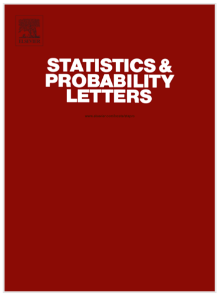
# Empirical likelihood for linear models in the presence of nuisance parameters

Mi-Ok Kim, Mai Zhou

# Accepted Manuscript




Empirical likelihood for linear models in the presence of nuisance parameters

Mi-Ok Kim, Mai Zhou

PII: S0167-7152(07)00428-2
DOI: 10.1016/j.spl.2007.12.007
Reference: STAPRO 4860

To appear in: *Statistics and Probability Letters*

Received date: 20 June 2006
Revised date: 28 August 2007
Accepted date: 11 December 2007

Please cite this article as: Kim, M.-O., Zhou, M., Empirical likelihood for linear models in the presence of nuisance parameters. *Statistics and Probability Letters* (2007), doi:10.1016/j.spl.2007.12.007

# Empirical Likelihood for Linear Models in the Presence of Nuisance Parameters

Mi-Ok Kim*and Mai Zhou
Center for Epidemiology and Biostatistics, Cincinnati Children's Medical Center
Cincinnati, OH 45229-3039 and
Department of Statistics, University of Kentucky
Lexington, KY 40506

**Abstract** We propose a simple alternative empirical likelihood (EL) method in linear regression which requires the same conditions of the ordinary profile EL but overcomes the challenge of maximizing the likelihood in the presence of high dimensional nuisance parameters. We adapt the idea of added variable plots. We regress the response and the independent variables of main interest on the ancillary variables and construct the likelihood based on the residuals. The hence constructed EL ratio has constraints only pertains to the parameters of interest and has a standard $\chi^2$ limiting distribution. Numerical results are included.

*Key words and phrases* Profile empirical likelihood, Added variable plots

## 1 Introduction

Empirical likelihood (EL) method is a general and effective nonparametric inference method. It can be seen as bootstrap that does not re-sample and assigns probability mass not necessarily evenly yet to every sample points under linear constraints. It has been extensively studied in the literature and has many application (see a monograph by Owen (2001) and reference therein for examples). We focus on EL application for linear regression in the presence of 'nuisance' parameters in this paper.

Inference in the presence of 'nuisance' parameters is an old problem where attention is often

*Corresponding author email: miok.kim@cchmc.org

focused on only one or two of the parameters, the other being considered 'nuisance' parameters necessary to characterize scientific problems but of no intrinsic interest. With parametric approaches, 'nuisance' parameters are often scale parameters in a linear model setting. On the contrary, nonparametric approaches do not usually necessitate scale 'nuisance' parameters unless prior information suggests variance modelling. Hence in this paper we use 'nuisance' parameters to refer to a part of regression coefficient vector that is not of an immediate interest.

Inference in linear regression is often concerned with the coefficient parameters individually or a subset of a few. Inference for a part of the parameter vector is particularly important with EL as it is not easy to characterize the confidence region for parameters of dimension $> 2$, since EL produces irregularly shaped confidence regions that are known only convex. Hence ellipsoidal properties cannot be used to characterize the EL confidence regions and a common practice is instead give two dimensional confidence regions or confidence intervals for subsets of parameters.

In such cases a standard approach that is analogous to its parametric counterpart is profiling the EL (e.g. Owen, 1991). Qin and Lawless (1994) demonstrated in a general estimating equation (GEE) framework that the profiling method yields EL ratio statistic with a standard $\chi^2$ limiting distribution. Chen (1994) investigated the profile EL for a linear regression where the intercept and slope are scalar. In spite of its nice theoretical properties, however, profiling EL is problematic in practice as its computational burden increases exponentially as the dimension of the nuisance parameters grows. In particular, confidence regions/intervals estimation is problematic as the EL ratio statistic is used inversely to decide whether a point under consideration to be included or not. For example, when the regression involves five independent variables and confidence intervals are to be estimated, the profiling method requires maximizing the likelihood over a five dimensional nuisance parameter vector (including intercept) for each value under consideration.

