Małgorzata K. Krzciuk
Tomasz Żądło

University of Economics in Katowice

ON SOME TESTS OF FIXED EFFECTS FOR LINEAR MIXED MODELS

Introduction

Let us introduce the assumptions of the general linear mixed model:

$$\begin{cases} \mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v} + \mathbf{e} \\ D^2(\mathbf{v}) = \mathbf{G} \\ D^2(\mathbf{e}) = \mathbf{R} \\ Cov(\mathbf{v}, \mathbf{e}) = \mathbf{0} \end{cases}, \tag{1}$$

where $\mathbf{Y}$ and $\mathbf{e}$ are random vectors of sizes $n \times 1$, $\mathbf{X}$ and $\mathbf{Z}$ are known matrices of sizes $n \times p$ and $n \times q$, respectively, the random vector $\mathbf{v}$ is of size $q \times 1$ and the vector of parameters $\boldsymbol{\beta}$ is of size $p \times 1$. Hence, the variance-covariance matrix of $\mathbf{Y}$ is given by:

$$D^2(\mathbf{Y}) = \mathbf{V} = \mathbf{V}(\boldsymbol{\delta}) = \mathbf{Z}\mathbf{G}\mathbf{Z}^{\mathbf{T}} + \mathbf{R}, \tag{2}$$

where $\boldsymbol{\delta}$ is a vector of unknown variance components.

If in the model (1) we additionally assume that elements of random vectors $\mathbf{E}$ and $\mathbf{v}$ are independent with zero expected values and variances σ_e^2 and σ_v^2, respectively, then $D^2(\mathbf{e}) = \mathbf{R} = \sigma_e^2 \mathbf{I}_{n\times n}$ and $D^2(\mathbf{v}) = \mathbf{G} = \sigma_v^2 \mathbf{I}_{q\times q}$, where $\mathbf{I}$ is the identity matrix. In this case, the variance-covariance matrix (2) simplifies to the following formula: $D^2(\mathbf{Y}) = \mathbf{V} = \mathbf{V}(\boldsymbol{\delta}) = \sigma_v^2 \mathbf{Z}\mathbf{Z}^{\mathbf{T}} + \sigma_e^2 \mathbf{I}_{n\times n}$, where $\boldsymbol{\delta} = \left[\sigma_e^2 \ \sigma_v^2\right]^T$.

In the paper the classic and permutation tests of fixed effects are studied including the situation when the assumption of normality of random effects and random components is not met.

1. Classic tests of significance of fixed effects

In the section we introduce classic tests of significance of fixed effects: Wald test, conditional t-test and likelihood ratio test. In particular, we present their assumptions, tests statistics and information on their properties. In all the tests the null hypothesis is $H_0 : \beta_i = 0$ and the alternative – $H_1 : \beta_i \neq 0$. It should be noted that alternative approach (which is not considered in the paper) called "the transformation method" can be used as well (Rao, 2003, pp. 110-11) – we can transform the mixed model into a standard linear regression model and apply standard linear regression methods for model validation.

The first classic test presented in this paper is Wald test, called also in the literature the Z-test (Verbeke, Molenberghs, 2000, p. 56). Moreover, this test can also be used to test fixed effects in more general class of mixed models (e.g. Wolny-Dominiak, 2011). This test is obtained assuming that the distribution of (Verbecke, Molenbergs, 2000, pp. 56-57):

$$\frac{\left(\hat{\beta}_j - \beta_j\right)}{\hat{D}\left(\hat{\beta}_j\right)}, \tag{3}$$

can be approximated by standard normal distribution. Hence, for a known matrix $\mathbf{K}$ the null and alternative hypotheses in Wald test can be written as: $H_0 : \mathbf{K\beta} = \mathbf{0}$, $H_1 : \mathbf{K\beta} \neq \mathbf{0}$ and Wald test statistic given by:

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \mathbf{K}' \left[\mathbf{K} \left(\mathbf{X}' \mathbf{V}^{-1} \left(\hat{\boldsymbol{\delta}} \right) \mathbf{X} \right)^{-1} \mathbf{K}' \right] \mathbf{K} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \tag{4}$$

may be approximated by χ^2 distribution with $rank(\mathbf{K})$ degrees of freedom. When we test hypothesis $H_0 : \beta_1 = 0$, matrix $\mathbf{K}$ has the form: $\mathbf{K} = [1\ 0\ 0\ \dots\ 0]$. It should be noted that in the Wald test the problem of estimation of variance components $\boldsymbol{\delta}$ is not taken into account (Verbeke, Molenbergs, 2000, pp. 56-57).

