Empirical likelihood ratio test for linear models with equality constraints

Yufang Li Jie Tang Yongsong Qin [1]

Department of Statistics, Guangxi Normal University

Guilin, Guangxi 541006, China

Abstract The empirical likelihood (EL) ratio test with equality restrictions for linear models are investigated. The likelihood ratio statistic and the EL ratio statistic are derived and their asymptotical distributions are given under some regularity conditions. Simulations are conducted to compare the performances of the two test methods. Simulation results show that the EL test method is generally superior to the likelihood test method. Especially, for the case of non-identical model errors, the advantage of the EL test method is significant.

Keywords: linear model, equality restriction, empirical likelihood ratio test, power

AMS 2010 subject classification: Primary 62G05 secondary 62E20

Short title: Empirical likelihood test for linear models

---

[1] Corresponding author.

Email address: ysqin@gxnu.edu.cn (Y. Qin)

1. Introduction

Linear regression models are the most important statistical models for explaining the relationship between response and explanatory variables. In particular, the statistical inference for linear models is of fundamental importance, and the inference method can be used as reference to the statistical inference for other models.

In this paper, we study the following linear model:

$$Y_n = X_n\beta + \epsilon_{(n)}, \tag{1}$$

where n is the sample size, β is the $k \times 1$ vector of regression parameters, $X_n = (x_1, x_2, \cdots, x_n)^\tau$ is the non-random $n \times k$ matrix of observations on the independent variable, $Y_n = (y_1, y_2, \cdots, y_n)^\tau$ is an $n \times 1$ vector of observations on the dependent variable, and $\epsilon_{(n)} = (\epsilon_1, \epsilon_2, \cdots, \epsilon_n)^\tau$ is the $n \times 1$ model error vector satisfying $E\epsilon_{(n)} = 0$.

We consider the following hypothesis problem A:

$$H_0 : H\beta = 0, \quad H_1 : H\beta \neq 0, \tag{2}$$

where H is a $s \times k$ known constant matrix.

One may consider a more general hypothesis problem B:

$$\widetilde{H}_0 : H\beta = \xi, \quad \widetilde{H}_1 : H\beta \neq \xi,$$

where H and ξ are $r \times k$ known constant matrix and r dimensional vector, respectively. We assume that the linear system $H\beta = \xi$ is consistent, i.e. there exists a solution, saying β_1. Let $\beta^* = \beta - \beta_1$ and $Y_n^* = Y_n - X_n\beta_1$. Then $H\beta = \xi$ becomes $H\beta^* = 0$ and $Y_n = X_n\beta + \epsilon_{(n)}$ changes into $Y_n^* = X_n\beta^* + \epsilon_{(n)}$. In other words, hypothesis problem B can be turned into hypothesis problem A.

Rao (1972) comprehensively introduced the backgrounds and statistical inference methods for above models and testing problems when the errors are independent, homoscedastic and normally distributed. Further, there are also some developments for above testing problems when the the errors are normally distributed with unknown error covariance matrix. See Magnus (1978), Breusch (1979) and Rothenberg (1984), among others.

For nonparametric settings, the empirical likelihood (EL) method proposed by Owen (1988, 1990) is an important statistical method. Due to its advantages over its counterparts like the normal-approximation-based method and the bootstrap method (e.g., Hall and La Scala, 1990; Hall, 1992), the EL method has also been used to different time series models for constructing tests and confidence regions. Chen and Qin (1993) and Zhong and Rao (2000) and Wu (2004) extended the EL method for finite populations. Owen (1991) extended the EL method to linear models. The results in Owen (1991) can be used to test $\beta = \beta_0$, where β_0 is an known vector. Chen and Cui (2003) further extended the EL method to generalized linear models. Chen and Keilegom (2009) presented a good review on the EL method for regression models. Qin and Lawless(1994) studied the EL method for estimating equations so that the scope of the EL method has been greatly promoted. A prerequisite to use the EL method is that the convex hull of scoring functions must have the zero vector as an interior point. To address this issue, Chen et al. (2008) proposed the adjusted EL (AEL). The AEL method not only guarantees the existence of EL statistics, but also solves the problem of low coverage of EL confidence regions. The under-coverage problem of the EL method was discussed by Tsao (2004). Under i.i.d. sample case, Qin and Lawless (1995) extended the results in Qin and Lawless (1994) to the case that the parameter space is constrained by a set of equations. Under non-random designs, $\{y_1, y_2, \cdots, y_n\}$ in model (1) are not i.i.d., which means that the results in Qin and Lawless (1995) can not be used to the testing problems for model (1).

