This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier’s archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright
Testing under local misspecification and artificial regressions

Anil K. Bera, Gabriel Montes-Rojas, Walter Sosa-Escudero

Abstract

An additivity property of LM tests is derived, linking joint, marginal and Bera–Yoon “adjusted” tests, hence the latter can be derived as the difference of the first two. An artificial regression framework provides an intuitive geometrical illustration of the Bera–Yoon principle.

1. Introduction

The Lagrange Multiplier (LM) or Rao’s score test is a popular choice when the estimation of a parametric “null” model is computationally convenient, and is extensively used in practice. Nevertheless Davidson and Mackinnon (1987) and Saikkonen (1989) showed that its performance may be negatively affected by the presence of misspecified alternatives, that is, when the true model does not correspond to the alternative hypothesis postulated by the researcher. Bera and Yoon (1993) proposed a modification of the LM test that is robust to locally misspecified alternatives, and that has been successfully implemented in many applications. For example, Bera et al. (2007) pointed out that the presence of first order serial correlation makes the standard Breusch and Pagan (1980) test for random effects reject its null hypothesis too often, thus implying that rejections may be due to random effects but also to serial correlation. Based on the Bera–Yoon (BY henceforth) principle, they derived an adjusted LM test for random effects that is not affected by the presence of local serial correlation. In a similar fashion, Baltagi and Li (1999) used the BY principle to obtain tests for functional form misspecification and spatial correlation.

This paper discusses a useful additivity property of LM tests, that links joint, marginal and BY “adjusted” tests. In particular, joint LM tests are shown to be the sum of the BY adjusted LM test for the parameters of interest and the marginal LM test for the nuisance parameters. This is a very useful result, since in many cases joint and marginal tests are easily available. Thus, the BY adjusted tests can be constructed by taking the appropriate differences. Bera et al. (2007) provide a pure algebraic proof of this result based on matrix algebra. We use an artificial regression framework that, besides leading to a very simple proof, provides an intuitive illustration of the relationship between joint, marginal and BY adjusted tests.

2. Joint, marginal and adjusted LM tests

Consider a simple parametric statistical model for an i.i.d. sample of size n, fully characterized by its log-likelihood function $l(\theta) = \sum_{i=1}^{n} l_i(\theta)$, where $\theta \in \mathbb{R}^p$, and $l_i(\theta)$ is the log-likelihood for each observation. Let $d(\theta) = \partial l(\theta)/\partial \theta$ be the score vector, and $J(\theta)$ the information matrix for a single observation. The parameter vector will be partitioned as $\theta = (\psi, \phi')$, where $\psi$ and $\phi$ are vectors in open subsets of $\mathbb{R}^{p_1}$ and $\mathbb{R}^{p_2}$, respectively, so $p = p_1 + p_2$. Let $d_\psi(\theta)$ and $d_\phi(\theta)$ denote the score vectors corresponding to each of these parameters, and partition the information matrix accordingly as

$$J(\theta) = \begin{bmatrix} J_{\psi\psi}(\theta) & J_{\psi\phi}(\theta) \\ J_{\phi\psi}(\theta) & J_{\phi\phi}(\theta) \end{bmatrix}.$$
no parameters are estimated, and later on we will provide a
generalization to the case of estimated parameters. The LM statistic
for \( h_{LM} \), which we will refer to as a “marginal” test, is
\[
LM_\phi = \frac{1}{n} d_\phi(\theta_0) J_{\phi\phi}^{-1}(\theta_0) d_\phi(\theta_0).
\]  
(1)

where \( \theta_0 = (\phi_0, \theta_0)^T \), and a standard result is that under the null hypothesis, \( LM_\phi \) has an asymptotic central chi-squared distribution with \( p_1 \) degrees of freedom.

Davidson and MacKinnon (1987) and Saikkonen (1989) showed
that when \( \phi \neq \phi_0 \), that is, when the alternative is misspecified, \( LM_\phi \) no longer has a central chi-squared distribution. In particular, they show that when \( H_0^* \) holds but \( \phi = \phi_0 + \delta / \sqrt{n} \), with \( 0 \leq \delta < \infty \), \( LM_\phi \) has asymptotically a central chi-squared distribution with \( p_1 \) degrees of freedom. That is, the LM test is insensitive
to the local departure of \( \phi \) away from \( \phi_0 \), hence providing a valid test framework under local misspecification. We will refer to this test as the
“adjusted” LM test.