To solve the problem we can use conditional t-test which is based on the approximation of (3) by Student's t-distribution (Verbeke, Molenbergs, 2000). If the assumption of normality of random effects and components is fulfilled, the test statistic has a distribution close to the Student t-distribution and only for special cases exact Student t-distribution (Frątczak, ed., 2012, pp. 412-413). Hence for $H_0 : \mathbf{K\beta} = \mathbf{0}$, $H_1 : \mathbf{K\beta} \neq \mathbf{0}$, distribution of the test statistic given by (Wolfinger, 1993, p. 1090; Littel et al., 2006, pp. 755-756, Frątczak, ed., 2012, pp. 412-413):

$$\mathrm{F} = \frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \mathbf{K}' \left[\mathbf{K} \left(\mathbf{X}' \mathbf{V}^{-1} \left(\hat{\boldsymbol{\delta}} \right) \mathbf{X} \right)^{-1} \mathbf{K}' \right] \mathbf{K} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{rank(\mathbf{K})} \tag{5}$$

can be approximated by F distribution with *rank*(**K**) numerator degrees of freedom, and the denumerator degrees of freedom is estimated from the data using e.g. Satterthwaite's (1946) approximation.

The last classic test for fixed effects presented in this paper is the likelihood ratio test (LRT). The basic assumption in this test is that two versions of the analyzed model – without and with the fixed effect − should be nested. The model without the considered effect should therefore be a special case of the model with this effect. It is fulfilled when the model is estimated using maximum likelihood method (Biecek, 2012, p. 160). The test statistic in this case can be written as:

$$2\log\left(\frac{L_2}{L_1}\right) = 2\left[\log\left(L_2\right) - \log\left(L_1\right)\right], \tag{6}$$

where L_1 is likelihood for the model without the analyzed fixed effect (nested model) and L_2 likelihood for the more general model (with this fixed effect). These values must meet the condition $L_2 > L_1$ ($\log(L_2) > \log(L_1)$), so the value of LRT statistic is positive. Under the null hypothesis, the distribution of this statistic is the χ^2 distribution with $k_2 - k_1$ degrees of freedom, where k_i is the number of parameters estimated in the i-th model (Pinheiro, Bates, 2000, p. 83). Hence, as written by Biecek (2012, p. 160), when we test one of the fixed effects we should use quantiles of χ^2 distribution with one degree of freedom.

However, it should be noted that LRT test is "anticonservative", what means that reported p-values for this test can be smaller than the true p-value. It means that the test does not hold its size. Pinheiro, Bates (2000, pp. 87-89) write: "Even though a likelihood ratio test for the ML fits of models with different fixed effects can be calculated, we do **not** recommend using such tests. Such likelihood ratio tests using the standard Chi-square reference distribution tend to be »anticonservative« – sometimes quite badly so". In the simulation study we will show that the properties of the LRT test for the considered problem are unsatisfactory.

2. Permutation tests of fixed effects based on log-likelihood

Biecek (2012, pp. 22-23, 160) studies permutation test of fixed effects, where the test statistics equals log-likelihood of the model. Similarly to other tests in this class, the procedure can be described in three stages (Moore, McCabe, 2005, p. 54). In the first step, we calculate value of the test statistics measuring the effect, in this case it is a log-likelihood for original data denoted by $\ln L_0$. In the second step we repeat B times following points:

– we generate $\boldsymbol{\pi}^{*,b}$ which is permutation of vector [1 2 … n] where n is the number of observations (where $b = 1, 2, \ldots, B$),
– for the test of the jth fixed effect, the jth column of $\mathbf{X}$ is permuted and we obtain permutation version of $\mathbf{X}$ denoted by $\mathbf{X}_{i,j}^{*,b}$ (where $b = 1, 2, \ldots, B$),
– we calculate log-likelihood for the model with $\mathbf{X}_{i,j}^{*,b}$ which is denoted by $\ln L_0^{*,b}$ (where $b = 1, 2, \ldots, B$).

