The Invariance of Score Tests to Measurement Error

By CHI-LUN CHENG

Institute of Statistical Science, Academia Sinica, Taipei, Republic of China

and CHIH-LING TSAI

Graduate School of Management, University of California, Davis, U.S.A.

Summary. The linear measurement error model is an alternative to the classical regression model, in which we assume that the independent variables are subject to error. This assumption can cause statistical inferences and parameter estimators to differ dramatically from those obtained from the classical regression model. However, inferences may remain unchanged even though the independent variables are assumed to be subject to error. This paper investigates the invariance property of score tests for assessing heteroscedasticity, first-order autoregressive disturbance, and the need for a Box-Cox power transformation. Under specific constraints, we show that the score tests for measurement error models are the same as the corresponding well-established tests derived from classical regression models. We also discuss some possible generalizations.

Keywords: Error-in-the-equation; Functional model; Structural model; Score test; Heteroscedasticity; Autocorrelation; Box-Cox transformation

1. Introduction

In past decades, measurement error (errors-in-variables) models have received great attention. Many of the early results are summarized in Fuller (1987), and updated results of linear and nonlinear measurement error models can be found in Cheng and Van Ness (1999) and Carroll *et al.* (1995). However, there are only a few articles investigating statistical inference in measurement error models, while the parameter estimators, predictions, interval estimations, and hypothesis tests associated with

classical regression models have been well established both theoretically and computationally. Due to the complexity resulting from the presence of measurement error, it is important and natural to consider whether we can adopt directly these established results when the true model is subject to measurement errors. It is known that ignoring measurement error can cause misleading results, such as bias and inconsistency in parameter estimation (see Cheng and Van Ness, 1999, p. 11 and p. 22). On the other hand, Fuller (1987, p. 74) and Cheng and Van Ness (1999, p. 70) showed that both least squares estimators and the maximum likelihood estimators obtained from the classical linear regression model (see 1.1 below) and the measurement errors model (see 1.1-1.2 below) lead to the same predictor. This finding, although admittedly uncommon, motivates us to study the invariance property of score tests for assessing heteroscedasticity, first-order autocorrelation and the necessity for Box-Cox transformation in the linear measurement error model.

The classical linear regression model can be described as

$$y_i = \beta_0 + \boldsymbol{\beta}'\boldsymbol{\xi}_i + \varepsilon_i, \quad (i = 1, \cdots, n), \tag{1.1}$$

where $\boldsymbol{\beta}$ and $\boldsymbol{\xi}_i$ are $p \times 1$ vectors, and the ε_i are random errors. If the independent variables $\boldsymbol{\xi}_i$ cannot be observed directly, and the observed independent variable $\mathbf{x}_i$ is the true value $\boldsymbol{\xi}_i$ plus some random measurement error $\boldsymbol{\delta}_i$:

$$\mathbf{x}_i = \boldsymbol{\xi}_i + \boldsymbol{\delta}_i, \quad (i = 1, \cdots, n), \tag{1.2}$$

then the model (1.1)-(1.2) is called a *measurement error (errors-in- variables)* model. The underlying predictors $\boldsymbol{\xi}_i$ can be regarded as fixed unknown parameters (a *functional model*) or as independent, identically distributed random variables (a *structural model*). Hereafter we will assume the structural model type, that $\boldsymbol{\xi}_i, \boldsymbol{\delta}_i$, and ε_i are independent, identically distributed normal random variables, and that ε_i and $\boldsymbol{\delta}_i$ have zero mean and a covariance matrix of

$$\mathbf{K} = \begin{bmatrix} \sigma_\varepsilon^2 & 0 \\ 0 & \boldsymbol{\Sigma}_{\delta\delta} \end{bmatrix}.$$

In addition, we assume that $(\varepsilon_i, \boldsymbol{\delta}_i')$ and $\boldsymbol{\xi}_j$ are independent for all i and j. Note that when normality is assumed, the parameters β_0 and $\boldsymbol{\beta}$ are not identifiable. Hence,

we adopt the common identifiability assumption that $\boldsymbol{\Sigma}_{\delta\delta}$ is known (Fuller 1987, Chapter 2). The model given by (1.1)-(1.2) is usually called a (structural) *error-in-the-equation* model when the last assumption is adopted, because the error ε_i may be the combination of measurement error and the equation error (Fuller, 1987, Section 2.2; Cheng and Van Ness, 1999, Section 1.5 and Gleser, 1992). It is this type of model that we will consider in this paper.

In the section that follows we will use Gleser's (1992) approach to reduce the complexity of the log-likelihood expression for model (1.1)-(1.2). In Section 3 we obtain score tests (Cox and Hinkley, 1974) for examining heteroscedasticity, first order autocorrelation, and the need for a Box-Cox power transformation in the measurement error model. Except for two constraints, which are satisfied when the sample size is large, these tests are the same as their analogues for classical regression models. Section 4 gives our concluding remarks on the invariance of the score test to the measurement error.