To overcome the limitation placed by the nuisance parameters, some have proposed plug-in method that replaces the nuisance parameters with their estimate. For example, recently Hjort, McKeague and Van Keilegom (2006) showed that the basic theorem of EL can be generalized to allow for plug-in estimates of nuisance parameters in the estimating equations. In linear regression a natural candidate for the estimates is the least squares. However, plugging in the least squares yields EL ratio statistic whose limiting distribution is not a standard $\chi^2$ and has to be calibrated via bootstrapping, which offsets the merit of the EL method. DiCiccio and Monti (2001) proposed two approximations to the profile empirical likelihood function for a scalar parameter of interest in the context of $M$-estimation using the third and fourth derivatives of the profile log empirical likelihood function at its maximizing point. Their method can be seen corresponding with various additive adjustments of the profile likelihood function in the parametric case, for example, as recent as Stern (1997) and Ferrari, Lucambio and Cribari-Neto (2005). The EL inference based on the approximation is no longer invariant under reparameterization.

We propose a simple alternative approach in case of linear regression that applies under the same conditions of the ordinary profile EL, invariant to reparametrization, and overcomes the computational challenges of the profiling method. As well known, added variable plots are devices to represent the effect of an independent variable on the regression after accounting for the effects of the other variables in the model. Two-dimensional plots are most common but the idea can be generalized to a higher dimension to jointly represent the effects of more than one independent variables on the regression after accounting for the effects of the other variables, although the results may be hard to visualize. For example, Cook (1994) discusses three-dimensional added variable plots. We adapt the idea to make the EL inference only pertain to the parameters of an immediate interest. We regress the response and the independent variables of main interest on the ancillary variables and construct the likelihood based on the residuals. The constraints of the EL only pertain to the parameters of interest. The resulting EL inference inherits some of the properties of ordinary EL: the likelihood is maximized at the same point where the ordinary profile EL is maximized and invariant to reparametrization, and $-2\log$ EL ratio is asymptotically $\chi^2$ distributed.

The rest of paper is organized as follows. Section 2 introduces the proposed method and section 3 presents numerical results. Proofs of the main results are deferred to section 4.

## 2 Partial Residuals and EL

For a vector $\nu$, we use $\nu_{(k)}$ to denote the $k$th element. Consider a linear model

$$y_i = x_i^\top \beta_0 + e_i \,, 1 \le i \le n \,,$$

where $e_i$ are mean zero independent random variables with finite variances $\sigma_i^2$. Suppose that the covariates and the regression coefficients are divided such that $x_i = (x_{1i}^\top, x_{2i}^\top)^\top \in R^p$ and $\beta_0 = (\beta_{10}^\top, \beta_{20}^\top)^\top$ with $x_{1i}$, $\beta_{10} \in R^{p_1}$, $x_{2i}$, $\beta_{20} \in R^{p_2}$ and $x_{1i(1)} = 1$. We are interested in a hypothesis test of $H_0 : \beta_2 = \beta_{20}$ versus $H_1 : \beta_2 \ne \beta_{20}$. The ordinary EL method defines the likelihood as

$$\mathbb{EL}(\beta_1, \beta_2) = \max_{w_i} \{ \prod_{i=1}^n w_i | \sum_{i=1}^n w_i z_i(\beta_1, \beta_2) = 0, \sum_{i=1}^n w_i = 1, w_i \ge 0 \} \,, \tag{1}$$

where $z_i(\beta_1, \beta_2) = x_i(y_i - x_{1i}^\top \beta_1 - x_{2i}^\top \beta_2)$. Then the profile EL method defines the ratio function for $\beta_2$ as $\mathcal{R}(\beta_2) = \max_{\beta_1 \in R^{p_1}} \mathbb{EL}(\beta_1, \beta_2) / \max_{\beta_1 \in R^{p_1}, \beta_2 \in R^{p_2}} \mathbb{EL}(\beta_1, \beta_2)$. As shown in Owen (1991) the denominator likelihood is maximized at $\prod_{i=1}^n 1/n$. Hence the profile EL ratio becomes

$$\mathcal{R}(\beta_2) = \max_{\beta_1 \in R^{p_1}} \{ \prod_{i=1}^n n w_i | \sum_{i=1}^n w_i z_i(\beta_1, \beta_2) = 0, \sum_{i=1}^n w_i = 1, w_i \ge 0 \} \,. \tag{2}$$

Its computation is usually done via iterative algorithms or a grid search algorithm. The computational challenges placed by the free parameter $\beta_1$ is inhibiting even when $p_1$ is as low as 3, while quite commonly regression analysis involves more than three independent variables.