Finally, consider the “joint” LM test for the null \( H_0 : \phi = \phi_0 \),
\( \phi_0 = \phi_0 \):
\[
LM_{\phi\phi} = \frac{1}{n} d(\theta_0) J^{-1}(\theta_0) d(\theta_0).
\]  
(3)

which under \( H_0 \) has asymptotically a central chi-squared distribution
with \( p_1 + p_2 \) degrees of freedom.

3. The additivity property and artificial regressions

The key result is the following.

**Lemma 1.** \( LM_{\phi\phi} = LM_\phi + LM_\phi = LM_\phi + LM_\phi \).

This is an important result, since it implies that BY adjusted tests
can be simply derived as the difference between joint and marginal tests. Bera et al. (2007) provide a pure algebraic proof of this result.

Nevertheless, our analysis here is based on artificial regressions, which provide a useful asymptotically equivalent framework that helps to intuitively interpret the relationship among the three tests, and leads to a very simple geometric proof of the result.

It is well known that LM tests can be obtained from artificial regressions (see MacKinnon, 1992, for a detailed exposition of artificial regressions and LM tests). Consider \( H_0 : \psi = \psi_0, \phi = \phi_0 \), and the corresponding test statistic given by (1). Consider the following artificial regression
\[
\tau = G(\theta) \psi + u,
\]  
(4)

where \( \tau \) is an \( n \times 1 \) vector of ones, \( G(\theta) \) is an \( n \times p \) matrix with typical element
\[
G(\theta)_{ik} = \frac{\partial l_k(\theta)}{\partial \theta_i}, \quad i = 1, \ldots, n, \quad k = 1, \ldots, p,
\]
and \( u \) is a residual vector. Note that \( d(\theta) = G(\theta)^T \tau \). The outer-product-of-gradientes (OPG) artificial regression relies on the fact that \( n^{-1/2} CG = J + o_p(1) \), so that the explained sum of squares (ESS) of regression model (4)
satisfies the following asymptotic equality
\[
\tau^T G(\theta)^{-1} G(\theta) \tau = \frac{1}{n} d(\theta)^T J^{-1} d(\theta) + o_p(1).
\]  
(5)

We will write the OPG approximation to the joint LM test as
\[
LM_{\phi\phi} = \tau^T G(\theta)^{-1} G(\theta) \tau - 1 G(\theta)^{-1} G(\theta) \tau.
\]

In order to express the marginal tests in terms of artificial regressions let \( G_1(\theta) \) and \( G_2(\theta) \), respectively, denote the columns of \( G(\theta) \) corresponding to \( \psi \) and \( \phi \), thus 
\( G(\theta) = [G_1(\theta) \ G_2(\theta)] \), and 
\( d_\phi(\theta) = G_2(\theta)^T \tau \) and \( d_\psi(\theta) = G_1(\theta)^T \tau \).
Asymptotically, the marginal test \( LM_\phi \) is simply the ESS from regressing \( \tau \) on \( G_1(\theta_0) \), and therefore, its OPG approximation is
\[
LM_{\phi\phi} = \tau^T G_1(\theta_0)^T G_1(\theta_0)^{-1} G_1(\theta_0)^T \tau.
\]

Interestingly, the BY statistic can also be expressed in terms of the
OPG regression, as the ESS of regressing \( \tau \) on the residuals obtained
from the multivariate regression of \( G_1(\theta_0) \) on \( G_2(\theta_0) \). In order to establish this result, let \( G_{\psi} \equiv G_{\psi} - G_{\psi} G_{22} G_{22}^{-1} G_{12} \), where
the dependence on \( \theta_0 \) has been omitted for simplicity. The ESS of regressing \( \tau \) on \( G_{\psi} \) is \( \tau^T G_{\psi}^T G_{\psi}^{-1} G_{\psi} \tau \), and it is easy to check
\[
\tau^T G_{\psi} \tau = \frac{1}{n} \left( G_{\psi} \tau - G_{12} G_{22}^{-1} G_{12} \tau \right)^T
\]  
= \frac{1}{n} \left( d_{\phi} - J_{\phi\phi} d_{\phi} \right) + o_p(1).

Consequently, the adjusted LM test can be seen as a simple LM test where in a first stage the effect of \( \phi \) is removed by orthogonal projection, i.e., considering the residuals from the regression of \( G_{\psi} \) on \( G_{\psi} \).