In this paper, we study the testing problem (2). The likelihood ratio statistic and the EL ratio statistic for (2) are derived and their asymptotical distributions are given under suitable regularity conditions. Simulations are conducted to compare the performances of these test methods. Our results indicate that the EL test method is generally superior to the likelihood test method. Especially, for the case of non-identical model errors, the advantage of the EL test method is significant.

The article is organized as follows. Section 2 presents the main results. Results from a simulation study are reported in Section 3. All technical details are presented in Section 4.

2. Main Results

In Section 2.1 and Section 2.2, the EL ratio test statistic and likelihood ratio test statistic are proposed, respectively. The limiting distributions of the two test statistics and the EL ratio test method al well as the likelihood ratio test method are given in Section 3.3.

2.1 EL ratio test statistic

Suppose that 0 is inside the convex hull of $\{x_i(y_i - x_i^\tau \beta), 1 \leq i \leq n\}$. Following Owen (1991), the (log) EL statistic for the model (1) without restriction on β is

$$\ell_U(\beta) = \sum_{i=1}^{n} \log\{1 + \lambda^\tau(\beta) x_i(y_i - x_i^\tau \beta)\},$$

where $\lambda(\beta) \in R^k$ is the solution of the following equation:

$$\frac{1}{n} \sum_{i=1}^{n} \frac{x_i(y_i - x_i^\tau \beta)}{1 + \lambda^\tau(\beta) x_i(y_i - x_i^\tau \beta)} = 0.$$

Suppose that the rank of H is $R(H) = r$. If $r < k$, then there exists a $k \times (k-r)$ matrix C_0 with $R(C_0) = k - r$ such that $\beta = C_0\theta$, where the vector $\theta \in R^{r-k}$. Therefore, the model (1) with restriction $H\beta = 0$ becomes an unrestricted model:

$$Y_n = \tilde{X}_n \theta + \epsilon_{(n)},$$

where $\tilde{X}_n = X_n C_0$. Suppose that $R(X_n) = k$ so that $R(\tilde{X}_n) = k - r$. Let $\tilde{x}_i^\tau$ be the i-th row of $\tilde{X}_n, 1 \leq i \leq n$. Suppose that 0 is inside the convex hull of $\{\tilde{x}_i(y_i - \tilde{x}_i^\tau \theta), 1 \leq i \leq n\}$. It follows that the (log) EL statistic for the model (1) with restriction $H\beta = 0$ is

$$\ell_R(\theta) = \sum_{i=1}^{n} \log\{1 + \eta^\tau(\theta) \tilde{x}_i(y_i - \tilde{x}_i^\tau \theta)\},$$

where $\eta(\theta) \in R^{k-r}$ is the solution of the following equation:

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\tilde{x}_i(y_i - \tilde{x}_i^\tau \theta)}{1 + \eta^\tau(\theta) \tilde{x}_i(y_i - \tilde{x}_i^\tau \theta)} = 0.$$

If $r = k$, then $H\beta = 0$ leads to $\beta = 0$. In this case, the (log) EL statistic for the model (1) with restriction $H\beta = 0$ is defined as $\ell_R(\theta) = 0$.

Finally, the EL ratio statistic for testing $H\beta = 0$ is defined as

$$\ell_n = 2\{\ell_U(\beta) - \ell_R(\theta)\}. \tag{3}$$

To define the AEL ratio statistic for testing $H\beta = 0$, Let $\omega_i(\beta) = x_i(y_i - x_i^\tau \beta), 1 \le i \le n$ and $\bar{\omega}(\beta) = n^{-1}\sum_{i=1}^n \omega_i(\beta)$, and $\omega_{n+1}(\beta) = -a_n\bar{\omega}_n(\beta)$ for some positive constant a_n. Then the AEL statistic for the model (1) without restriction on β is

$$\ell_{AU}(\beta) = \sum_{i=1}^{n+1} \log\{1 + \lambda_A^\tau(\beta)\omega_i(\beta)\},$$

where $\lambda_A(\beta) \in R^k$ is the solution of the following equation:

$$\frac{1}{n+1}\sum_{i=1}^{n+1} \frac{\omega_i(\beta)}{1 + \lambda_A^\tau(\beta)\omega_i(\beta)} = 0.$$

