A General Asymptotic Result Useful in Models with Nuisance Parameters

Edsel A. Peña* Ma. Zenia Agustin†

May 18, 2001

Abstract

In statistical models with nuisance parameters, a general asymptotic result of test statistics obtained by substitution of estimators of the nuisance parameters for their unknown values in quantities whose asymptotic distributions are known is provided. Applications of the main result in obtaining asymptotic distributions of statistics formed from residuals of the classical linear model as well as in goodness-of-fit testing and model validation for Andersen and Gill's (1982) multiplicative intensity model are illustrated. The asymptotic results can be viewed as further extensions of those by Pierce (1982) in models where the plug-in estimators may not be fully efficient.

1. Introduction

Consider a statistical model $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ with the class of probability measures of form $\mathcal{P} = \{P_{(\eta,\beta)} : (\eta, \beta) \in \mathcal{N} \times \mathcal{B} \subseteq \Re^r \times \Re^s\}$. Denote by X the random entity observed in this model, so $X \in \mathcal{X}$. The parameter vector of the model, (η, β), is viewed as a nuisance parameter, and it will be assumed that the true, but unknown, value of (η, β) is (η_0, β_0). Of interest is the distribution of a $\mathcal{A}$-measurable quantity

$$Q : \mathcal{X} \times \mathcal{N} \times \mathcal{B} \to \Re^q$$

after the substitution of an estimator $(\hat{\eta}, \hat{\beta})$, which is a $\mathcal{A}$-measurable mapping from $\mathcal{X}$ into $\mathcal{N} \times \mathcal{B}$, for the unknown (η, β) in $Q(X, \eta, \beta))$. In the sequel and for economy of notation, provided no confusion could arise, we suppress writing X in the expressions of these quantities, e.g., $Q(\eta, \beta)$ is $Q(X, \eta, \beta)$.

The structures of the estimator $(\hat{\eta}, \hat{\beta})$ of (η, β) are as follows: The estimator $\hat{\beta}$ of β is obtained by solving an estimating equation of form

$$S(X, \beta) = 0, \tag{1.1}$$

where the mapping $S : \mathcal{X} \times \mathcal{B} \to \Re^s$ does not depend on η. After obtaining $\hat{\beta}$ from (1.1), the estimator of η, denoted by $\hat{\eta}$, is obtained by solving the estimating equation

$$R(X, \eta, \hat{\beta}) = 0, \tag{1.2}$$

*E. Peña is Professor in the Department of Statistics, University of South Carolina, Columbia, SC 29208. He acknowledges DHHS/NIH Grant 7 R01 GM56182 which partially supports his research. His e-mail address is pena@stat.sc.edu.

†Z. Agustin is Assistant Professor in the Department of Mathematics and Statistics, Southern Illinois University, Edwardsville, IL 62026. Her e-mail address is zagusti@siue.edu.

where $R : \mathcal{X} \times \mathcal{N} \times \mathcal{B} \to \Re^r$. The estimator $\hat{\eta}$ may therefore be written via $\hat{\eta} = \hat{\eta}(\hat{\beta})$. The (test) statistic whose distribution is of main interest is

$$\hat{Q} \equiv Q(X, \hat{\eta}, \hat{\beta}). \tag{1.3}$$

A setting where the results of this paper are relevant and important is in the context of models where the estimator $\hat{\beta}$ of β is obtained from a partial likelihood for β. The use of a partial likelihood may be practically necessary because of the difficulty of obtaining an estimator from the full likelihood of (η, β). In some situations, the estimating equations in (1.1) and (1.2) may be obtained from intuitive considerations and they need not coincide with estimating equations arising from likelihood functions. Upon obtaining an estimator of $\hat{\beta}$ from an estimating equation that depends only on β, the estimator of η may then be obtained through the use of the profile likelihood, which is obtained by plugging in for β the estimator $\hat{\beta}$ in the full likelihood for (η, β). More generally, the result is useful when the estimators of η and β are obtained via an estimating equation approach. In addition, the results are applicable in determining the asymptotic distributions of model validation statistics based on estimated generalized residuals, such as residuals arising from the Cox proportional hazards model and its extensions, as well as in other models. Some of the earlier papers that have dealt with these issues are those of Cox and Snell (1971), Durbin (1973), Pierce and Kopecky (1979), Pierce (1982), and Randles (1982, 1984). More recently in the context of survival analysis and reliability models, the papers by Baltazar-Aban and Peña (1995), Peña (1998), and Aban and Peña (1999) dealt with some consequences of plugging-in estimators for unknown parameters to obtain model residuals, and the use of these residuals in model validation.