2. Model Structure and Parameter Estimators

The usual form of the log-likelihood function for (1.1)-(1.2), based on the observed data $(\mathbf{x_1}', y_1), \cdots, (\mathbf{x_n}', y_n)$, is complicated. To simplify our analysis we thus adopt Gleser's (1992) approach in order to obtain an alternative expression for this function. Let

$$y_i = \beta_0^* + \boldsymbol{\beta}^{*\prime}\mathbf{x}_i + \varepsilon_i^*, \quad (i = 1, \cdots, n), \tag{2.1}$$

where $\beta_0^* = \beta_0 + \boldsymbol{\beta}'(\mathbf{I}_p - \boldsymbol{\Lambda}')\boldsymbol{\mu}$, $\quad \boldsymbol{\beta}^* = \boldsymbol{\Lambda}\boldsymbol{\beta}$, $\quad \varepsilon_i^* = \boldsymbol{\beta}'\{\boldsymbol{\xi}_i - (\mathbf{I}_p - \boldsymbol{\Lambda}')\boldsymbol{\mu} - \boldsymbol{\Lambda}'\mathbf{x}_i\} + \varepsilon_i$, $\boldsymbol{\mu} = E\boldsymbol{\xi}_i = E\mathbf{x}_i$, $\mathbf{I}_p$ is the identity matrix, $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{\xi\xi} = (\boldsymbol{\Sigma}_{\xi\xi} + \boldsymbol{\Sigma}_{\delta\delta})^{-1}\boldsymbol{\Sigma}_{\xi\xi}$, which is called the *reliability matrix* by Gleser (1992), and $\boldsymbol{\Sigma}_{xx}$ and $\boldsymbol{\Sigma}_{\xi\xi}$ are the covariance matrices of $\mathbf{x}_i$ and $\boldsymbol{\xi}_i$, respectively. Further, ε_i^* are independent, identically distributed normal random variables with mean zero and variance $\sigma^2 = \boldsymbol{\beta}'\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\Lambda}\beta + \sigma_\varepsilon^2$ for $i = 1, \cdots, n$, and ε_i^* is independent of $\mathbf{x}_i$. In addition, let h be the density function of ε^*, and $g(\mathbf{x})$ be the density function of $\mathbf{x}$ with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}_{\xi\xi} + \boldsymbol{\Sigma}_{\delta\delta}$. Then,

applying Cheng and Van Ness's (1999, p. 138) result, $f(\mathbf{x}, y) = g(\mathbf{x})h(\varepsilon^*)$, we obtain the log-likelihood function (ignoring constants) for model (1.1)-(1.2). It is

$$L(\mathbf{y}, \mathbf{X}; \beta_0, \boldsymbol{\beta}, \sigma_\varepsilon^2, \boldsymbol{\mu}, \boldsymbol{\Sigma}_{\xi\xi}) = \sum log\ g(\mathbf{x}_i) - \frac{1}{2}\ log \mid \boldsymbol{\Omega}^* \mid$$

$$-\frac{1}{2}(\mathbf{y} - \beta_0^*\mathbf{1} - \mathbf{X}\boldsymbol{\beta}^*)'(\boldsymbol{\Omega}^*)^{-1}(\mathbf{y} - \beta_0^*\mathbf{1} - \mathbf{X}\boldsymbol{\beta}^*), \tag{2.2}$$

where $\mathbf{y} = (y_1, ..., y_n)'$, $\mathbf{1}= (1, ..., 1)'$, $\mathbf{X} = (\mathbf{x_1}', ..., \mathbf{x_n}')'$, and $\boldsymbol{\Omega}^* = \sigma^2\mathbf{I}_n$.

Based on (2.2), Gleser (1992) showed that the maximum likelihood estimators of β_0^* and $\boldsymbol{\beta}^*$ are the ordinary least squares estimators, provided $\mathbf{S_{xx}} - \boldsymbol{\Sigma}_{\delta\delta}$ is positive definite and the constraint given below (Gleser, 1992, equation (3.4)) is satisfied:

$$\hat{\sigma}_\varepsilon^2 = S_{yy} - \mathbf{S}_{y\mathbf{x}}(\mathbf{S_{xx}} - \boldsymbol{\Sigma}_{\delta\delta})^{-1}\mathbf{S}_{\mathbf{x}y} \geq 0, \tag{2.3}$$

where $\mathbf{S_{xx}}, S_{yy}$, and $\mathbf{S}_{\mathbf{x}y}$ are the sample covariances. When the sample size is large, Gleser (1992) showed that this constraint is satisfied and that $\mathbf{S_{xx}} - \boldsymbol{\Sigma}_{\delta\delta}$ is positive definite. For small to moderate sample sizes, if $\mathbf{S_{xx}} - \boldsymbol{\Sigma}_{\delta\delta}$ fails to be positive definite or (2.3) is violated, then we recommend either collecting more data, or proceeding only after conducting model checking and performing appropriate remedies (see Fuller, 1987). In this paper, we assume these constraints are satisfied.