In the third step we calculate p-value for this test (Biecek, 2012, pp. 22-23):

$$p = \frac{1 + \#\left\{b : \ln L_0^{*,b} > \ln L_0\right\}}{1 + B}. \tag{7}$$

It is the fraction of cases where log-likelihood for the model based on permutations is larger than model based on original data.

In the simulation study we will also consider the permutation version of the conditional t-test and the permutation version of the log-likelihood ratio test. To obtain the permutation version of the conditional t-test we should replace in the three stage procedure presented in this section the log-likelihood by the t-test statistic. Similarly, to obtain the permutation version of the LRT test we should replace in the procedure the log-likelihood by the LRT statistic.

3. Simulation study

In this section we present results of the simulation study which is widely used in many applications to assess properties of new methods (e.g. Domański, Jędrzejczak, 2003; Gamrot, 2013; Krzciuk, Mierzwa, Wywiał, 2013). The simulation is conducted in R (R Development Core Team, 2013). Similarly to Kończak (2010, 2012) we will compare properties of classic and permutation tests. In those analyzes we use data on revenues from municipal taxation in 284 Swedish municipalities, presented in Särndal, Swensson, Wretman (1992). In this paper we analyze six variables: RMT85 – revenues from municipal taxation in 1985 in millions kronor, P75 – population in municipalities in thousands in 1985, REV84 – real estate values in 1984 (in millions of kronor), CL – indicator of cluster, REG – indicator of geographic region.

We consider five models: nested error model (called also the basic unit level model by Rao, 2003, p.78)

$$Y_{id} = \beta_0 + \beta_1 X_{1id} + \beta_2 X_{2id} + v_d + e_{id}, \tag{8}$$

two random regression coefficient models with constant:

$$Y_{id} = \beta_0 + (\beta_1 + v_d) X_{1id} + \beta_2 X_{2id} + e_{id} \tag{9}$$

and

$$Y_{id} = \beta_0 + \beta_1 X_{1id} + (\beta_2 + v_d) X_{2id} + e_{id}, \tag{10}$$

and two random regression coefficient models without constant:

$$Y_{id} = (\beta_1 + v_d) X_{1id} + \beta_2 X_{2id} + e_{id} \tag{11}$$

and

$$Y_{id} = \beta_1 X_{1id} + (\beta_2 + v_d) X_{2id} + e_{id}. \tag{12}$$

In all of the models the variable of interest is RMT85 and explanatory variables are P75 and REV84. We considered two grouping variables – CL and REG, therefore we analyzed ten models.

Model selection was made based on the Akkaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). The smallest value of AIC and BIC where obtained for the model $Y_{id} = \beta_0 + (\beta_1 + v_d) X_{1id} + \beta_2 X_{2id} + e_{id}$, where the grouping variable was the indicator of cluster (CL).

We make two experiments to examine the significance of the parameter β_2 corresponding to the variable REV84. In the first experiment we generate data based on the model:

$$Y_{id} = \beta_0 + (\beta_1 + v_d) X_{1id} + e_{id}, \tag{13}$$

where values of β_0, β_1, σ_v^2 and σ_e^2 are obtained based on the real data under assumption of (13). Hence, the parameter β_2 is omitted. In this experiment we analyze type I error for classic and permutation tests. In the second experiment we generate data based on the model:

$$Y_{id} = \beta_0 + (\beta_1 + v_d) X_{1id} + \beta_2 X_{2id} + e_{id}, \tag{14}$$

where values of β_0, β_1, β_2, σ_v^2 and σ_e^2 are obtained based on the real data under assumption of (14). Hence, we consider type II error for classic and permutation tests.

Both experiments are conducted in two variants – (a) and (b). In the variant (a) random components and random effects are generated from the normal distribution. In the variant (b) shifted exponential distribution is used to generate values of random components and random effects. In both variants expectations were 0, variances – σ_e^2 and σ_v^2.