A specific application of this procedure is in the context of goodness-of-fit testing in the Cox proportional hazards model where the problem is to test the composite null hypothesis that the baseline hazard function belongs to a parametric class of hazard functions parameterized by η. The parameter β is the regression coefficient in the Cox model. In this testing problem, (η, β) is considered a nuisance parameter. Under this model, the parameter β is usually estimated using the maximum partial likelihood estimator, while η is estimated from the resulting profile likelihood. The functional form of the quantity $Q(X, \eta, \beta)$ is usually chosen such that the (asymptotic) distribution of $Q(X, \eta_0, \beta_0)$ is fully known. This specific setting will be used for our demonstration of the utility of the asymptotic result in section 3, aside from an application dealing with determining asymptotic distributions of test statistics formed from estimated model residuals in the classical linear model.

2. Asymptotic Results

Generally, finite-sample properties of $\hat{Q}$ are not usually easy to obtain, so we instead focus on its asymptotic distribution. As such we will be considering a sequence of models $(\mathcal{X}^n, \mathcal{A}^n, \mathcal{P}^n)$ with $n = 1, 2, 3, \ldots$. For the nth model, the associated estimators and quantities are Q^n, R^n, S^n, $\hat{\beta}^n$, $\hat{\eta}^n$, and $\hat{Q}^n$. We will assume that the true parameter value (η_0, β_0) does not change with n. From hereon, for brevity, we shall drop the superscript ^{n} if no confusion could arise. In order to make progress we state conditions which are needed to get the asymptotic distribution of $\hat{Q}$. For our notation, if $f(x,y)$ is a differentiable function of (x,y), and if the partial derivative with respect to x is $f_x(x,y)$, then $f_x(x_0,y_0)$ will represent this partial derivative evaluated at $(x,y)=(x_0,y_0)$. The following conditions will be assumed.

(A) As $n \to \infty$, $\frac{1}{\sqrt{n}}\begin{bmatrix} Q(\eta_0,\beta_0) \\ R(\eta_0,\beta_0) \\ S(\beta_0) \end{bmatrix} \xrightarrow{\text{d}} N_{q+r+s}\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \Sigma \equiv \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{pmatrix}\right]$, where Σ is a positive definite matrix.

(B) Each of the components of the mappings $(\eta,\beta) \mapsto Q(\eta,\beta)$, $(\eta,\beta) \mapsto R(\eta,\beta)$, and $(\eta,\beta) \mapsto S(\beta)$ have first-order partial derivatives with respect to the components of η and β, and these partial derivatives are continuous at (η_0,β_0). A subscript of η or β for Q, R, or S will then represent the partial derivative of the matrix function with respect to the component of the parameter. For example, $Q_\eta(\eta,\beta)$ is a $q \times r$ matrix consisting of the partial derivatives $\frac{\partial}{\partial \eta_k} Q_j(\eta,\beta)$ for $j=1,2,\ldots,q$ and $k=1,2,\ldots,r$.