3. Score tests

Rao (1947) and Cox and Hinkley (1974) established the general form for score statistics as

$$S = \mathbf{V}_1'(\mathbf{I}_{11} - \mathbf{I}_{12}\mathbf{I}_{22}^{-1}\mathbf{I}_{21})^{-1}\mathbf{V}_1, \tag{3.1}$$

where $\mathbf{V}_1 = \partial L(\boldsymbol{\theta})/\partial\boldsymbol{\theta}_1$, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$, $\boldsymbol{\theta}_1$ is the set of parameters to be tested, $\boldsymbol{I}_{ij} = -E(\partial^2 L(\boldsymbol{\theta})/\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j')$ for $i, j = 1$ and 2, and L is the log-likelihood function for the given measurement error model with the vector of unknown parameters $\boldsymbol{\theta}$. In addition, the quantities on the right side of (3.1) are evaluated at $(\boldsymbol{\theta}_{10}', \hat{\boldsymbol{\theta}}_2(\boldsymbol{\theta}_{10})')'$, $\boldsymbol{\theta}_{10}$ is the parameter specified under the null hypothesis, and $\hat{\boldsymbol{\theta}}_2(\boldsymbol{\theta}_{10})$ is the maximum likelihood estimator of $\boldsymbol{\theta}_2$ when $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_{10}$.

Under the null hypothesis, the score test statistic asymptotically has a chi-squared distribution with p_1 degrees of freedom, where p_1 is the dimension of $\boldsymbol{\theta}_1$. The advantage of the score test is that the maximum likelihood estimator needs to be computed only under the null hypothesis, not under the alternative hypothesis. We will make use of equation (3.1) to obtain score tests for assessing heteroscedasticity, autocorrelation, and the need for Box-Cox transformation (see Sections 3.1, 3.2, and 3.3, respectively) in measurement error models.

3.1 Score Test for Heteroscedasticity

We first assume that the disturbances, ε_i, are independently distributed random variables with mean zero and variance

$$var(\varepsilon_i) = w_i\sigma_\varepsilon^2 = w(\mathbf{t}_i, \boldsymbol{\lambda})\sigma_\varepsilon^2 \quad (i = 1, ..., n), \tag{3.2}$$

where $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_q)'$, $\mathbf{t}_i = (t_{i1}, ..., t_{iq})'$ is a known vector, and σ_ε^2 is an unknown constant. We further assume that the weight w_i is a twice-differentiable function of $\boldsymbol{\lambda}$, and that there is a unique value $\boldsymbol{\lambda}^*$ of $\boldsymbol{\lambda}$ such that $w(\mathbf{t}_i, \boldsymbol{\lambda}^*) = 1$ for all $i = 1, \cdots, n$.

The log-likelihood function for the structural model (see equation (2.2)), ignoring constants and incorporating the weighted function in (3.2), is

$$L(\mathbf{y}, \mathbf{X}; \beta_0, \boldsymbol{\beta}, \sigma_\varepsilon^2, \boldsymbol{\mu}, \boldsymbol{\Sigma}_{\xi\xi}, \boldsymbol{\lambda}) = \sum log\ g(\mathbf{x}_i) - \frac{1}{2}\sum log\ (\boldsymbol{\beta}'\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\Lambda}\boldsymbol{\beta} + \sigma_\varepsilon^2 w_i)$$

$$-\frac{1}{2}\sum \frac{(y_i - \beta_0^{*\prime} - \boldsymbol{\beta}^{*\prime}\mathbf{x}_i)^2}{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\Lambda}\boldsymbol{\beta} + \sigma_\varepsilon^2 w_i}.$$

Now let $\mathbf{U}^*$ be an n vector with elements $\hat{\varepsilon}_i^{*2}/\hat{\sigma}^2$, where $\hat{\varepsilon}_i^* = y_i - \hat{\beta}_0^* - \hat{\boldsymbol{\beta}}^{*\prime} x_i$, $\hat{\sigma}^2 = \sum \hat{\varepsilon}_i^{*2}/n$, and $(\hat{\beta}_0^*, \hat{\boldsymbol{\beta}}^{*\prime})$ are the maximum likelihood estimators of $(\beta_0^*, \boldsymbol{\beta}^{*\prime})$ calculated under the null hypothesis ($H_0 : \boldsymbol{\lambda} = \boldsymbol{\lambda}^*$) and the log-likelihood function in (2.2). Also let $\mathbf{D}$ be the $n \times q$ matrix $[\partial w(\mathbf{t}_i, \boldsymbol{\lambda})/\partial\lambda_j]$ evaluated under the null hypothesis, and let $\bar{\mathbf{D}} = \mathbf{D} - \mathbf{1}\mathbf{1}'\mathbf{D}/n$. Then by applying (3.1), we obtain the score test statistic for testing the null hypothesis:

$$S_1 = \frac{1}{2}\mathbf{U}^{*\prime}\bar{\mathbf{D}}(\bar{\mathbf{D}}'\bar{\mathbf{D}})^{-1}\bar{\mathbf{D}}'\mathbf{U}^*.$$

It has the same form as that for testing heteroscedasticity in the classical weighted regression model, but the computations are quite different (see the derivation of S_1 in the appendix and the result in Breusch and Pagan, 1979, and Cook and Weisberg, 1983).