In all experiments we made 1000 iterations and additionally for permutation tests 500 permutations in each iteration. In the variant (a) of the first experiment we consider the data generated based on the normal distributions and the model without the analyzed parameter.

Firstly, we should remind that LRT test does not hold its size. Although it is known property of LRT test of fixed effects in mixed models as it was stated at the end of the section 2, we would like to check in the simulation study what "quite badly" behavior means (see the opinion of Pinheiro and Bates, 2000, cited at the end of the section 2). Moreover, we have checked that the permutation version of LRT test does not hold its size, too.

For classic tests, except for LRT, we obtained similar results for all of levels of significance and in some cases values of the type I error are greater than the assumed α. Classic tests tend to be "anti-conservative". It should be noted that in permutation lnL test and permutation version of t-test we reject null hypothesis exactly or less times than it is assumed (than α). Log-likelihood ratio test in the group of classic tests gave the worst results, similarly to its permutation version among permutation tests.

Table 1

Summary of experiment 1(a) – values of type I error

Test	Assumed level of significance (α)		
	α = 0,01	α = 0,05	α = 0,1
Conditional t-test	0,011	0,047	0,150
Wald test	0,011	0,047	0,106
LRT	0,694	0,739	0,757
Permutation lnL test	0,010	0,049	0,097
Permutation t-test	0,010	0,046	0,090
Permutation LRT	0,154	0,209	0,270

Table 2

Summary of experiment 1(b) – values of type I error

Test	Assumed level of significance (α)		
	α = 0,01	α = 0,05	α = 0,1
Conditional t-test	0,014	0,044	0,087
LRT	0,255	0,278	0,287
Wald test	0,014	0,045	0,088
Permutation lnL test	0,005	0,046	0,106
Permutation t-test	0,005	0,041	0,098
Permutation LRT	0,622	0,684	0,711

In the variant (b) we assume the same model as in the variant (a), but random effects and random components are generated based on shifted exponential distribution. Similarly to the variant (a) the worst results are obtained for classic and permutation LRT. For LRT this type of error was higher than 0,25 for all of assumed level of significance, in the permutation version of this test – higher than 0,6. As in the case of the variant (a), conditional t-test and Wald test gave similar results for all levels of significance. These tests should not be used because they are based on the normality assumption. Results obtained for permutation lnL test and permutation t-test in this variant of the first experiment were similar.

In the second experiment – variant (a), we generate data from normal distribution, but based on the model with the analyzed parameter. Although the best results for permutation tests, and all test too, we obtain for permutation LRT test, we omit both LRT tests in the interpretations because of their very high type I error obtained in the experiments 1(a) and 1(b). All of the permutation tests give better results than classis t-test and Wald test. Comparing classic t-test and permutation t-test we should note that permutation version test give better results for all of assumed levels of significance.

In variant (b) of the second experiment we use the same model as in variant (a), but random effects and random components are generated based on shifted exponential distributions. Firstly, classic tests should not be used because they are based on the normality assumption. Secondly, we omit both LRT tests in the interpretations because of their very high type I error obtained in the experiments 1(a) and 1(b). For all of the considered cases we obtain smaller values of type II errors for permutation t-test than for permutation lnL test.

To sum up the results of the second simulation study, we should stress that the values of the type II error are large. But it is not unusual in econometrics. For example, it is known (see Maddala, 2006, p. 615) that the powers of some widely used unit root tests of stationarity of time series are very low – for some tests even less then 0,1 (what implies values of type II error larger than 0,9).