(C) There exist a $q \times r$ matrix A_1, a $q \times s$ matrix A_2, an $r \times r$ nonsingular matrix B_1, an $r \times s$ matrix B_2, and an $s \times s$ matrix C such that as $n \to \infty$,

$$\frac{1}{n}\begin{pmatrix} Q_\eta(\eta_0,\beta_0) & Q_\beta(\eta_0,\beta_0) \end{pmatrix} \xrightarrow{\text{pr}} \begin{pmatrix} A_1 & A_2 \end{pmatrix};$$

and

$$\frac{1}{n}\begin{bmatrix} R_\eta(\eta_0,\beta_0) & R_\beta(\eta_0,\beta_0) \\ 0 & S_\beta(\beta_0) \end{bmatrix} \xrightarrow{\text{pr}} \begin{bmatrix} B_1 & B_2 \\ 0 & C \end{bmatrix}.$$

(D) There exists a sequence $\{\hat{\beta}\}$ satisfying $S(\hat{\beta})=0$ with $\sqrt{n}(\hat{\beta}-\beta_0)=O_p(1)$ as $n \to \infty$. In addition, there exists a sequence $\{\hat{\eta} \equiv \hat{\eta}(\hat{\beta})\}$ satisfying $R(\hat{\eta},\hat{\beta})=0$ with $\sqrt{n}(\hat{\eta}-\eta_0)=O_p(1)$. These conditions require that derivatives up to the second-order of the components of R and S with respect to the components of η and β exist, and furthermore, that the second derivatives are bounded in a neighborhood of (η_0,β_0).

We now state and prove the main asymptotic results.

Theorem 2.1: *As $n \to \infty$, $\frac{1}{\sqrt{n}}\hat{Q} = \frac{1}{\sqrt{n}}Q(\hat{\eta},\hat{\beta})$ has representation*

$$\frac{1}{\sqrt{n}}\hat{Q} = \begin{bmatrix} I & -A_1B_1^{-1} & -\{A_2 - A_1B_1^{-1}B_2\}C^{-1} \end{bmatrix} \frac{1}{\sqrt{n}}\begin{bmatrix} Q(\eta_0,\beta_0) \\ R(\eta_0,\beta_0) \\ S(\beta_0) \end{bmatrix} + o_p(1). \tag{2.1}$$

Consequently, $\frac{1}{\sqrt{n}}\hat{Q}$ converges to a $q \times 1$ Gaussian random vector with asymptotic mean vector $\nu = \mathbf{0}$ and asymptotic covariance matrix Ξ given by

$$\begin{aligned}\Xi = {} & \Sigma_{11} + A_1B_1^{-1}\Sigma_{22}B_1^{-1}A_1^{\mathrm{t}} + (A_2 - A_1B_1^{-1}B_2)C^{-1}\Sigma_{33}C^{-1}(A_2 - A_1B_1^{-1}B_2)^{\mathrm{t}} \\ & -2A_1B_1^{-1}\Sigma_{21} - 2(A_2 - A_1B_1^{-1}B_2)C^{-1}\Sigma_{31} + 2A_1B_1^{-1}\Sigma_{23}C^{-1}(A_2 - A_1B_1^{-1}B_2)^{\mathrm{t}}.\end{aligned} \tag{2.2}$$

Proof : To establish the above results, first we note that by first-order Taylor expansion, $S(\beta) = S(\beta_0) + S_\beta(\beta^*)(\beta-\beta_0)$ where $\beta^* \in [\beta,\beta_0]$. Since $S(\hat{\beta})=0$, then $0 = S(\beta_0) + S_\beta(\beta^*)(\hat{\beta}-\beta_0)$ with

$\beta^* \in [\hat{\beta}, \beta_0]$. Because $\sqrt{n}(\hat{\beta} - \beta_0) = O_p(1)$, by the continuity of the first partial derivative $S_\beta(\beta)$, and since $\frac{1}{n} S_\beta(\beta_0) \xrightarrow{\text{pr}} C$, then $0 = \frac{1}{\sqrt{n}} S(\beta_0) + C\sqrt{n}(\hat{\beta} - \beta_0) + o_p(1)$. Therefore, we have the representation

$$\sqrt{n}(\hat{\beta} - \beta_0) = -C^{-1} \frac{1}{\sqrt{n}} S(\beta_0) + o_p(1). \tag{2.3}$$