3.2 Score Test for Autocorrelation

We next investigate the score test for first-order autoregressive AR(1) measurement-error models. Here we assume that the errors ε_i follow a first-order autoregressive process, *i.e.*,

$$\varepsilon_i = \rho\varepsilon_{i-1} + u_i \quad (i = 2, \cdots, n),$$

where ρ is the unknown autocorrelation coefficient and the u_i are independent and normally distributed with mean zero and common variance σ_u^2.

The log-likelihood given in equation (2.1) still holds, except that the covariance matrix $\tilde{\mathbf{\Omega}}^*$ of ε_i^* becomes

$$\tilde{\omega}_{ii}^* = \sigma^2 = \boldsymbol{\beta}'\mathbf{\Sigma}_{\delta\delta}\mathbf{\Lambda}\boldsymbol{\beta} + \tilde{\sigma}_\varepsilon^2, \qquad \tilde{\omega}_{ij}^* = \rho^{|j-i|}\tilde{\sigma}_\varepsilon^2, \quad \forall\ i \neq j, \tag{3.3}$$

where $\tilde{\sigma}_\varepsilon^2 = \sigma_u^2/(1-\rho^2)$. Thus the reformulated log-likelihood function, ignoring constants, is

$$L(\mathbf{y}, \mathbf{X}; \beta_0, \boldsymbol{\beta}, \sigma_u^2, \boldsymbol{\mu}, \mathbf{\Sigma}_{\xi\xi}, \rho) = \sum log\ g(\mathbf{x}_i) - \frac{1}{2}\ log \mid \tilde{\mathbf{\Omega}}^* \mid$$

$$-\frac{1}{2}(\mathbf{y} - \beta_0^*\mathbf{1} - \mathbf{X}\boldsymbol{\beta}^*)'(\tilde{\mathbf{\Omega}}^*)^{-1}(\mathbf{y} - \beta_0^*\mathbf{1} - \mathbf{X}\boldsymbol{\beta}^*).$$

Again applying (3.1), the score statistic for testing the null hypothesis $H_0 : \rho = 0$ is

$$S_2 = \frac{\left(\sum{}' \hat{\varepsilon}_i^* \hat{\varepsilon}_{i-1}^*\right)^2}{(n-1)(\hat{\sigma}_u^2 + \hat{\boldsymbol{\beta}}'\mathbf{\Sigma}_{\delta\delta}\hat{\mathbf{\Lambda}}\hat{\boldsymbol{\beta}})^2} = \frac{\left(\sum{}' \hat{\varepsilon}_i^* \hat{\varepsilon}_{i-1}^*\right)^2}{(n-1)\hat{\sigma}^4},$$

where $\sum'$ indicates summation over $i = 2, \cdots, n$. The derivation of S_2 is shown in the appendix, and it has the same form as the score test statistic for evaluating first-order autocorrelation in the classical AR(1) regression model (see Tsai, 1986).

3.3 Approximate Score Test for the Box-Cox transformation

Finally, we consider the Box-Cox transformation model (Box and Cox, 1964), and its associated score statistic. Let

$$y_i(\nu) = \beta_0^* + \beta^{*\prime} x_i + \varepsilon_i^* \quad (i = 1, \cdots, n), \tag{3.4}$$

where β_0^*, β^* and ε_i^* are defined as in (2.1), $y_i(\nu) = (y_i^\nu - 1)/\nu$ if $\nu \neq 0$, and $y_i(\nu) = log\ y_i$ if $\nu = 0$. Based on model (3.4), we can adapt Atkinson's (1973) approach to obtain the approximate score test statistic for testing the null hypothesis $H_0 : \nu = \nu_0$ which is

$$S_3 = -(z' A\dot{z}/\dot{z}' A\dot{z}')s_z,$$

where $z(\nu) = y(\nu)/J^{1/n}$, J is the Jacobian of the transformation, $z = z(\nu)|_{\nu=\nu_0}$, $A = I_n - X(X'X)^{-1}X'$, and $s_z = [\{z'Az - (z'A\dot{z})^2/(\dot{z}'A\dot{z})\}/(n-p-1)]^{1/2}$. This test is the same as Atkinson's (1973) approximate score test, since not only is the model structure the same (3.4) but also the parameter estimates for these two tests are identical.