Similarly, one can define the AEL statistic $\ell_{AR}(\theta)$ for the model (1) with restriction $H\beta = 0$. The AEL ratio statistic for testing $H\beta = 0$ is defined as

$$\ell_{An} = 2\{\ell_{AU}(\beta) - \ell_{AR}(\theta)\}. \tag{4}$$

2.2 Likelihood ratio test statistic

To derive the likelihood ratio test, at this moment, we suppose that $\epsilon_{(n)}$ is normally distributed and $Var(\epsilon_{(n)}) = \sigma^2 I_n$ with $0 < \sigma^2 < \infty$, which is used to derive the likelihood statistic only. Then the (log) likelihood function for the model (1) without restriction on β is

$$\tilde{L}_U = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(Y_n - X_n\beta)^\tau(Y_n - X_n\beta).$$

Using the results in Section 2.1, if $r < k$, one can obtain the (log) likelihood function for the model (1) with restriction $H\beta = 0$ as

$$\tilde{L}_R = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(Y_n - \tilde{X}_n\theta)^\tau(Y_n - \tilde{X}_n\theta).$$

Therefore, if σ^2 is known, one can define the (log) likelihood ratio statistic for testing $H\beta = 0$ as

$$\tilde{L}_n = 2(\tilde{L}_U - \tilde{L}_R) = \frac{1}{\sigma^2}\{(Y_n - \tilde{X}_n\hat{\theta})^\tau(Y_n - \tilde{X}_n\hat{\theta}) - (Y_n - X_n\hat{\beta})^\tau(Y_n - X_n\hat{\beta})\},$$

where $\hat{\beta}$ and $\hat{\theta}$ are the least square estimators of β and θ under the unrestricted model $Y_n = X_n\beta + \epsilon_{(n)}$ and the restricted model $Y_n = X_n\beta + \epsilon_{(n)}$ with $H\beta = 0$, respectively. Since σ^2 is unknown, we use unrestricted model to estimate σ^2, i.e.

$$\hat{\sigma}^2 = \frac{1}{n-k}(Y_n - X_n\hat{\beta})^\tau (Y_n - X_n\hat{\beta}). \tag{5}$$

At last, if $r < k$, we define the likelihood ratio statistic for testing $H\beta = 0$ as

$$L_n = \frac{1}{\hat{\sigma}^2}\{(Y_n - \tilde{X}_n\hat{\theta})^\tau (Y_n - \tilde{X}_n\hat{\theta}) - (Y_n - X_n\hat{\beta})^\tau (Y_n - X_n\hat{\beta})\}, \tag{6}$$

If $r = k$, define $L_n = \frac{1}{\hat{\sigma}^2}\{Y_n^\tau Y_n - (Y_n - X_n\hat{\beta})^\tau (Y_n - X_n\hat{\beta})\}$.

2.3 Limiting distributions of test statistics

Use $||a||$ to denote the L_2-norm of a vector a and let $\lambda_{min}(A)$ denote the least eigenvalue of a matrix A. To obtain the asymptotical distribution of ℓ_n in (3) and L_n in (6), we need following assumptions.

(A1). $\{\epsilon_i, 1 \le i \le n\}$ are independent random variables with mean 0, $E\epsilon_i^2 = \sigma_i^2$, and $E\epsilon_i^4 < \infty, 1 \le i \le n$.

(A2). $R(X_n) = k$, $n^{-2}\sum_{i=1}^n ||x_i||^4 E\epsilon_i^4 \to 0$, and there are constants C_1, C_2 such that $\lambda_{min}\left(\sum_{i=1}^n x_i x_i^\tau \sigma_i^2\right)/n \ge C_1 > 0$ and $n^{-1}\sum_{i=1}^n ||x_i||^3 E(|\epsilon_i|^3) \le C_2 < \infty$.

(A3). $R(H) = r, r \ge 1$.