The asymptotically equivalent of Lemma 1 in the artificial regressions framework is:

**Lemma 2.** \( LM_{\phi\phi} = LM_\phi^a + LM_\phi^a = LM_\phi^a + LM_\phi^a \).

We use the artificial regression framework to provide a simple and intuitive geometrical illustration of this result. Consider Fig. 1. Three vectors are represented with solid arrows. \( \psi \) represents the vector of ones, and \( G_1\) and \( G_2\) represent the vectors of scores. The joint test \( LM_{\phi\phi} \) is represented by the segment \( \overline{\psi} \). Its Euclidean length is the ESS of projecting \( \tau \) on the space spanned by \( G_1\) and \( G_2\), as established by the artificial regression result. The marginal test \( LM_\phi \) is represented by the segment \( \overline{\phi} \), whose length is the ESS of projecting \( \tau \) on the space spanned

\[\text{Fig. 1. Geometrical representation of joint, marginal and BY adjusted tests.}\]
by $G_0$. Finally, the adjusted LM test $LM^*_j$ is represented by the segment $\overline{ab}$. The line containing $\overline{ab}$ is the orthogonal complement of the space spanned by $G_0$, so the residuals of projecting $G_j$ on $G_0$ must lie along this ray. Hence $\overline{ab}$ is the ESS of projecting $j$ on this space.

Based on this picture it easy to see that $TE$ (representing $LM^*_0$) can be constructed as the sum of the orthogonal vectors $\overline{ab} = TE$ and $\overline{d}\overline{a}$ that represent, respectively, $LM^*_0$ and $LM^*_j$, and which is the desired result.

When $G_0$ and $G_j$ are already orthogonal, there is no need for adjustments, and the joint test is simply the sum of the marginal tests $LM^*_0$ and $LM^*_j$, in which case the adjusted and unadjusted test statistics coincide.

A proof of Lemma 2 can be based on two standard results on OLS projections. The first one is that if in the linear model $Y = b_1X_1 + b_2X_2 + u$, $X_1$ and $X_2$ are orthogonal, then $ESS = ESS_1 + ESS_2$, where $ESS_1$ is the explained sum of squares of projecting $Y$ on the space spanned by $X_j$. The second one is the regression

$$Y = a_1X_1 + a_2X_2 + u_2,$$

where $X_2 \equiv M_1X_1$, with $M_1 \equiv I - X_1(X_1'X_1)^{-1}X_1'$, that is, $X_2$ are the residuals from running a regression of $X_2$ on $X_1$. Then the residuals of both regression models coincide, and hence their $ESS$s coincide too, since both models have the same explained variable. To establish Lemma 2, note that the joint test is asymptotically the $ESS$ of regressing $i$ on $G_0$ and $G_j$, so by our previous results, this $ESS$ remains unaltered if $i$ is instead regressed on $G_0$ and $G_j$, because by construction $G_0$ and $G_j$ are orthogonal. Finally, using the first OLS result, and the fact that $LM^*_j$ is the $ESS$ of regressing $i$ on $G_0$, and $LM^*_j$ is the $ESS$ of regressing $j$ on $G_0$, we complete the proof.

In practice there will be unknown parameters that need to be estimated, say $\gamma$, such that the complete parameter vector can be decomposed as $\theta = (\gamma', \psi', \phi')$. Let $\hat{\gamma}$ be the MLE of $\gamma$ under the joint null $H_0: \psi = \psi_0$, $\phi = \phi_0$, and let $\hat{\theta} = (\hat{\gamma}', \hat{\psi_0}', \hat{\phi_0}')$. The marginal LM test for $H_0: \psi = \psi_0$ assuming $\phi = \phi_0$ becomes:

$$LM_{0}^{\gamma} = \frac{1}{n}d_0'\left(\hat{\gamma}ight)J\hat{\gamma}\left(\hat{\gamma}\right)d_0\left(\hat{\gamma}\right),$$

where $J_{\hat{\gamma}} \equiv J_0 - J_{0\gamma}\hat{\gamma}J_{0\gamma}$. The joint LM test for $H_0$: $\psi = \psi_0$, $\phi = \phi_0$ is