When the true value of $\beta_1$ is known, we can avoid profiling using the following EL and EL ratio:

$$\widetilde{\mathbb{EL}}(\beta_2) = \max_{w_i} \{ \prod_{i=1}^n w_i | \sum_{i=1}^n w_i \tilde{z}_i(\beta_2) = 0, \sum_{i=1}^n w_i = 1, w_i \ge 0 \} \,, \tag{3}$$

where $\tilde{z}_i(\beta_2) = x_{2i}(y_i - x_{1i}^\top \beta_{10} - x_{2i}^\top \beta_2)$ and

$$\widetilde{\mathcal{R}}(\beta_2) = \max_{w_i} \{ \prod_{i=1}^n n w_i | \sum_{i=1}^n w_i \tilde{z}_i(\beta_2) = 0, \sum_{i=1}^n w_i = 1, w_i \ge 0 \} \,. \tag{4}$$

However, $\beta_{10}$ is not always known and we propose following alternative method based on the idea of added variable plots. Define matrices $X_{1n} = (x_{11}, \cdots, x_{1n})^\top$, $X_{2n} = (x_{21}, \cdots, x_{2n})^\top$, $H_n = X_{1n}(X_{1n}^\top X_{1n})^{-1} X_{1n}^\top$ and $Y_n = (y_1, \cdots, y_n)^\top$. We regress $x_{2i}$ and $y_i$ on $x_{1i}$ and denote the respective residuals by $x_{2i}^*$ and $y_i^*$. Specifically $x_{2i}^* = x_{2i} - X_{2n}^\top h_i$ and $y_i^* = y_i - Y_n^\top h_i$, where $h_i$ denote the $i$th column of $H_n$. With $(x_{2i}^*, y_i^*)$, we have an induced linear model $y_i^* = x_{2i}^{*\top}\beta_2 + e_i^*$. We define the likelihood based on $(x_{2i}^*, y_i^*)$ as

$$\mathbb{EL}^*(\beta_2) = \max_{w_i}\{\prod_{i=1}^n w_i | \sum_{i=1}^n w_i z_i^*(\beta_2) = 0, \sum_{i=1}^n w_i = 1, w_i \geq 0\}, \tag{5}$$

where $z_i^*(\beta_2) = x_{2i}^*(y_i^* - x_{2i}^{*\top}\beta_2)$. Accordingly we define the EL ratio as

$$\mathcal{R}^*(\beta_2) = \max_{w_i}\{\prod_{i=1}^n n w_i | \sum_{i=1}^n w_i z_i^*(\beta_2) = 0, \sum_{i=1}^n w_i = 1, w_i \geq 0\}. \tag{6}$$

We note that the optimization problem in (6) is free of the nuisance parameter $\beta_1$. We have the following results.

Some notation is needed. Let $mineig(V)$ denote the minimum eigenvalue of the symmetric matrix $V$ and $||\nu||$ denote the $L_2$ norm of a vector $\nu$. Let $ch(A)$ denote the convex hull of the set $A \subseteq R^p$.

**Theorem 1** *Let $n_0 \geq p$, $\gamma \geq 0$, $a, b > 0$, and let $N = \{x_i | y_i - x_i^\top \beta_0 < 0\}$ and $P = \{x_i | y_i - x_i^\top \beta_0 > 0\}$. Assume that $ch(N) \cap ch(P) \neq \emptyset$ with probability tending to 1 as $n \to \infty$. Also assume that $\frac{1}{n^2}\sum_{i=1}^n ||x_i||^4 E(e_i^4) \to 0$. Suppose $a < \sigma_i < b||x_i||^\gamma$ for all $i$ and that for all $n \geq n_0$, $a < mineig(X_n^\top X_n/n)$ and $(1/n)\sum_{i=1}^n ||x_i||^{2+\gamma} < b$. Then, $-2\log \mathcal{R}^*(\beta_{20}) \to \chi^2_{p_2}$ in distribution as $n \to \infty$.*

Note that the conditions of the above theorem is same as those of corollary 2 of Owen (1991). Hence this alternative EL method works under the same requirements of the profile EL method.