Again, by first-order Taylor expansion,

$$\begin{aligned} R(\eta, \hat{\beta}) &= R(\eta, \beta_0) + R_\eta(\eta, \beta^*)(\hat{\beta} - \beta_0) \\ &= R(\eta_0, \beta_0) + R_\eta(\eta^*, \beta_0)(\eta - \eta_0) + \left\{ R_\beta(\eta_0, \beta^*) + (\eta - \eta_0)^{\mathrm{t}} R_{\beta\eta}(\eta^*, \beta^*) \right\} (\hat{\beta} - \beta_0), \end{aligned}$$

with $\beta^* \in [\hat{\beta}, \beta_0]$ and $\eta^* \in [\eta, \eta_0]$. [Note the slight abuse of notation in the last term. If one is to be more precise in writing this, then there is a need to consider each of the component of R, and the jth component of the term $(\eta - \eta_0)^{\mathrm{t}} R_{\beta\eta}(\eta^*, \beta^*)(\hat{\beta} - \beta_0)$ should be written as the quadratic form $(\eta - \eta_0)^{\mathrm{t}} R_{j\beta\eta}(\eta^*, \beta^*)(\hat{\beta} - \beta_0)$.] Now, since $\hat{\eta}$ satisfies $R(\hat{\eta}, \hat{\beta}) = 0$, then

$$\begin{aligned} 0 &= \frac{1}{\sqrt{n}} R(\eta_0, \beta_0) + \left[\frac{1}{n} R_\eta(\eta^*, \beta_0) \right] \sqrt{n}(\hat{\eta} - \eta_0) + \\ & \left[\frac{1}{n} R_\beta(\eta_0, \beta^*) \right] \sqrt{n}(\hat{\beta} - \beta_0) + \frac{1}{\sqrt{n}} [\sqrt{n}(\hat{\eta} - \eta_0)]^{\mathrm{t}} \left[\frac{1}{n} R_{\beta\eta}(\eta^*, \beta^*) \right] [\sqrt{n}(\hat{\beta} - \beta_0)], \end{aligned}$$

where $\eta^* \in [\hat{\eta}, \eta_0]$ and $\beta^* \in [\hat{\beta}, \beta_0]$. Because $\sqrt{n}(\hat{\eta} - \eta_0) = O_p(1)$, $\sqrt{n}(\hat{\beta} - \beta_0) = O_p(1)$, the partial derivatives are continuous, $\frac{1}{n} R_\eta(\eta_0, \beta_0) \xrightarrow{\text{pr}} B_1$, $\frac{1}{n} R_\beta(\eta_0, \beta_0) \xrightarrow{\text{pr}} B_2$, the boundedness of $R_{\beta\eta}(\eta, \beta)$ in a neighborhood of (η_0, β_0), and using the representation (2.3), it follows after an obvious sequence of manipulations that

$$0 = \frac{1}{\sqrt{n}} R(\eta_0, \beta_0) + B_1 \sqrt{n}(\hat{\eta} - \eta_0) - B_2 C^{-1} \frac{1}{\sqrt{n}} S(\beta_0) + o_p(1).$$

This implies the following representation for $\hat{\eta}$:

$$\sqrt{n}(\hat{\eta} - \eta_0) = -B_1^{-1} \frac{1}{\sqrt{n}} R(\eta_0, \beta_0) + B_1^{-1} B_2 C^{-1} \frac{1}{\sqrt{n}} S(\beta_0) + o_p(1). \tag{2.4}$$

By first-order Taylor expansion,

$$\frac{1}{\sqrt{n}} \hat{Q} = \frac{1}{\sqrt{n}} Q(\eta_0, \beta_0) + \left[\frac{1}{n} Q_\eta(\eta^*, \beta^*) \right] \sqrt{n}(\hat{\eta} - \eta_0) + \left[\frac{1}{n} Q_\beta(\eta^*, \beta^*) \right] \sqrt{n}(\hat{\beta} - \beta_0),$$

where $\eta^* \in [\hat{\eta}, \eta_0]$ and $\beta^* \in [\hat{\beta}, \beta_0]$. Since $\frac{1}{n} Q_\eta(\eta^*, \beta^*) = A_1 + o_p(1)$ and $\frac{1}{n} Q_\beta(\eta^*, \beta^*) = A_2 + o_p(1)$, then

$$\frac{1}{\sqrt{n}} \hat{Q} = \frac{1}{\sqrt{n}} Q(\eta_0, \beta_0) + A_1 \sqrt{n}(\hat{\eta} - \eta_0) + A_2 \sqrt{n}(\hat{\beta} - \beta_0) + o_p(1).$$