Theorem 1 states the asymptotic distributions of the EL and the AEL ratio tests ℓ_n and ℓ_{An} in (3) and (4) and Theorem 2 presents the limiting distribution of the likelihood ratio test L_n in (6).

THEOREM 1. *Suppose that Assumptions (A1)-(A3) are satisfied and $a_n = o(n^{2/3})$. Then if $H_0 : H\beta = 0$, as $n \to \infty$,*

$$\ell_n \xrightarrow{d} \chi^2(r) \tag{7}$$

and

$$\ell_{An} \xrightarrow{d} \chi^2(r) \tag{8}$$

where $\chi^2(r)$ is a χ^2-distribution with r degrees of freedom.

THEOREM 2. *Suppose that Assumptions (A1)-(A3) are satisfied, and* $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ *are identically distributed. Then if* $H_0 : H\beta = 0$*, as* $n \to \infty$*,*

$$L_n \xrightarrow{d} \chi^2(r). \tag{9}$$

Let $z_\alpha(r)$ satisfy $P(\chi^2(r) \le z_{1-\alpha}(r)) = 1-\alpha$ for $0 < \alpha < 1$. It follows from Theorems 1 and 2 that the EL (AEL) based and the likelihood based reject regions with significant level α are respectively as $\{\ell_n > z_{1-\alpha}(r)\}(\{\ell_{An} > z_{1-\alpha}(r)\})$ and $\{L_n > z_{1-\alpha}(r)\}$.

3. Simulations

In the simulations, we use the model: $Y_n = X_n\beta + \epsilon_{(n)}$ with $X_n = (x_1, x_2, \cdots, x_n)^\tau, x_i = (1, \frac{i}{n+1}, 3 + \frac{i^2}{n^2})^\tau, 1 \le i \le n$, $\beta = (1.5, 0.5, 1 + \Delta)^\tau$, Δ is taken as $0, 0.1, 10$, respectively, $\epsilon_{(n)} = (\epsilon_1, \epsilon_2, \cdots, \epsilon_n)^\tau$, and $\epsilon_i's$ are taken from $N(0,1)$, $U(-0.5, 0.5)$, $t(5)$ and $\chi_4^2 - 4$, respectively. The sample size n is taken as $50, 100, 200, 300, 500$, respectively and the hypothesis to test is $H_0 : \beta_3 = 2\beta_2$.

At first, under above i.i.d. model errors $\epsilon_{(n)}$, we compare the powers of the EL, AEL and LR based tests by reporting the frequency of rejection at the 5% significant level as Δ varies, i.e. we present the proportions of $\ell_n > z_{0.95}(1)$, $\ell_{An} > z_{0.95}(1)$ with $a_n = 1.3\log(n)$ and $L_n > z_{0.95}(1)$ respectively in $2,000$ replications. These simulation results are reported in Table 1.

Secondly, under non-identically distributed model errors $\epsilon_{(n)}$ indicated in Table 2, we again compare the powers of the the EL, AEL with $a_n = 1.3\log(n)$ and LR based tests by reporting the frequency of rejection at the 5% significant level in $2,000$ replications.

Results in Table 1 indicate that for i.i.d. model errors, in terms of the probability of type 1 error, the LR test is slightly better than the EL test, and as the sample size increases, the advantage of LR gradually disappears. On the other hand, the AEL test performs are similar with the EL test. In terms of the powers of these three tests, both of the AEL test and EL test perform similarly and are significantly better than the LR test. As can be seen from Table 2, in the case of non-identically distributed model errors, the EL test and AEL test are comparable in terms of the probability of type 1 error and the power of the test, and are significantly better than the LR test, which also shows that in this case, the LR test is not appropriate.