It follows from the theorem that for any $0 < \alpha < 1$ an EL confidence region for $\beta_{20}$ with an asymptotic coverage probability $1 - \alpha$ is given by $\{\beta_2 | -2\log \mathcal{R}^*(\beta_2) < c_{1-\alpha}\}$ where $c_{1-\alpha}$ is defined such that $P(\chi^2_{p_2} > c_{1-\alpha}) = \alpha$. Also it can be shown trivially that $\hat{\beta}_2^* = \hat{\beta}_2$ as NPMLEs (non-parametric maximum likelihood estimate) in linear regression are ordinary least squares estimates, where $\hat{\beta}_2$ and $\hat{\beta}_2^*$ denote an NPMLE of $\beta_2$ with respect to the EL defined in (1) and (5). This implies that while the proposed EL is different from the one by the profile EL, they provide the same point estimate and the confidence intervals are centered at the same value in the sense that the p-values are 1 at the same point. Moreover, the confidence regions and intervals are almost identical as examples show later. An asymptotic level $\alpha$ test can be specified by using the complement of the confidence region as a rejection region.

# 3 Simulation Results

## 3.1 Simulation study 1

We present results that empirically verify the results of Theorem 1. We use following simulation setup where $M1$ and $M2$ represent homoscedastic and heteroscedastic errors model:

$$M1 : y_i = 3 + 2z_{1i} + z_{2i} + z_{3i} + z_{4i} + z_{5i} + e_i$$

$$M2 : y_i = 3 + 2z_{1i} + z_{2i} + z_{3i} + z_{4i} + z_{5i} + 0.3(s_{1i} + s_{2i} + s_{3i} + s_{4i} + s_{5i})e_i$$

where $z_{1i} \sim \chi^2_1$, $z_{2i} \sim N(-1, 1)$, $z_{3i} \sim \exp(1)$, $z_{4i} \sim N(1, 1)$, $z_{5i} \sim \chi^2_3$ and $z_{ji}$, $j = 1, \cdots, 5$ are independent of one another, and $s_{ji}$ are truncated $z_{ji}$ so that $|s_{ji}| \leq 5$. Two distributions, $N(0, 1)$ and $\exp(1) - 1$ are considered for $e_i$. In each of 5000 simulated samples we considered two hypotheses: $\beta_{20} = 2$ and $\beta_{20} = (2, 1)^\top$ where the covariate(s) of main interest is/are $z_{1i}$ and $(z_{1i}, z_{2i})$ and in the notation of the previous section $x_{2i} = z_{1i}$ and $x_{2i} = (z_{1i}, z_{2i})^\top$ respectively. We calculated $-2 \log \mathcal{R}^*(\beta_{20})$ for each $\beta_{20}$.

We note that the simulation setup reflects quite a common situation in regression that analysis concerning even mildly sophisticated problems would involve four or five independent variables, while only one or two of them are of main interest. Even with the nuisance parameters of a moderate dimension as low as three or four, computing (2) is challenging and the usual profile EL method may be deemed impractical. The proposed alternative method comes to its rescue. As the computational burden of the usual profile EL is inhibiting, we instead compare the proposed alternative EL with $\widetilde{\mathcal{R}}(\beta_2)$. All the computations were conducted using *emplik* package in R.2.0.1.

Q-Q plots in figure 1 and 2 empirically validate the results of Theorem 1 for homoscedastic and heteroscedastic errors model. As the errors in the heteroscedastic model are functions of all the five independent variables, the heteroscedasticity in the model (M2) is strong. Therefore the empirical distributions of $\mathcal{R}^*(\beta_{20})$ show a rather weak agreement with the respective $\chi^2$ asymptotic reference distributions in figure 2. This also explains less than desirable performance of the proposed method for the heteroscedasticity model reported in table 1. The performance can be improved certainly with increasing sample size.

Table 1 show that the proposed EL method performs comparably to a method with the EL in (3) and that the asymptotic results in Theorem 1 can be used to construct level appropriate tests or confidence intervals. Note that the EL method with the EL in (3) should be best as true values are plugged in for $\beta_1$ in the EL (3). Hence the comparable performance of the proposed method is all that can be desired.