Using the representations in (2.3) and (2.4), we arrive at the representation

$$\frac{1}{\sqrt{n}} \hat{Q} = \left[\begin{array}{ccc} I & -A_1 B_1^{-1} & -(A_2 - A_1 B_1^{-1} B_2) C^{-1} \end{array} \right] \frac{1}{\sqrt{n}} \left[\begin{array}{c} Q(\eta_0, \beta_0) \\ R(\eta_0, \beta_0) \\ S(\beta_0) \end{array} \right] + o_p(1).$$

Finally, it follows from condition (A) that $\frac{1}{\sqrt{n}} \hat{Q}$ converges in distribution to a Gaussian random vector whose mean vector is zero and whose covariance matrix

$$\Xi = \left[\begin{array}{ccc} I & -A_1 B_1^{-1} & -(A_2 - A_1 B_1^{-1} B_2) C^{-1} \end{array} \right] \left[\begin{array}{ccc} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{array} \right] \left[\begin{array}{c} I \\ -B_1^{-1} A_1^{\mathrm{t}} \\ -C^{-1}(A_2 - A_1 B_1^{-1} B_2)^{\mathrm{t}} \end{array} \right],$$

which is the expression in (2.2). This completes the proof. $\|$

An immediate extension of Theorem 2.1 is when the asymptotic mean vector of

$$\frac{1}{\sqrt{n}} \left(\begin{array}{ccc} Q(\eta_0, \beta_0)^{\mathrm{t}} & R(\eta_0, \beta_0)^{\mathrm{t}} & S(\eta_0, \beta_0)^{\mathrm{t}} \end{array} \right)^{\mathrm{t}}$$

is $\mu = \left(\begin{array}{ccc} \mu_1^{\mathrm{t}} & \mu_2^{\mathrm{t}} & \mu_3^{\mathrm{t}} \end{array} \right)^{\mathrm{t}}$. This situation is relevant in problems where the distribution of $\frac{1}{\sqrt{n}}\hat{Q}$ is desired under a sequence of local alternatives converging to (η_0, β_0). It is easy to see that the only change in Theorem 2.1 pertains to the asymptotic mean ν of $\frac{1}{\sqrt{n}}\hat{Q}$, which will now equal

$$\nu = \mu_1 - A_1 B_1^{-1} \mu_2 - (A_2 - A_1 B_1^{-1} B_2) C^{-1} \mu_3. \tag{2.5}$$

We remark that one may view the asymptotic result in Theorem 2.1 as a further extension of results in Pierce (1982). However, note that his result may not apply if the plug-in estimators are not efficient, and indeed, his proof is shorter because the nuisance parameter estimators are efficient.

3. Two Applications

We now demonstrate the applicability of the preceding results in two situations. The first situation deals with the classical linear model, while the second one deals with a popular counting process model in survival analysis and reliability settings.

3.1 On Statistics Based on Linear Model Residuals

Consider the classical linear model $\mathbf{Y} = \mathbf{X}\beta + \epsilon$, where $\mathbf{Y}$ is an observable $n \times 1$ vector, $\mathbf{X}$ is an observable and fixed $n \times p$ design matrix of full rank, β is a $p \times 1$ vector of unknown regression coefficients, and ϵ is an unobserved $n \times 1$ error vector whose distribution is multivariate normal with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{Cov}\{\epsilon, \epsilon\} = \sigma^2 \mathbf{I}_n$, with σ^2 unknown. For our notation, for $i = 1, 2, \ldots, n$, let $\mu_i(\beta) = \mathbf{X}_i \beta$ be the mean of Y_i. Let