$$LM_{0}^{\gamma, \phi} = \frac{1}{n}d_0'\left(\hat{\gamma}\right)J_{0\gamma\phi}\left(\hat{\gamma}\right)d_0\left(\hat{\gamma}\right),$$

where $J_{0\gamma\phi} \equiv J_0 - J_{0\gamma}J_{0\gamma}\phi J_{0\gamma\phi}$, and, finally, the adjusted LM test is

$$LM'_{0}^{\gamma, \phi} = \frac{1}{n}d_0'\left(\hat{\gamma}\right)J_{0\gamma^{-1}}\phi J_{0\gamma}\left(\hat{\gamma}\right)d_0\left(\hat{\gamma}\right),$$

where $d_0 = d_0 - J_{0\gamma}J_{0\gamma}^{-1}d_0$ and $J_{0\gamma^{-1}}\phi J_{0\gamma} = J_0^{-1} - J_{0\gamma^{-1}}\phi J_{0\gamma}^{-1}J_{0\gamma}$.

A more general result is then:

**Lemma 3.** $LM_{0}^{\gamma, \phi} = LM'_{0}^{\gamma, \phi} + LM_{0}^{\gamma, \gamma} + LM_{0}^{\gamma, \phi} + LM_{0}^{\phi, \gamma} + LM_{0}^{\phi, \phi}.$

Our approach can be extended to prove this result using an OPG-based asymptotically equivalent version of Lemma 3. First note that by standard Taylor expansions and by properties of MLE,

$$\frac{1}{\sqrt{n}}d_0'\left(\hat{\gamma}\right) = \frac{1}{\sqrt{n}}d_0'(\theta_0) - J_{0\gamma}(\theta_0)\sqrt{n}(\gamma - \gamma_0) + o_p(1)$$

and

$$\frac{1}{\sqrt{n}}d_0'\left(\hat{\phi}\right) = \frac{1}{\sqrt{n}}d_0'(\theta_0) - J_{0\phi}(\theta_0)\sqrt{n}(\gamma - \gamma_0) + o_p(1).$$

and $\sqrt{n}(\gamma - \gamma_0) = J_0^\gamma(\theta_0)\sqrt{n}d_0'(\theta_0) + o_p(1)$. Second, the OPG approximation implies

$$\frac{1}{\sqrt{n}}d_0'\left(\hat{\theta}\right) = \frac{1}{\sqrt{n}}G_{1}(\theta_0)\psi_0 + (\theta_0 - G_{1}(\theta_0)G_{1}(\theta_0)^{-1}G_{1}(\theta_0))\psi_0 + o_p(1),$$

and

$$\frac{1}{\sqrt{n}}d_0'\left(\hat{\phi}\right) = \frac{1}{\sqrt{n}}G_{2}(\theta_0)\phi_0 + (\theta_0 - G_{2}(\theta_0)G_{2}(\theta_0)^{-1}G_{2}(\theta_0))\phi_0 + o_p(1).$$

It is worth to notice that $G_0 = (G_0 G_p)(G_p G_p)^{-1}G_p$ and $G_0 = (G_0 G_p)(G_p G_p)^{-1}G_p$ are the residuals of an artificial regression of $G_0$ and $G_p$, respectively, on $G_p$. That is, the modified score vectors are orthogonal to the influence of deviations in $\gamma$ through $G_p$. Finally, we can apply Lemma 2 to these orthogonal scores, and the proof concludes by noting that the OPG asymptotic approximation consistently estimates the information matrix.

**4. Concluding remarks**

The additivity property discussed in this paper provides an interesting link between joint, marginal, and the adjusted LM tests proposed by Bera and Yoon (1993), where the latter, unlike standard LM tests, are insensitive to local misspecification. This is a very useful result, since in many practical instances joint and marginal tests are available, so Bera–Yoon robust test statistics can be computed quite easily. As an application, consider the recent paper by Baltagi et al. (2006) that derived marginal and joint tests to detect heteroscedasticity in an error components model. Their marginal tests for each component assume homoscedasticity in the remaining component. Using our additivity result, an adjusted test for homoscedasticity in one error component that is insensitive to the local presence of heteroscedasticity in the other component can be obtained by taking the difference of the joint and a marginal test. As a relevant extension, Bera et al. (2008) recently derived adjusted tests under a more general estimation approach, such as the generalized method of moments (GMM). An interesting exercise for further research would be to generalize the additivity property for the MLE based tests to the GMM framework.

**References**