4. Conclusions

For the linear normal structural model, we apply Gleser's (1992) formulation of the log-likelihood function to obtain score tests for testing heteroscedasticity and auto-correlation, respectively. The resulting score tests are the same as those obtained in the classical regression model with no measurement errors. Hence, practitioners can assess assumptions of constant variance and independent errors, as well as the need for a Box-Cox transformation, without caring whether or not the variables are measured with error.

Two natural questions should be raised: (i). Why are score tests S_1, S_2, and S_3 invariant to errors in variables? (ii). Are all score tests invariant to measurement error? An answer to the first question is given below.

Suppose that V_1 and $\mathbf{I}_{11} - \mathbf{I}_{12}\mathbf{I}_{22}^{-1}\mathbf{I}_{21}$ in (3.1) computed from the model with measurement errors are the same as constants m and m^2 multiplied respectively by V_1 and $\mathbf{I}_{11} - \mathbf{I}_{12}\mathbf{I}_{22}^{-1}\mathbf{I}_{21}$ computed from the corresponding model without measurement errors. Then the resulting score tests computed from the models with and without measurement errors are the same. It can be seen that the score tests S_1 and S_2 have m equal to $\hat{\sigma}_\varepsilon^2$ and $\hat{\sigma}_u^2/(\hat{\sigma}_u^2 + \hat{\boldsymbol{\beta}}'\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}})^2$, respectively. Note that the insensitivity to measurement error usually only holds for testing the specific null hypothesis. For example, if we test $\rho = \rho_0 \neq 0$ for model (3.3), then the resulting score test obtained from the measurement error model is different from the corresponding score test obtained from the classical autoregressive regression model. This is because only under $\rho = 0$ do we have $\sigma_\varepsilon^2 = \sigma_u^2$, which leads to the invariance of S_2 to measurement errors. It is also worth noting that the computations for S_1 in measurement error models (see appendix) is much more complicated than the corresponding score test obtained from the classical weighted regression model (see Cook and Weisberg, 1983) even though both tests are identical. Finally, the transformation model (3.4) has the same structure as the classical transformation model, which leads to the score test S_3 being invariant to measurement errors.

The answer to the second question really depends on the model structure and the parameter to be tested. Consider the score test for testing $H_0 : \boldsymbol{\beta}_2 = 0$ in model (1.1)-(1.2), where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$, and $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are $p_1 \times 1$ and $p_2 \times 1$ vectors. Applying (3.1), we obtain

$$S_4 = \hat{\boldsymbol{\varepsilon}}_{\mathbf{1}}^{*\prime}\tilde{X}_2(\tilde{X}_2'\tilde{X}_2)^{-1}\tilde{X}_2'\hat{\boldsymbol{\varepsilon}}_{\mathbf{1}}^{*}/\hat{\sigma}_1^2,$$

where $X = (X_1', X_2')'$, X_1 and X_2 are $n \times p_1$ and $n \times p_2$ matrices, $\hat{\boldsymbol{\varepsilon}}_{\mathbf{1}}^{*} = \mathbf{y} - \hat{\boldsymbol{\beta}}_0^{*} - X_1\hat{\boldsymbol{\beta}}_1^{*}$, $\hat{\sigma}_1^2 = \hat{\boldsymbol{\varepsilon}}_{\mathbf{1}}^{*\prime}\hat{\boldsymbol{\varepsilon}}_{\mathbf{1}}^{*}/n$, $\tilde{X}_2 = (I - H_1^{*})X_2^{*}$, $H_1^{*} = X_1^{*}(X_1^{*\prime}X_1^{*})^{-1}X_1^{*\prime}$, and $(X_1^{*}, X_2^{*}) = X\Lambda$. If $\Sigma_{\delta\delta}$ is 0, then S_4 is the same as the score test obtained from the classical regression model. Otherwise, S_4 is sensitive to measurement errors. This example clearly indicates that score tests may be sensitive to measurement error. In conclusion, since score tests depend on the model structure and the hypothesis setting, they may not be invariant.

