HYPOTHESES TESTING FOR ERROR-IN-VARIABLES MODELS

PATRICIA GIMENEZ\textsuperscript{1,2}, HELENO BOLFARINE\textsuperscript{1} AND ENRICO A. COLOSIMO\textsuperscript{1,3}

\textsuperscript{1}Departamento de Estatística, IME/USP, Caixa Postal 20570, 01452-990, São Paulo-SP, Brazil
\textsuperscript{2}Departamento de Matemática - FCEYN/UNMDP, Funes 3350, C.P. 7600, Mar del Plata-BS.AS., Argentina
\textsuperscript{3}Departamento de Estatística, ICEa/UFMG, Caixa Postal 702, 30461-970, Belo Horizonte-MG, Brazil

(Received December 4, 1997; revised April 12, 1999)

Abstract. In this paper, hypotheses testing based on a corrected score function are considered. Five different testing statistics are proposed and their asymptotic distributions are investigated. It is shown that the statistics are asymptotically distributed according to the chi-square distribution or can be written as a linear combination of chi-square random variables with one degree of freedom. A small scale numerical Monte Carlo study is presented in order to compare the empirical size and power of the proposed tests. A comparative calibration example is used to illustrate the results obtained.

Key words and phrases: Asymptotic tests, comparative calibration, consistent estimator, measurement error, naive test.

1. Introduction

Most studies in life sciences, biology, engineering, demography and economics involve covariates that can not be recorded exactly. Errors arise, most notably as measurement errors. Examples include a follow-up study of A-bomb survivors where the variable radiation received is measured with error (Okajima et al. (1985), Pierce et al. (1992)), amount of nitrogen in the soil in a study related to the yield of a certain grain (Fuller (1987)), biologic covariates, such as systolic blood pressure, daily intake of saturated fat in the famous Framingham Heart prospective study dealing with cardiovascular disease (Gordon and Kannel (1968)). Frequently, interests are on assessing the statistical relationship between the unobserved covariates and the response.

The present paper is primarily concerned with testing for association between the true covariates and the response variable. A simple approach considers the naive test obtained from substituting the unobserved covariates with the observed ones. Tosteson and Tsiatis (1988) have compared the local power, assuming a general measurement error structure, of the naive score test and the optimal score test obtained by a flexible procedure in generalized linear models. Lagakos (1988) has also computed the efficiency loss for naive tests in univariate linear, Cox and logistic regression models. Stefanski and Carroll (1990) have considered Wald tests. They have compared the naive Wald and a corrected Wald test (Stefanski (1985)) assuming an additive measurement error structure.

Nakamura (1990) introduced an approach which allows the derivation of consistent and asymptotically normal estimators for the parameters of a linear or nonlinear measurement errors-in-variables model. Additional results on corrected score functions are established by Gimenez and Bolfarine (1997). We recall that most of the approaches considered for estimation in such models produces only approximate unbiased estimates,
with no formal theoretical justification, such as the regression calibration (Carroll and Stefanski (1990)) or James-Stein (Whittenmore (1989)) type estimators. These less biased estimators are used to avoid the attenuation problem typically associated with the naive or ordinary regression estimators. Resampling techniques are then required for obtaining the estimated standard errors associated with such estimates, making it difficult to obtain general valid asymptotic results to be used in conjunction with such estimators. Nakamura’s approach allows its use in more general situations without making assumptions concerning the true covariates, having associated general expressions for the asymptotic covariance matrix. The main object of the paper is to derive asymptotic valid tests (Carroll et al. (1995), p. 207) for some measurement error models, which are validated by the asymptotic distributions associated with the procedures. A review of the approach is considered in Section 2. In Section 3 the asymptotic tests are formally obtained by using the asymptotic properties of the estimators. Wald, score and likelihood type statistics are proposed. A small scale numerical study is presented in Section 4 for comparing the asymptotic tests. The applicability of these results is illustrated in a comparative calibration model in Section 5.

2. Corrected score estimator approach

Let \( Z = (z_1, \ldots, z_r)' \) denote the matrix of independent variables (covariates), \( Y = (y_1, \ldots, y_r)' \) the vector of dependent variables and \( \theta = (\theta_1, \ldots, \theta_p)' \) the \( p \)-dimensional vector of unknown parameters, lying in a parameter space \( \Theta \). The notation considered above is used for simplicity. However, more general situations where \( Z \) and \( Y \) are matrices, leading to multivariate models, for example, can be handled similarly. Moreover, let \( I(\theta; Z, Y) \) be the log-likelihood function corresponding to the sample and

\[
U(\theta; Z, Y) = \frac{\partial I(\theta; Z, Y)}{\partial \theta} \quad \text{and} \quad I(\theta; Z, Y) = -\frac{\partial U(\theta; Z, Y)}{\partial \theta}, \quad \theta \in \mathcal{F},
\]

the score and information matrix, respectively, where \( \mathcal{F} \) is an open convex subspace of \( \Theta \). Let \( \theta_0 \) be the maximum likelihood estimator of \( \theta \), that is, the value of \( \theta \) such that \( U(\theta_0; Z, Y) = 0 \) and \( \theta_0 \in \mathcal{F} \), be the true parameter value. Let \( E^+ (\cdot) \) denote the expectation with respect to the vector \( Y \) given \( Z \). Under some regularity conditions the maximum likelihood estimator (MLE) is consistent and asymptotically normal. These important properties of the MLE are based strongly on the fact that under the true parameter value \( E^+ \{U(\theta_0; Z, Y)\} = 0 \).

We are concerned with the situation that \( Z \) can not be recorded directly, but instead we observe a surrogate \( X = (x_1, \ldots, x_r)' \) having measurement error (Carroll et al. (1995)). In this paper an additive error model

\[
x_i = z_i + u_i, \quad i = 1, \ldots, n
\]

is considered, where the random errors \( u_1, \ldots, u_n \), are mutually independent and also are independent of \( Z \) and \( Y \), each having normal distribution with zero mean and covariance matrix \( \Sigma_u \). This covariance matrix may be assumed known or estimated from validation studies (Fuller (1987)). Thus, calling \( U(\theta; X, Y) \) the naive score function, we have that, in general, \( E\{U(\theta_0; X, Y)\} \neq 0 \), implying that \( \theta_0 \) which solves \( U(\theta; X, Y) = 0 \) is not necessarily a consistent estimator of \( \theta \).

Nakamura (1990) considers a correction for score functions. The corrected score method depends on the existence of a corrected score function \( U^*(\theta; X, Y) \) such that

\[
E\{U^*(\theta; X, Y) \mid Y, Z\} = U(\theta; Z, Y),
\]

(2.1)