### 3.2 Simulation study 2

This study empirically examines the properties of the proposed likelihood and confidence intervals constructed. We simulate one sample of $n = 100$ from the following setup and construct a 95% confidence interval for the slope coefficient: $y_i = 3 + 2z_{1i} + e_i$, where $z_1 \sim \chi^2_1$ and $e_i \sim N(0, 1)$. Figure 3 presents two $-2\log$ EL ratio functions. We note that $-2\log\mathcal{R}^*(\beta_{20})$ (solid line) and $-2\log\mathcal{R}(\beta_2)$ (dashed line) are almost identical and both reach the minimum 0 at the same value. This empirically verifies that their NPMLEs for the slope coefficient are same.

## 4 Appendix

*Proof of Theorem 1* As we have $\hat{\beta}^*_2 = \hat{\beta}_2$, $\hat{\beta}^*_2$ is a consistent estimator of $\beta_2$.

Let $\tilde{e}_n = (e_1, \cdots, e_n)^\top$ and $\Sigma_n = E_e(\tilde{e}_n \tilde{e}_n^\top) = diag(\sigma_i^2)$ where $E_e$ denotes the conditional expectation given $(x_{1i}, x_{2i})$. There exists a positive definite matrix $S$ under the conditions of the theorem such that $S = \lim_{n\to\infty} \frac{1}{n} X_{2n}^\top (I_n - H_n)\Sigma_n (I_n - H_n) X_{2n}$. Define $\bar{Z}^* = \frac{1}{n}\sum_{i=1}^n z_i^*(\beta_{20})$ and $S_n = \frac{1}{n}\sum_{i=1}^n z_i^*(\beta_{20}) z_i^{*\top}(\beta_{20})$.

Let $a_n \approx b_n$ mean that there are constants $0 < A < B < \infty$ such that $A \leq a_n/b_n \leq B$ for all $n$ and let $\lim_{n\to\infty} n^{-1} X_{1n}^\top X_{1n} = I$ without loss of generality. Then, $||x_{2i}^*|| \approx ||x_{2i} - (n^{-1} X_{2n}^\top X_{1n}) x_{1i}||$ except on the events whose probability tends to 0 with increasing $n$. Note that

$$||x_{2i} - (n^{-1} X_{2n}^\top X_{1n}) x_{1i}|| \leq ||x_{2i}|| + \max_j ||(n^{-1} X_{2n}^\top X_{1n})_{(j)}|| \, ||x_{1i}|| \approx ||x_{2i}|| + ||x_{1i}|| \, ,$$

where $(n^{-1} X_{2n}^\top X_{1n})_{(j)}$ denotes the $j$th row of $(n^{-1} X_{2n}^\top X_{1n})$. Also for all $1 \leq i, j \leq n$, $H_{n(i,j)} \approx n^{-1} x_{1i}^\top x_{1j} = o(1)$, where $H_{n(i,j)}$ denote the $i,j$th element of $H_n$. Hence $mineig(X_{2n}^\top (I - H_n) X_n/n) \approx mineig(X_{2n}^\top X_{2n}/n)$. It follows that under the conditions of the theorem, $\frac{1}{n^2}\sum_{i=1}^n ||x_{2i}^*||^4 E(e_i^4) \to 0$ and there exist positive constants $a^*$ and $b^*$ such that $a^* < mineig(X_{2n}^\top (I - H_n) X_n/n)$ for all $n \geq n_0$, and $(1/n)\sum_{i=1}^n ||x_{2i}^*||^{2+\gamma} < b^*$. Therefore, $x_{2i}^* e_i$, $1 \leq i \leq n$, satisfy the Lindberg's condition under the conditions of the theorem as well as $x_i e_i$ of the original model do. As

$$\sqrt{n}\bar{Z}^* = \frac{1}{\sqrt{n}} X_{2n}^\top (I_n - H_n)(I_n - H_n)^\top \tilde{e}_n = \frac{1}{\sqrt{n}} \sum x_{2i}^* e_i \, ,$$

$\sqrt{n}\bar{Z}^*$ is asymptotically normally distributed with mean 0 and variance $S$ by the multivariate central limit theorem.