Appendix

Derivation of S_1

For the purpose of simplicity, we assume that $\boldsymbol{\mu}$, β_0 and $\boldsymbol{\Sigma}_{\xi\xi}$ (and hence $\boldsymbol{\Lambda}$) are known. As for the tedious computations of the score test under the full parameters $\boldsymbol{\mu}$, β_0, $\boldsymbol{\Sigma}_{\xi\xi}$, $\boldsymbol{\beta}$, σ_ε^2 and $\boldsymbol{\lambda}$, it can be obtained from the first author upon request. Here, the log-likelihood function $L(\boldsymbol{\beta}, \sigma_\varepsilon^2, \boldsymbol{\lambda})$ of the measurement error model (1.1)-(1.2) with weight function w_i is

$$L(\boldsymbol{\beta}, \sigma_\varepsilon^2, \boldsymbol{\lambda}) = const. - \frac{1}{2}\sum log\ (\boldsymbol{\beta}'\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\Lambda}\boldsymbol{\beta} + \sigma_\varepsilon^2 w_i) - \frac{1}{2}\sum \frac{(y_i - \boldsymbol{\beta}^{*\prime}\mathbf{x}_i)^2}{\boldsymbol{\beta}'\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\Lambda}\boldsymbol{\beta} + \sigma_\varepsilon^2 w_i},$$

where $\boldsymbol{\beta}^* = \boldsymbol{\Lambda}\boldsymbol{\beta}$. Let

$$\mathbf{J}(\boldsymbol{\beta}, \sigma_\varepsilon^2, \boldsymbol{\lambda}) = \begin{bmatrix} \mathbf{I_{11}} & \mathbf{I_{21}}' \\ \mathbf{I_{21}} & \mathbf{I_{22}} \end{bmatrix}$$

be the expected Fisher information matrix partitioned so that $\mathbf{I_{11}}$ corresponds to second partial derivatives with respect to $(\boldsymbol{\beta}, \sigma_\varepsilon^2)$ and $\mathbf{I_{22}}$ corresponds to the second partial derivatives with respect to the elements of $\boldsymbol{\lambda}$.

Denote the inverse matrix of $\mathbf{J}$ by

$$\mathbf{J}^{-1}(\boldsymbol{\beta}, \sigma_\varepsilon^2, \boldsymbol{\lambda}) = \begin{bmatrix} \mathbf{I_{11}}^* & \mathbf{I_{21}}^{*\prime} \\ \mathbf{I_{21}}^* & \mathbf{I_{22}}^* \end{bmatrix}.$$

Straightforward but somewhat tedious computations show that

$$\mathbf{I_{11}} = \begin{bmatrix} \mathbf{T} & \mathbf{V}' \\ \mathbf{V} & \mathbf{F} \end{bmatrix},$$

where

$$\mathbf{T} = -E\frac{\partial^2 L}{\partial\boldsymbol{\beta}\boldsymbol{\beta}'} = \frac{(\mathbf{X}\boldsymbol{\Lambda})'(\mathbf{X}\boldsymbol{\Lambda})}{\hat{\sigma}^2} + \frac{2n\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}'\boldsymbol{\Lambda}\boldsymbol{\Sigma}_{\delta\delta}}{\hat{\sigma}^4},$$

$$\mathbf{V}' = -E\frac{\partial^2 L}{\partial\boldsymbol{\beta}\partial\sigma_\varepsilon^2} = \frac{n\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}}{\hat{\sigma}^4} \quad \text{and} \quad \mathbf{F} = -E\frac{\partial^2 L}{\partial^2\sigma_\varepsilon^2} = \frac{n}{2\hat{\sigma}^4},$$

where $\hat{\sigma}^2 = \hat{\boldsymbol{\beta}}'\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}} + \hat{\sigma}_\varepsilon^2 = n^{-1}\sum \hat{\varepsilon}_i^{*2}$ and all the expectations are evaluated at the maximum likelihood estimators of $\boldsymbol{\beta}$, σ_ε^2 and $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$. Analogously, we obtain

$$\mathbf{I_{21}} = \frac{\hat{\sigma}_\varepsilon^2}{\hat{\sigma}^4}(\mathbf{D}'\mathbf{1}\hat{\boldsymbol{\beta}}'\boldsymbol{\Lambda}\boldsymbol{\Sigma}_{\delta\delta}, \frac{1}{2}\mathbf{D}'\mathbf{1}) \quad \text{and} \quad \mathbf{I_{22}} = \frac{\hat{\sigma}_\varepsilon^4}{2\hat{\sigma}^4}\mathbf{D}'\mathbf{D}.$$

Using equations (9) and (10) of Henderson and Searle (1981), we have

$$\mathbf{I_{11}}^{-1} = \begin{bmatrix} (\mathbf{T}-\mathbf{V}'\mathbf{F}^{-1}\mathbf{V})^{-1} & -(\mathbf{T}-\mathbf{V}'\mathbf{F}^{-1}\mathbf{V})^{-1}\mathbf{V}'\mathbf{F}^{-1} \\ -\mathbf{F}^{-1}\mathbf{V}(\mathbf{T}-\mathbf{V}'\mathbf{F}^{-1}\mathbf{V})^{-1} & (\mathbf{F}-\mathbf{V}\mathbf{T}^{-1}\mathbf{V}')^{-1} \end{bmatrix}.$$