$$\hat{\beta} = (\mathbf{X}^{\mathrm{t}}\mathbf{X})^{-1}(\mathbf{X}^{\mathrm{t}}\mathbf{Y}) \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{n}(\mathbf{Y} - \mathbf{X}\hat{\beta})^{\mathrm{t}}(\mathbf{Y} - \mathbf{X}\hat{\beta}) \tag{3.1}$$

be the maximum likelihood estimators of β and σ^2. Let $\psi(\cdot)$ be a differentiable real-valued function on $\Re$ with derivative $\psi'(\cdot)$, and such that if Z is the standard normal variable, then $\mathbf{E}\{\psi(Z)^2\} < \infty$, $\mathbf{E}\{\psi(Z)\} = 0$, and $\mathbf{E}\{[\psi'(Z)]^2\} < \infty$. Of interest is the asymptotic distribution of the statistic

$$\hat{Q} = \sum_{i=1}^{n} \psi\left(\frac{Y_i - \mu_i(\hat{\beta})}{\hat{\sigma}}\right). \tag{3.2}$$

Notice that the statistics $R_i = (Y_i - \mu_i(\hat{\beta}))/\hat{\sigma}, i = 1, 2, \ldots, n$, are the estimated model residuals. We now put this in the framework in which Theorem 2.1 is directly applicable.

The correspondence of the parameters in this model and those in Theorem 2.1 will be $\sigma^2 \leftrightarrow \eta$ and $\beta \leftrightarrow \beta$. The starting quantity leading to $\hat{Q}$ is

$$Q(\sigma^2, \beta) = \sum_{i=1}^{n} \psi\left(\frac{Y_i - \mu_i(\beta)}{\sigma}\right).$$

The relevant estimating equations for obtaining the estimators $\hat{\beta}$ and $\hat{\sigma}^2$ are

$$\mathbf{S}(\beta) \equiv \mathbf{X}^{\mathrm{t}}(\mathbf{Y} - \mathbf{X}\beta) = \sum_{i=1}^{n} \mathbf{X}_i^{\mathrm{t}}(Y_i - \mu_i(\beta)) = \mathbf{0} \tag{3.3}$$

where $\mathbf{X}_i$ is the ith row of $\mathbf{X}$, and

$$R(\sigma^2, \beta) \equiv \sum_{i=1}^{n} \left[\left(\frac{Y_i - \mu_i(\beta)}{\sigma} \right)^2 - 1 \right] = 0. \tag{3.4}$$

If β_0 and σ_0^2 are the true values of β and σ^2, respectively, then by noting that $Z_i \equiv (Y_i - \mu_i(\beta_0))/\sigma_0, i = 1, 2, \ldots, n$ are independent and identically distributed standard normal variables, it follows by the multivariate central limit theorem that

$$\frac{1}{\sqrt{n}} \begin{bmatrix} Q(\sigma_0^2, \beta_0) \\ R(\sigma_0^2, \beta_0) \\ \mathbf{S}(\beta_0) \end{bmatrix} \xrightarrow{\mathrm{d}} N \left[\begin{pmatrix} 0 \\ 0 \\ \mathbf{0} \end{pmatrix}, \mathbf{\Sigma} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \mathbf{\Sigma}_{13} \\ \Sigma_{12}^{\mathrm{t}} & \Sigma_{22} & \mathbf{0} \\ \mathbf{\Sigma}_{13}^{\mathrm{t}} & \mathbf{0} & \mathbf{\Sigma}_{33} \end{pmatrix} \right],$$

where $\Sigma_{11} = \mathbf{Var}\{\psi(Z)\} = \mathbf{E}\{\psi(Z)^2\}$, $\Sigma_{22} = 2$, $\mathbf{\Sigma}_{33} = \sigma^2\mathbf{V}$, $\Sigma_{12} = \mathbf{Cov}\{\psi(Z), Z^2\} = \mathbf{E}\{Z^2\psi(Z)\}$, and $\mathbf{\Sigma}_{13} = \sigma\mathbf{Cov}\{\psi(Z), Z\}\nu = \sigma\mathbf{E}\{Z\psi(Z)\}\nu$, where ν and $\mathbf{V}$ are such that, as $n \to \infty$,

$$\left| \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i - \nu \right| \to 0 \quad \text{and} \quad \left| \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i^{\mathrm{t}}\mathbf{X}_i - \mathbf{V} \right| \to 0.$$