Table 1: Frequencies of rejection of the EL, AEL and LR tests under i.i.d. errors

ϵ_i	n	$\Delta = 0$			$\Delta = 0.1$			$\Delta = 10$		
		EL	AEL	LR	EL	AEL	LR	EL	AEL	LR
$N(0,1)$	50	0.111	0.041	0.060	0.746	0.613	0.059	1.000	0.999	0.382
	100	0.064	0.044	0.046	0.886	0.870	0.057	1.000	1.000	0.694
	200	0.061	0.052	0.049	0.982	0.978	0.051	1.000	1.000	0.922
	300	0.046	0.043	0.040	0.997	0.997	0.056	1.000	1.000	0.989
	500	0.054	0.051	0.052	1.000	0.999	0.062	1.000	1.000	1.000
$U(-\frac{1}{2}, \frac{1}{2})$	50	0.096	0.037	0.059	1.000	0.992	0.051	1.000	0.999	1.000
	100	0.061	0.046	0.054	1.000	1.000	0.058	1.000	1.000	1.000
	200	0.055	0.046	0.052	1.000	1.000	0.051	1.000	1.000	1.000
	300	0.054	0.048	0.052	1.000	1.000	0.059	1.000	1.000	1.000
	500	0.056	0.050	0.054	1.000	1.000	0.056	1.000	1.000	1.000
$t(5)$	50	0.117	0.043	0.057	0.616	0.454	0.059	1.000	1.000	0.278
	100	0.094	0.068	0.057	0.787	0.754	0.052	1.000	1.000	0.492
	200	0.071	0.063	0.059	0.927	0.920	0.055	1.000	1.000	0.768
	300	0.071	0.063	0.059	0.966	0.964	0.048	1.000	1.000	0.887
	500	0.055	0.050	0.052	0.993	0.992	0.046	1.000	1.000	0.988
$\chi^2(4) - 4$	50	0.136	0.051	0.056	0.288	0.110	0.048	1.000	0.993	0.113
	100	0.081	0.060	0.054	0.353	0.277	0.043	1.000	1.000	0.146
	200	0.066	0.054	0.048	0.494	0.468	0.056	1.000	1.000	0.242
	300	0.061	0.054	0.051	0.608	0.586	0.046	1.000	1.000	0.326
	500	0.060	0.055	0.052	0.754	0.746	0.048	1.000	1.000	0.496

Table 2: Frequencies of rejection of the EL, AEL and LR tests under non-identically distributed errors

ϵ_i	n	$\Delta = 0$			$\Delta = 0.1$			$\Delta = 10$		
		EL	AEL	LR	EL	AEL	LR	EL	AEL	LR
$N(0, i/n)$	50	0.104	0.039	0.034	1.000	0.979	0.038	1.000	0.998	0.832
	100	0.066	0.045	0.043	1.000	1.000	0.042	1.000	1.000	0.993
	200	0.059	0.048	0.045	1.000	1.000	0.040	1.000	1.000	1.000
	300	0.044	0.037	0.034	1.000	1.000	0.032	1.000	1.000	1.000
	500	0.047	0.044	0.044	1.000	1.000	0.038	1.000	1.000	1.000
$U(-\frac{i}{n}, \frac{i}{n})$	50	0.092	0.032	0.049	1.000	0.99	0.050	1.000	0.997	0.999
	100	0.073	0.050	0.038	1.000	1.000	0.044	1.000	1.000	1.000
	200	0.061	0.052	0.035	1.000	1.000	0.040	1.000	1.000	1.000
	300	0.061	0.056	0.042	1.000	1.000	0.043	1.000	1.000	1.000
	500	0.046	0.043	0.033	1.000	1.000	0.042	1.000	1.000	1.000
$t(i)$	50	0.128	0.030	0.132	0.723	0.563	0.142	1.000	0.998	0.400
	100	0.090	0.059	0.132	0.875	0.853	0.134	1.000	1.000	0.599
	200	0.059	0.051	0.106	0.965	0.964	0.120	1.000	1.000	0.848
	300	0.058	0.054	0.088	0.996	0.996	0.120	1.000	1.000	0.947
	500	0.056	0.051	0.093	1.000	1.000	0.094	1.000	1.000	0.988
$\chi^2(i) - i$	50	0.110	0.043	0.036	0.148	0.062	0.038	1.000	0.995	0.043
	100	0.078	0.052	0.038	0.096	0.066	0.028	1.000	1.000	0.040
	200	0.063	0.050	0.034	0.078	0.062	0.025	1.000	1.000	0.038
	300	0.066	0.058	0.028	0.068	0.059	0.033	1.000	1.000	0.043
	500	0.054	0.050	0.032	0.071	0.065	0.027	1.000	1.000	0.036

4. Proofs

In the sequel, C is used to denote a positive constant that may vary for different occasions. Lemma 1 will be used in the proof of Theorem 1.