Further algebraic simplifications yield

$$\mathbf{T}-\mathbf{V}'\mathbf{F}^{-1}\mathbf{V} = (\mathbf{X}\boldsymbol{\Lambda})'(\mathbf{X}\boldsymbol{\Lambda})/\hat{\sigma}^2,$$

and

$$-(\mathbf{T}-\mathbf{V}'\mathbf{F}^{-1}\mathbf{V})^{-1}\mathbf{V}'F^{-1} = -2\hat{\sigma}^2\{(\mathbf{X}\boldsymbol{\Lambda})'(\mathbf{X}\boldsymbol{\Lambda})\}^{-1}\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\Lambda}\boldsymbol{\beta}.$$

To compute $(\mathbf{F}-\mathbf{V}\mathbf{T}^{-1}\mathbf{V}')^{-1}$, we use equation (17) of Henderson and Searle (1981), namely,

$$(\mathbf{F}-\mathbf{V}\mathbf{T}^{-1}\mathbf{V}')^{-1} = \mathbf{F}^{-1}+\mathbf{F}^{-1}\mathbf{V}(\mathbf{T}-\mathbf{V}'\mathbf{F}^{-1}\mathbf{V})^{-1}\mathbf{V}'\mathbf{F}^{-1} = 2\hat{\sigma}^4+4\hat{\sigma}^2 a,$$

where $a = \hat{\boldsymbol{\beta}}'\boldsymbol{\Lambda}\boldsymbol{\Sigma}_{\delta\delta}\{(\mathbf{X}\boldsymbol{\Lambda})'(\mathbf{X}\boldsymbol{\Lambda})\}^{-1}\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}$. Thus

$$\mathbf{I_{11}}^{-1} = \hat{\sigma}^2\begin{bmatrix} \{(\mathbf{X}\boldsymbol{\Lambda})'(\mathbf{X}\boldsymbol{\Lambda})\}^{-1} & -2\{(\mathbf{X}\boldsymbol{\Lambda})'(\mathbf{X}\boldsymbol{\Lambda})\}^{-1}\boldsymbol{\Sigma}_{\delta\delta}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}} \\ -2\hat{\boldsymbol{\beta}}'\boldsymbol{\Lambda}\boldsymbol{\Sigma}_{\delta\delta}\{(\mathbf{X}\boldsymbol{\Lambda})'(\mathbf{X}\boldsymbol{\Lambda})\}^{-1} & 2\hat{\sigma}^2/n+4a \end{bmatrix},$$

and hence

$$\mathbf{\hat{I}^*_{22}} = (\mathbf{I_{22}}-\mathbf{I_{21}}\mathbf{I_{11}}^{-1}\mathbf{I_{21}}')^{-1} = 2\hat{\sigma}^4(\bar{\mathbf{D}}'\bar{\mathbf{D}})^{-1}/\hat{\sigma}^4_\varepsilon.$$

Let $L_{\boldsymbol{\lambda}}(\boldsymbol{\lambda} \mid \boldsymbol{\beta}, \sigma^2_\varepsilon)$ be the $q \times 1$ score vector with the jth component

$$[\partial L(\boldsymbol{\beta}, \sigma^2_\varepsilon)/\partial\lambda_j]_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^*}.$$

It can be shown that $L_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}^* \mid \hat{\boldsymbol{\beta}}, \hat{\sigma}^2_\varepsilon) = \hat{\sigma}^2_\varepsilon\mathbf{D}'(\mathbf{U}-\mathbf{1})/2\hat{\sigma}^2$, where $\mathbf{U}$ is the n vector with elements $\hat{\varepsilon}^{*2}_i/\hat{\sigma}^2$. Then the score statistic for testing the hypothesis $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$ is

$$S_1 = \{L_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}^* \mid \hat{\boldsymbol{\beta}}, \hat{\sigma}^2_\varepsilon)\}'\mathbf{\hat{I}^*_{22}}\{L_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}^* \mid \hat{\boldsymbol{\beta}}, \hat{\sigma}^2_\varepsilon)\} = \frac{1}{2}(\mathbf{U}-\mathbf{1})'\mathbf{D}(\bar{\mathbf{D}}'\bar{\mathbf{D}})^{-1}\mathbf{D}'(\mathbf{U}-\mathbf{1})$$

$$= \frac{1}{2}\mathbf{U}'\bar{\mathbf{D}}(\bar{\mathbf{D}}'\bar{\mathbf{D}})^{-1}\bar{\mathbf{D}}'\mathbf{U}.$$