Straightforward calculations yield the following quantities:

$$Q_{\sigma^2}(\sigma^2, \beta) = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(\frac{Y_i - \mu_i(\beta)}{\sigma} \right) \psi' \left(\frac{Y_i - \mu_i(\beta)}{\sigma} \right);$$

$$\mathbf{Q}_\beta(\sigma^2, \beta) = -\frac{1}{\sigma} \sum_{i=1}^{n} \mathbf{X}_i \psi' \left(\frac{Y_i - \mu_i(\beta)}{\sigma} \right);$$

$$R_{\sigma^2}(\sigma^2, \beta) = -\frac{1}{\sigma^2} \sum_{i=1}^{n} \left(\frac{Y_i - \mu_i(\beta)}{\sigma} \right)^2;$$

$$\mathbf{R}_\beta(\sigma^2, \beta) = -2 \sum_{i=1}^{n} \mathbf{X}_i \left(\frac{Y_i - \mu_i(\beta)}{\sigma} \right);$$

$$\mathbf{S}_\beta(\beta) = -\sum_{i=1}^{n} \mathbf{X}_i^{\mathrm{t}}\mathbf{X}_i.$$

Consequently,

$$\frac{1}{n} \left(Q_{\sigma^2}(\sigma_0^2, \beta_0), \quad \mathbf{Q}_\beta(\sigma_0^2, \beta_0) \right) \xrightarrow{\mathrm{pr}} \left(A_1 \equiv -\frac{1}{2\sigma_0^2}\mathbf{E}\{Z\psi'(Z)\}, \quad \mathbf{A}_2 \equiv -\frac{1}{\sigma_0}\mathbf{E}\{\psi'(Z)\}\nu \right);$$

$$\frac{1}{n} \begin{bmatrix} R_{\sigma^2}(\sigma_0^2, \beta_0) & \mathbf{R}_\beta(\sigma_0^2, \beta_0) \\ 0 & \mathbf{S}_\beta(\sigma_0^2, \beta_0) \end{bmatrix} \xrightarrow{\mathrm{pr}} \begin{bmatrix} B_1 \equiv -\frac{1}{\sigma_0^2} & \mathbf{B}_2 \equiv \mathbf{0} \\ \mathbf{0} & \mathbf{C} \equiv -\mathbf{V} \end{bmatrix}.$$

Direct substitution of these expressions for Ξ in (2.2) and straightforward simplifications now yield

$$\begin{aligned} \Xi &= \mathbf{Var}\{\psi(Z)\} + \frac{1}{2}\mathbf{E}\{Z\psi'(Z)\} \left\{ \mathbf{E}\{Z\psi'(Z)\} - 2\mathbf{E}\{Z^2\psi(Z)\} \right\} + \\ &\quad \mathbf{E}\{\psi'(Z)\} \left\{ \mathbf{E}\{\psi'(Z)\} - 2\mathbf{E}\{Z\psi(Z)\} \right\} (\nu\mathbf{V}^{-1}\nu^{\mathrm{t}}), \end{aligned} \tag{3.5}$$

where we recall that Z is a standard normal variable. Note that the asymptotic variance of $Q(\sigma_0^2, \beta_0)$ equals $\mathbf{Var}\{\psi(Z)\}$, so that from the above expression, the effect of substituting the maximum likelihood estimators for the associated unknown parameters in the quantity Q to obtain the statistic $\hat{Q}$ is contained in the last two sets of terms in (3.5).

