1–3.

From the simulation results, the empirical likelihood method performs better than ANCOVA II, especially when the sample size is small.

Table 1 Coverage probabilities of confidence interval of treatment effect β in case (C1).

sample size (n)	ANCOVA II	empirical likelihood
30	0.74	0.94
50	0.94	0.95
80	0.81	0.96
100	0.89	0.95

Table 2 Coverage probabilities of confidence interval of treatment effect β in case (C2).

sample size (n)	ANCOVA II	empirical likelihood
30	0.74	0.93
50	0.88	0.95
80	0.84	0.94
100	0.91	0.95

Table 3 Coverage probabilities of confidence interval of treatment effect β in case (C3).

sample size (n)	ANCOVA II	empirical likelihood
30	0.81	0.94
50	0.77	0.94
80	0.89	0.96
100	0.89	0.95

During the simulating process, we also find that when the sample size is small, e.g., 30 or 50, the estimators of variance of the treatment effect β by ANCOVA II are sometimes negative. So the proposed empirical likelihood approach performs well for practical finite sample sizes.

References

[1] Chen, S. X., Empirical likelihood confidence intervals for linear regression coefficients, *J. Multivariate Anal.*, **49**, 1994, 24–40.

[2] Chen, S. X. and Hall, P., Smoothed empirical likelihood confidence intervals for quantiles, *Ann. Staatist.*, **21**, 1993, 621–637.

[3] Loen, S., Tsiatis, A. A. and Davidian, M., Semiparametric estimation of treatment effect in a pretest-protese study, *Biometrics*, **59**, 2003, 1046–1055.

[4] Owen, A. B., Empirical likelihood ratio confidence intervals for single functional, *Biometrika*, **75**, 1988, 237–249.

[5] Owen, A. B., Empirical likelihood ratio confidence region, *Ann. Statist.*, **18**, 1990, 90–120.

[6] Owen, A. B., Empirical likelihood for linear model, *Ann. Statist.*, **19**, 1991, 1725–1747.

[7] Qin, J. and Lawless, J. F., Empirical likelihood and general estimating equations, *Ann. Statist.*, **22**, 1994, 300–325.

[8] Thomas, D. R. and Grunkemeier, G. L., Confidence interval estimation of survival probabilities for censored data, *J. Amer. Statist. Assoc.*, **70**, 1975, 865–871.

[9] Yang, L. and Tsiatis, A. A., Efficiency study of estimators for a treatment effect in a pretest-posttest trial, *Amer. Statist.*, **4**, 2001, 314–321.