Derivation of S_2

Again, we assume that $\boldsymbol{\mu}$, β_0 and $\boldsymbol{\Sigma}_{\xi\xi}$ (and hence $\boldsymbol{\Lambda}$) are known. Then, the log-likelihood function $L(\boldsymbol{\beta}, \sigma_\varepsilon^2, \rho)$ of (1.1)-(1.2) with the AR(1) error structure is

$$L(\boldsymbol{\beta}, \sigma_u^2, \rho) = const. - \frac{1}{2}\, log \mid \boldsymbol{\Omega}^* \mid -\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^*)'(\boldsymbol{\Omega}^*)^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^*),$$

where $\boldsymbol{\Omega}^*$ is defined in Section 3.2 and $\mathbf{X}$ is the $n \times p$ matrix of the independent variables. Using the formulae (cf. Searle, 1981 Ch. 12)

$$\frac{\partial log \mid \boldsymbol{\Omega}^* \mid}{\partial \rho} = tr\{(\boldsymbol{\Omega}^*)^{-1}\frac{\partial \boldsymbol{\Omega}^*}{\partial \rho}\} \quad \text{and} \quad \frac{\partial}{\partial \rho}\{(\boldsymbol{\Omega}^*)^{-1}\} = -(\boldsymbol{\Omega}^*)^{-1}\frac{\partial \boldsymbol{\Omega}^*}{\partial \rho}(\boldsymbol{\Omega}^*)^{-1},$$

it can be shown that

$$L_\rho(\rho \mid \hat{\boldsymbol{\beta}}, \hat{\sigma}_u^2) = \frac{\hat{\sigma}_u^2 \ \sum' \hat{\varepsilon}_i^* \hat{\varepsilon}_{i-1}^*}{\left(\hat{\sigma}_u^2 + \hat{\boldsymbol{\beta}}' \boldsymbol{\Sigma}_{\delta\delta} \boldsymbol{\Lambda} \hat{\boldsymbol{\beta}}\right)^2}.$$

Furthermore, we obtain

$$\mathbf{I_{21}} = (0, 0) \qquad and \qquad \mathbf{I_{22}} = \frac{(n-1)\hat{\sigma}_u^4}{\left(\hat{\sigma}_u^2 + \hat{\boldsymbol{\beta}}' \boldsymbol{\Sigma}_{\delta\delta} \boldsymbol{\Lambda} \hat{\boldsymbol{\beta}}\right)^2}.$$

Since $\mathbf{I_{21}} = (0, 0)$, the matrix $\mathbf{I_{11}}^*$ is not needed to be computed for obtaining

$$\hat{\mathbf{I}}_{\mathbf{22}}^* = \frac{\left(\hat{\sigma}_u^2 + \hat{\boldsymbol{\beta}}' \boldsymbol{\Sigma}_{\delta\delta} \boldsymbol{\Lambda} \hat{\boldsymbol{\beta}}\right)^2}{(n-1)\hat{\sigma}_u^4}.$$

Under the null hypothesis, $\rho = 0$, we have $\sigma_u^2 = \sigma_\varepsilon^2$. Then applying (3.1), we obtain

$$S_2 = \frac{\left(\sum' \hat{\varepsilon}_i^* \hat{\varepsilon}_{i-1}^*\right)^2}{(n-1)\left(\hat{\sigma}_u^2 + \hat{\boldsymbol{\beta}}' \boldsymbol{\Sigma}_{\delta\delta} \boldsymbol{\Lambda} \hat{\boldsymbol{\beta}}\right)^2} = \frac{\left(\sum' \hat{\varepsilon}_i^* \hat{\varepsilon}_{i-1}^*\right)^2}{(n-1)\hat{\sigma}^4}.$$

REFERENCES

Atkinson, A. C. (1973). Testing transformations to normality. *J. R. Statist. Soc. B* **35**, 473-9.

Atkinson, A. C. (1985). *Plots, Transformations and Regression.* Oxford: Oxford University Press.

Box, G. E. P. (1980). Sampling and Bayes inference in scientific modeling and robustness (with discussion). *J. R. Statist. Soc. A* **143**, 383-430.

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations (with discussion). *J. R. Statist. Soc. B* **26**, 211-52.

Breusch, T. S. and Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica* **47**, 1287-94.

Carroll, R. J., Ruppert, D. and Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models*. London: Chapman & Hall.

Cheng. C.-L. and Van Ness, J. W. (1999). *Statistical Regression with Measurement Error*. London: Arnold.

Cook, R. D. and Weisberg, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika* **70**, 1-10.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. London: Chapman & Hall.

Fuller, W. A. (1987). *Measurement Error Models.* New York: Wiley.

Gleser, L. J. (1992). The importance of assessing measurement reliability in multivariate regression. *J. Am. Statist. Assoc.* **87**, 696-707.

Henderson, H. V. and Searle, S. R. (1981). On deriving the inverse of a sum of matrices. *SIAM Review* **23**, 53-60.

Rao, C. R. (1947). Large sample tests of statistical hypotheses concerning several parameters with application to problems of testing. *Proc. Camb. Phil. Soc.* **44**, 50-7.

Searle, S. R. (1982). *Matrix Algebra Useful for Statistics*. New York: Wiley.

Tsai, C. L. (1986). Score test for the first order autoregressive model with heteroscedasticity. *Biometrika* **73**, 455-60.