Table 3

Summary of experiment 2(a) − values of type II error

Test	**Assumed level of significance (α)**		
	α = 0,01	**α = 0,05**	**α = 0,1**
Conditional t-test	0,907	0,763	0,637
LRT	0,754	0,718	0,703
Wald test	0,905	0,761	0,634
Permutation lnL test	0,901	0,761	0,626
Permutation t-test	0,849	0,628	0,491
Permutation LRT	0,360	0,279	0,217

Table 4

Summary of experiment 2(b) − values of type II error

Test	Assumed level of significance (α)		
	α = 0,01	α = 0,05	α = 0,1
Conditional t-test	0,912	0,791	0,689
LRT	0,720	0,695	0,688
Wald test	0,910	0,791	0,687
Permutation lnL test	0,938	0,789	0,667
Permutation t-test	0,937	0,738	0,574
Permutation LRT	0,369	0,283	0,229

Conclusions

In the paper we study three classic and three permutation tests of fixed effects. Permutation tests can be used even if the assumptions of normality of distributions of random effects and random components are not met. Their properties are considered in the Monte Carlo simulation study based on the real data on revenues of municipal taxations in Swedish municipalities. In most of the studied cases in the simulation analysis, permutation lnL test and permutation version of the t-test have the best properties.

References

Biecek P. (2012), Analiza danych z programem R. Modele liniowe z efektami stałymi, losowymi i mieszanymi, WN PWN, Warszawa.

Domański Cz., Jędrzejczak A. (2003), O teście zgodności χ^2dla prób nieprostych, [w:] Metoda reprezentacyjna w badaniach ekonomiczno-społecznych, Wydawnictwo Akademii Ekonomicznej, Katowice, pp. 27-36.

Frątczak E., red. (2012), Zaawansowane metody analiz statystycznych, Szkoła Główna Handlowa − Oficyna Wydawnicza, Warszawa.

Gamrot W. (2013), On Exact Computation of Minimum Sample Size for Restricted Estimation of a Binomial Parameter, „Journal of Statistical Planning and Inference", Vol. 143, pp. 852-866.

Kończak G. (2010), The Moving Average Control Chart Based on the Sequence of Permutation Tests, [in:] Proceedings in Computational Statistics 2010, ed. Y. Lechevallier, G. Saporta, Physica-Verlag, Berlin Heilderberg, pp. 1199-1206.

Kończak G. (2012), On Testing Multi-directional Hypotheses in Categorical Data Analysis, [in:] Proceedings of COMPSTAT 2012, ed. A. Colubi, E.J. Kontoghiorghes, K. Pokianos, G. Gonzalez-Rodriguez, pp. 427-436.

Krzciuk, M., Mierzwa M. and Wywiał J. (2013), Symulacyjne badanie szybkości zbieżności rozkładu statystyk do rozkładu normalnego, „Śląski Przegląd Statystyczny" nr 11 (17), pp. 201-208.

Littell R.C. et al. (2012), SAS for Mixed Models, SAS Institute. Cary, NC.

Maddala G.S. (2006), Ekonometria, WN PWN, Warszawa.

Moore D.S., McCabe G.P. (2005), Introduction to the Practice of Statistics, W.H. Freeman, New York.

Pinheiro J.C., Bates D.M. (2000), Mixed-Effects Models in S and S-PLUS, Springer Verlag, New York.

R Development Core Team (2013), A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna.

Rao J.N.K. (2003), Small Area Estimation, John Wiley & Sons, New Jersey.

Särndal C.E., Swensson B., Wretman J. (1992), Model Assisted Survey Sampling, Springer Verlag, New York.

Satterthwaite F.E. (1946), An Approximate Distribution of Estimates of Variance Components, „Biometrics Bulletin", Vol. 2, No. 6, pp. 110-114.

Verbeke G., Molenberghs G. (2000), Linear Mixed Models for Longitudinal Data, Springer Verlag, New York.

Wolfinger R. (1993), Covariance Structure Selection in General Mixed Models, „Communications in Statistics – Simulation and Computation", Vol. 22, pp. 1079-1106.

Wolny-Dominiak A (2011), Analiza porównawcza modeli mieszanych szacowania stóp taryf w ubezpieczeniach majątkowych z wykorzystaniem kroswalidacji, Wydawnictwo Uniwersytetu Ekonomicznego, Wrocław, pp. 229-237.

ON SOME TESTS OF FIXED EFFECTS FOR LINEAR MIXED MODELS

Summary

In the paper, we consider the problem of testing significance of fixed effects in the class of general linear mixed models. The problem is important especially when the assumption of normality of random effects and random components is not met what is typical for economic applications. In the Monte Carlo simulation studies we compare properties of classic and permutation tests.