LEMMA 1. *Suppose that Assumptions (A1)-(A3) are satisfied. Then as $n \to \infty$,*

$$\max_{1 \le i \le n} ||x_i \epsilon_i|| = o_p(n^{1/2}), \tag{10}$$

$$n^{-1/2} \left\{ n^{-1} \sum_{i=1}^{n} x_i x_i^{\tau} \sigma_i^2 \right\}^{-1/2} \sum_{i=1}^{n} x_i \epsilon_i \xrightarrow{d} N(0, I_k), \tag{11}$$

$$n^{-1} \sum_{i=1}^{n} x_i x_i^{\tau} \epsilon_i^2 = n^{-1} \sum_{i=1}^{n} x_i x_i^{\tau} \sigma_i^2 + o_p(1), \tag{12}$$

$$\sum_{i=1}^{n} ||x_i \epsilon_i||^3 = O_p(n). \tag{13}$$

Proof. To prove (10), take $\delta > 0$. Then

$$P(\max_{1 \le i \le n} ||x_i \epsilon_i|| > n^{1/2}\delta) \le \sum_{i=1}^{n} P(||x_i \epsilon_i|| > n^{1/2}\delta) \le \frac{1}{n^2 \delta^4} \sum_{i=1}^{n} ||x_i||^4 E\epsilon_i^4$$

Then (10) follows by Assumption (A2).

To prove (11), it suffices to show, for any given $l \in R^k$ with $||l|| = 1$, that

$$S_n \hat{=} n^{-1/2} l^{\tau} \left(n^{-1} \sum_{i=1}^{n} x_i x_i^{\tau} \sigma_i^2 \right)^{-1/2} \sum_{i=1}^{n} x_i \epsilon_i \xrightarrow{d} N(0, 1). \tag{14}$$

From Assumptions (A1) and (A2), one can show that

$$Var(S_n) = l^{\tau} l = 1$$

and

$$\sum_{i=1}^{n} E \left| n^{-1/2} l^{\tau} \left(n^{-1} \sum_{i=1}^{n} x_i x_i^{\tau} \sigma_i^2 \right)^{-1/2} x_i \epsilon_i \right|^3 \le C n^{-3/2} \sum_{i=1}^{n} ||x_i||^3 E(|\epsilon_i|^3) = o(1).$$

We thus have (14) by the Liapounov's central limit theorem.

By Chebychev's inequality and Assumptions (A1) and (A2), one can show that (12) holds true.

Note that $\sum_{i=1}^{n} ||x_i\epsilon_i||^3 = \sum_{i=1}^{n} ||x_i||^3|\epsilon_i|^3$. Then By Assumptions (A1) and (A2), we have

$$E\left\{n^{-1}\sum_{i=1}^{n}||x_i\epsilon_i||^3\right\} = E\left\{n^{-1}\sum_{i=1}^{n}||x_i||^3E(|\epsilon_i|^3)\right\} \leq C,$$

which leads to (13).

Proof of Theorem 1. Let

$$U_n = n^{-1}\sum_{i=1}^{n}x_i\epsilon_i, V_n = n^{-1}\sum_{i=1}^{n}x_ix_i^{\tau}\sigma_i^2, \tilde{U}_n = n^{-1}\sum_{i=1}^{n}\tilde{x}_i\epsilon_i, \tilde{V}_n = n^{-1}\sum_{i=1}^{n}\tilde{x}_i\tilde{x}_i^{\tau}\sigma_i^2.$$

Based on Lemma 1 and following the proof of Theorem 1 in Qin (2021), one can show that

$$2\ell_U(\beta) = nU_n^{\tau}V_n^{-1}U_n + o_p(1). \tag{15}$$

Similar to the proof of (15), we have

$$2\ell_R(\theta) = n\tilde{U}_n^{\tau}\tilde{V}_n^{-1}\tilde{U}_n + o_p(1). \tag{16}$$