On the other hand, $S_n = \frac{1}{n}\sum x_{2i}^* x_{2i}^{*\top}(e_i^2 - 2e_i \tilde{e}_n^\top h_i + (\tilde{e}_n^\top h_i)^2)$, where $e_i \tilde{e}_n^\top h_i \approx \frac{1}{n}\sum_{k=1}^n e_i e_k x_{1k}^\top x_{1i}$ and $(\tilde{e}_n^\top h_i)^2 \approx (\frac{1}{n}\sum_{k=1}^n e_k x_{1k}^\top x_{1i})^2$ except on the events whose probability tends to 0 with increasing

$n$. For all $i$,

$$E_e(\frac{1}{n}\sum_{k=1}^{n} e_i e_k x_{1k}^{\top} x_{1i}) = \frac{1}{n}\sigma_i^2 \|x_{1i}||^2 = o(1)$$
$$E_e\left(\frac{1}{n}\sum_{k=1}^{n} e_i e_k x_{1k}^{\top} x_{1i}\right)^2 = \frac{1}{n^2}\{E_e(e_i^4)||x_{1i}||^4 + \sum_{k\neq i}\sigma_i^2\sigma_k^2(x_{1k}^{\top}x_{1i})^2\} = o(1)$$
$$E_e\left(\frac{1}{n}\sum_{k=1}^{n} e_k x_{1k}^{\top} x_{1i}\right)^2 = \frac{1}{n^2}\sum_{k=1}^{n}\sigma_k^2(x_{1k}^{\top}x_{1i})^2 = o(1)\,.$$

We have $|e_i\tilde{e}_n^{\top}h_i| = o_p(1)$ and $(\tilde{e}_n^{\top}h_i)^2 = o_p(1)$ by the Chebychev's and Markov's inequality respectively. Also $\frac{1}{n}\sum x_{2i}^* x_{2i}^{*\top} e_i^2 \to S$ by the Chebychev's inequality and it follows that $S_n \to S$ in probability. The rest of the proof follows similar arguments in the proof of theorem 2 of Owen (1991).

## 5 Reference

Chen, S. X., (1994), Empirical likelihood confidence intervals for linear regression coefficients, *J. Multivariate Anal.* **49**, 24-40.

Cook, R. D. and Weisberg, S. (1994), *An introduction to regression graphics*, John Wiley & Sons, Inc., New York, NY.

DiCiccio, T. J. and Monti, A. C. (2001), Approximations to the profile empirical likelihood function for a scalar parameter in the context of $M$-estimation, *Biometrika*, **88**, 337-351.

Ferrari, S. L. P., Lucambio, F. and Cribari-Neto, F. (2004), Improved profile likelihood inference, *J. Stat. Plan. Infer.*, **134**, 373-391.

Hjort, N. L., McKeague, I. W., and Van Keilegom, I., (2006), Extending the scope of empirical likelihood, *Ann. Statist.* under revision.

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

Owen, A. B., (2001), *Empirical likelihood* (Chapman & Hall/CRC, London)

Qin, J. and Lawless, J., (1994), Empirical likelihood and general estimating equations, *Ann. Statist.* **22**, 300-325.

Stern, S. E. (1997), A second-order adjustment to the profile likelihood in the case of a multidimensional parameter of interest, *J. R. Statist. Soc. B*, **59**, 653-665.