Note that $\tilde{x}_i^{\tau} = x_i^{\tau}C_0, 1 \leq i \leq n$, which implies

$$\tilde{U}_n = C_0^{\tau}U_n, \tilde{V}_n = C_0^{\tau}V_nC_0. \tag{17}$$

From (15)-(17), we have

$$\begin{aligned}\ell_n &= nU_n^{\tau}\left\{V_n^{-1} - C_0(C_0^{\tau}V_nC_0)^{-1}C_0^{\tau}\right\}U_n + o_p(1)\\ &= \left\{\sqrt{n}V_n^{-1/2}U_n\right\}^{\tau}\left\{I - V_n^{1/2}C_0(C_0^{\tau}V_nC_0)^{-1}C_0^{\tau}V_n^{1/2}\right\}\left\{\sqrt{n}V_n^{-1/2}U_n\right\} + o_p(1).\end{aligned}$$

Note that $I - V_n^{1/2}C_0(C_0^{\tau}V_nC_0)^{-1}C_0^{\tau}V_n^{1/2}$ is symmetric and idempotent with trace $k - (k-r) = r$. Hence $\ell_n \xrightarrow{d} \chi^2(r)$ by Lemma 1. We thus have (7). ((9)) can be proved by following the proof of (7) and the proof of Theorem 1 in Chen, et al. (2008).

Proof of Theorem 2. Denote $\sigma_i^2 = \sigma^2, 1 \leq i \leq n$. From $R(X_n) = k$, we know that $\hat{\beta} = (X_n^{\tau}X_n)^{-1}X_n^{\tau}Y_n$. It can be shown that

$$(Y_n - X_n\hat{\beta})^{\tau}(Y_n - X_n\hat{\beta}) = \epsilon_{(n)}^{\tau}A_n\epsilon_{(n)}, \tag{18}$$

where $A_n = I - X_n(X_n^\tau X_n)^{-1}X_n^\tau$. Noticing that $\tilde{X}_n = X_n C_0$, one can show that

$$(Y_n - \tilde{X}_n\hat{\theta})^\tau(Y_n - \tilde{X}_n\hat{\theta}) = \epsilon_{(n)}^\tau B_n \epsilon_{(n)},$$

where $B_n = I - X_n C_0(C_0^\tau X_n^\tau X_n C_0)^{-1}C_0^\tau X_n^\tau$. Therefore, using the notations in the proof of Theorem 1, we have

$$\begin{aligned} & (Y_n - \tilde{X}_n\hat{\theta})^\tau(Y_n - \tilde{X}_n\hat{\theta}) - (Y_n - X_n\hat{\beta})^\tau(Y_n - X_n\hat{\beta}) = \epsilon_{(n)}^\tau D_n \epsilon_{(n)} \\ = \; & \sigma^2 \times n U_n^\tau \left\{V_n^{-1} - C_0(C_0^\tau V_n C_0)^{-1}C_0^\tau\right\} U_n, \end{aligned}$$

where $D_n = X_n(X_n^\tau X_n)^{-1}X_n^\tau - X_n C_0(C_0^\tau X_n^\tau X_n C_0)^{-1}C_0^\tau X_n^\tau$. Then from the proof of Theorem 1, we have

$$\frac{1}{\sigma^2}\left\{(Y_n - \tilde{X}_n\hat{\theta})^\tau(Y_n - \tilde{X}_n\hat{\theta}) - (Y_n - X_n\hat{\beta})^\tau(Y_n - X_n\hat{\beta})\right\} \xrightarrow{d} \chi^2(r). \tag{19}$$

Let $\Sigma_n = n^{-1}X_{(n)}^\tau X_{(n)}$. Then $n^{-1}\sum_{i=1}^n x_i x_i^\tau = \Sigma_n$ and $n^{-1}\sum_{i=1}^n x_i^\tau \Sigma_n^{-1} x_i = k$. It follows that

$$E\left\{\epsilon_{(n)}^\tau A_n \epsilon_{(n)}\right\} = \sigma^2 tr(A_n) = \sigma^2 \sum_{i=1}^n (1 - n^{-1}x_i^\tau \Sigma_n^{-1} x_i) = \sigma^2(n-k). \tag{20}$$