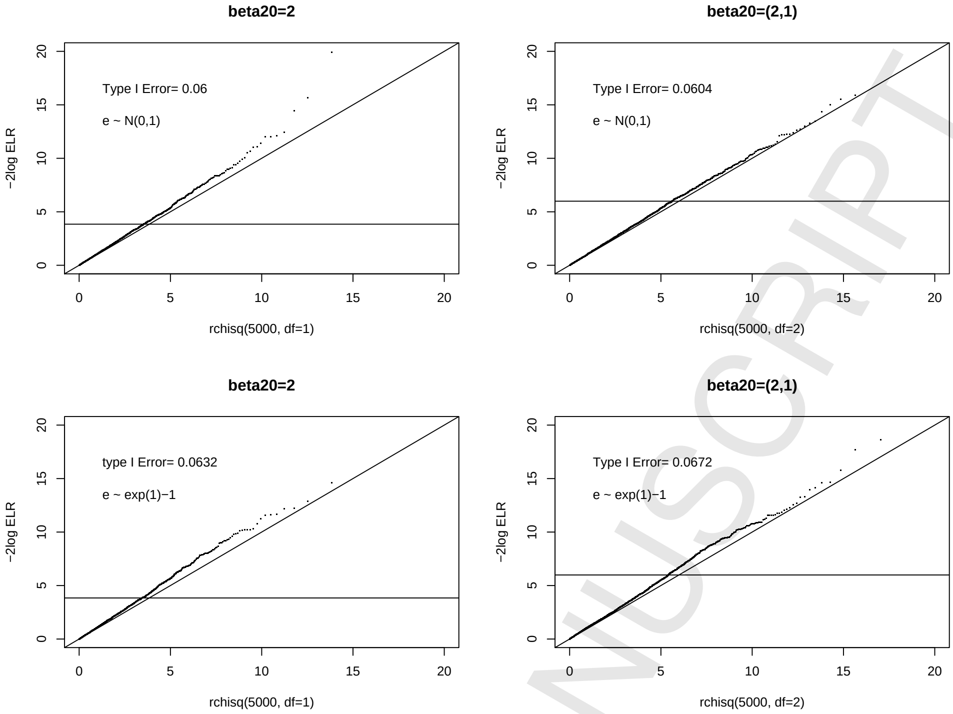



Figure 1: Q-Q plots of $-2\log \mathcal{R}^*(\beta_{20})$ and quantiles of respective $\chi^2$ reference distributions for homoscedastic errors model (M1) based on 5000 simulations with $n = 400$.

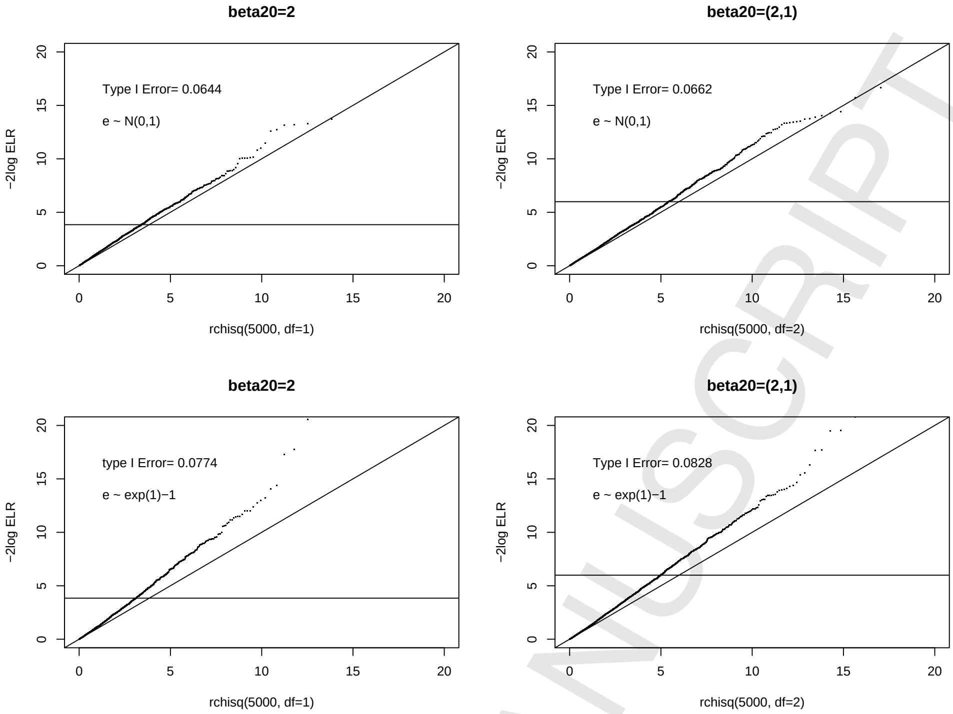



Figure 2: Q-Q plots of $-2\log\mathcal{R}^*(\beta_{20})$ and quantiles of respective $\chi^2$ reference distributions for heteroscedastic errors model (M2) based on 5000 simulations with $n = 400$.

Table 1: Observed Type I Error Rates for Homoscedastic (M1) and Heteroscedastic Errors Model (M2): $\mathcal{R}^*$ indicates the results by the proposed method and $\tilde{\mathcal{R}}$ by the method with true values plugged in for the nuisance parameters.