On the other hand, letting $\mu_4 = E\epsilon_1^4$ and $A_n = (a_{ij})_{n\times n}$, by Assumptions (A1)-(A2) and Lemma B.3 in Su and Yang (2015), we have

$$\begin{aligned} & Var\left\{\epsilon_{(n)}^\tau A_n \epsilon_{(n)}\right\} = (\mu_4 - 3\sigma^4)\sum_{i=1}^n a_{ii}^2 + 2\sigma^4 tr(A_n) \\ = \; & (\mu_4 - 3\sigma^4)\sum_{i=1}^n (1 - n^{-1}x_i^\tau \Sigma_n^{-1} x_i)^2 + 2\sigma^4(n-k) \\ = \; & (\mu_4 - 3\sigma^4)\sum_{i=1}^n \left\{1 - 2n^{-1}x_i^\tau \Sigma_n^{-1} x_i + n^{-2}(x_i^\tau \Sigma_n^{-1} x_i)^2\right\} + 2\sigma^4(n-k) \\ = \; & (\mu_4 - 3\sigma^4)(n-2k) + o(1) + 2\sigma^4(n-k) = n(\mu_4 - \sigma^4) + O(1). \end{aligned} \tag{21}$$

Hence from (5), (18), (20) and (21), we have

$$\hat{\sigma}^2 = \sigma^2 + o_p(1). \tag{22}$$

From (6), (19) and (22), we have (9).

Acknowledgements This work was partially supported by the National Natural Science Foundation of China (12061017), the Science and Technology Project of Guangxi (Guike AD21220114) and the Research Fund of Guangxi Key Lab of Multi-source Information Mining & Security (22-A-01-01).

References

Breusch, T, 1979, Conflict among criteria for testing hypotheses: extensions and comments, Econometrica, 47, 203-207.

Chen, J. and Qin, J., 1993, Empirical likelihood estimation for finite populations and the effective usage of auxiliary information, Biometrika, 80, 107-116.

Chen, J., Variyath, A. M. and Abraham, B., 2008, Adjusted empirical likelihood and its properties, Journal of Computational and Graphical Statistics, 17, 426-443.

Chen, S. X. and Cui, H. J., 2003, An extended empirical likelihood for generalized linear models, Statistica Sinica, 13, 69-81.

Chen, S. X. and Keilegom, I. V. 2009, A review on empirical likelihood for regressions (with discussions), Test, 3, 415-447.

Hall, P., 1992, The Bootstrap and Edgeworth Expansion, Springer-Verlag, New York.

Hall, P. and La Scala, B., 1990, Methodology and algorithms of empirical likelihood, Int. Statist. Rev., 58, 109-127.

Magnus, J., 1978, Maximum likelihood estimation of the GLS model with unknown parameters in the disturbance covariance matrix, Journal of Econometrics, 7, 281-312.

Owen, A. B., 1988, Empirical likelihood ratio confidence intervals for a single functional, Biometrika, 75, 237-249.

Owen, A. B., 1990, Empirical likelihood ratio confidence regions, Ann. Statist., 18, 90-120.

Owen, A. B., 1991, Empirical likelihood for linear models, Ann. Statist., 19, 1725-1747.

Owen, A. B., 2001, Empirical Likelihood, London: Chapman & Hall.

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

Qin, J. and Lawless, J., 1995, Estimating equations, empirical likelihood and constraints on parameters, Can. J. Stat., 23, 145-159.

Qin, Y. S., 2021, Empirical likelihood for spatial autoregressive models with spatial autoregressive disturbances, Sankhy$\bar{a}$ A: Indian J. Statist., 83, 1-25.

Rao, C. R., 1972, Linear Statistical Inference and Its Applications (2nd Edition). New York: John Wiley & Sons.

Rothenberg, T. J., 1984, Hypothesis testing in linear models when the error covariance matrix is nonscalar, Econometrica, 52, 827-842.

Su, L. J. and Yang, Z. L., 2015, QML estimation of dynamic panel data models with spatial errors, Journal of Econometrics, 185, 230-258.

Tsao, M., 2004, Bounds on coverage probabilities of the empirical likelihood ratio confidence regions, Ann. Statist, 32, 1215-1221.

Wu, C. B., 2004, Weighted empirical likelihood inference, Statistics & Probability Letters, 66, 67-79.

Zhong, B. and Rao, J.N.K., 2000, Empirical likelihood inference under stratified random sampling using auxiliary population information, Biometrika, 87, 929-938.