| M1 | $H_0$ | | $\beta_{20} = 2$ | | | | $\beta_{20} = (2, 1)$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\mathcal{R}^*$ | | $\widetilde{\mathcal{R}}$ | | $\mathcal{R}^*$ | | $\widetilde{\mathcal{R}}$ | |
| $e_i$ | $n$ | $\alpha$ | 0.1 | 0.05 | 0.1 | 0.05 | 0.1 | 0.05 | 0.1 | 0.05 |
| | 75 | | 0.1478 | 0.0840 | 0.1338 | 0.0776 | 0.1792 | 0.1064 | 0.1612 | 0.1010 |
| $N(0, 1)$ | 100 | | 0.1424 | 0.0802 | 0.1284 | 0.0710 | 0.1576 | 0.0962 | 0.1464 | 0.0808 |
| | 200 | | 0.1320 | 0.0742 | 0.1164 | 0.0588 | 0.1414 | 0.0806 | 0.1302 | 0.0730 |
| | 75 | | 0.1660 | 0.0966 | 0.1474 | 0.0842 | 0.1892 | 0.1154 | 0.1806 | 0.1132 |
| $\exp(1) - 1$ | 100 | | 0.1466 | 0.0836 | 0.1334 | 0.0736 | 0.1752 | 0.1060 | 0.1686 | 0.1002 |
| | 200 | | 0.1250 | 0.0676 | 0.1094 | 0.0566 | 0.1452 | 0.0876 | 0.1368 | 0.0776 |
| M2 | $H_0$ | | $\beta_{20} = 2$ | | | | $\beta_{20} = (2, 1)$ | | | |
| | | | $\mathcal{R}^*$ | | $\widetilde{\mathcal{R}}$ | | $\mathcal{R}^*$ | | $\widetilde{\mathcal{R}}$ | |
| $e_i$ | $n$ | $\alpha$ | 0.1 | 0.05 | 0.1 | 0.05 | 0.1 | 0.05 | 0.1 | 0.05 |
| | 75 | | 0.1922 | 0.1246 | 0.1664 | 0.1044 | 0.1950 | 0.1312 | 0.1704 | 0.1094 |
| $N(0, 1)$ | 100 | | 0.1716 | 0.1076 | 0.1498 | 0.0892 | 0.1848 | 0.1144 | 0.1668 | 0.1002 |
| | 200 | | 0.1328 | 0.0746 | 0.1156 | 0.0606 | 0.1482 | 0.0848 | 0.1342 | 0.0768 |
| | 75 | | 0.2128 | 0.1394 | 0.1800 | 0.1164 | 0.2352 | 0.1588 | 0.2054 | 0.1370 |
| $\exp(1) - 1$ | 100 | | 0.1890 | 0.1180 | 0.1560 | 0.0930 | 0.2148 | 0.1342 | 0.1910 | 0.1230 |
| | 200 | | 0.1442 | 0.0858 | 0.1214 | 0.0648 | 0.1628 | 0.0984 | 0.1470 | 0.0870 |

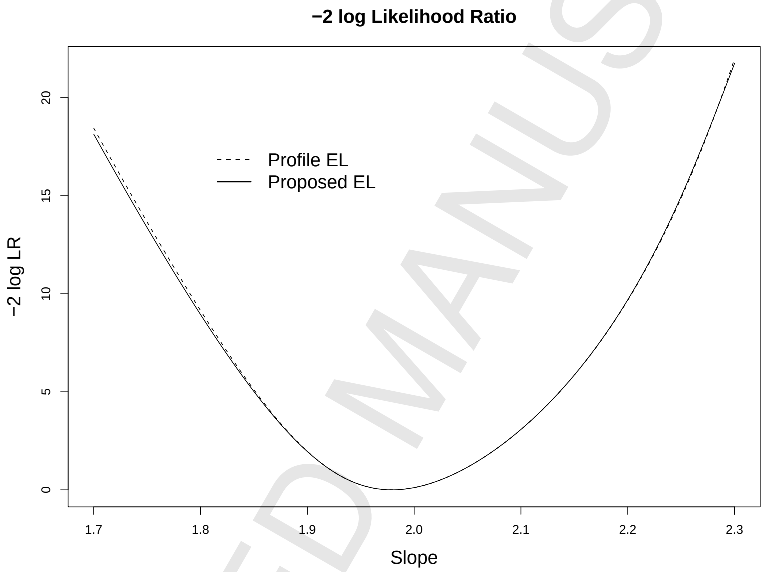



Figure 3: Plots of empirical likelihood ratio functions: solid line for $-2\log\mathcal{R}^*(\beta_{20})$ and dashed line for $-2\log\mathcal{R}(\beta_2)$