We now consider some specific choices of $\psi(\cdot)$. First, suppose that

$$\psi(z) \equiv \psi_1(z) = z. \tag{3.6}$$

Then, it is immediate that $\mathbf{E}\{Z\psi'(Z)\} = 0$, $\mathbf{E}\{\psi'(Z)\} = 1$, $\mathbf{E}\{Z\psi(Z)\} = 1$, and $\mathbf{E}\{Z^2\psi(Z)\} = 0$. Consequently, from (3.5),

$$\Xi \equiv \Xi_{\psi_1} = 1 - \nu \mathbf{V}^{-1} \nu^{\mathrm{t}}. \tag{3.7}$$

Therefore, under the linear model,

$$\frac{1}{\sqrt{n}} \hat{Q}_{\psi_1} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{Y_i - \mathbf{X}_i \hat{\beta}}{\hat{\sigma}} \right] \tag{3.8}$$

converges in distribution to a zero-mean normal with variance $1 - \nu\mathbf{V}^{-1}\nu^{\mathrm{t}}$. The effect of the substitution of estimators for the unknown parameters is the reduction in the variance given by $\Delta = \nu\mathbf{V}^{-1}\nu^{\mathrm{t}}$. By the definitions of ν and $\mathbf{V}$, we must have $|\hat{\Delta} - \nu\mathbf{V}^{-1}\nu^{\mathrm{t}}| \to 0$, where

$$\hat{\Delta} = \left[\frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i \right] \left[\frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i^{\mathrm{t}} \mathbf{X}_i \right]^{-1} \left[\frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i \right]^{\mathrm{t}}.$$

Interestingly, note that if the $\mathbf{X}_i$'s are properly centered so that $\nu = \mathbf{0}$, then the plug-in procedure will not have an effect, at least asymptotically.

However, this is usually not the case since oftentimes the model will have an intercept term, which will make the design matrix $\mathbf{X}$ to have as its first column $\mathbf{1}_n = (1, 1, \ldots, 1)^{\mathrm{t}}$. Let us examine the value of $\hat{\Delta}$ under this linear model with intercept term. We have

$$\hat{\Delta} = \frac{1}{n} \mathbf{1}_n^{\mathrm{t}} \mathbf{X} (\mathbf{X}^{\mathrm{t}} \mathbf{X})^{-1} \mathbf{X}^{\mathrm{t}} \mathbf{1}_n = \frac{1}{n} \mathbf{1}_n^{\mathrm{t}} \mathbf{1}_n = 1, \tag{3.9}$$

with the second to last equality obtaining since $\mathbf{1}_n$ is in the column space of $\mathbf{X}$ and $\mathbf{X}(\mathbf{X}^{\mathrm{t}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{t}}$ is a projection matrix. Thus, consequently, when the linear model contains an intercept term, the statistic in (3.8) has asymptotic variance equal to 0. This is not surprising, and indeed is to be expected, since the sum of the estimated model residuals in this linear model should always be zero, hence even for finite n, the variance of $\hat{Q}_{\psi_1}$ is always zero.

If one is to take $\psi(z) = \psi_2(z) = z^2 - 1$, then it is easily seen by direct calculation from (3.5) that $\Xi_{\psi_2} = 0$. This, of course, is expected since for this choice of ψ, $\frac{1}{\sqrt{n}}\hat{Q}_{\psi_2} = 0$ as seen by plugging in the estimators. However, notice that the quantity

$$\frac{1}{\sqrt{n}} Q_{\psi_2}(\sigma_0^2, \beta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\left(\frac{Y_i - \mathbf{X}_i \beta_{\mathbf{0}}}{\sigma_0} \right)^2 - 1 \right]$$

converges in distribution to a normal variable with mean zero and variance 2. Thus, in this situation as is the case with ψ_1, there is a dramatic effect by replacing the unknown parameters by their maximum likelihood estimates!

Next, consider the $\psi(\cdot)$ function given by

$$\psi(z) \equiv \psi_3(z) = z^3. \tag{3.10}$$

Since $\psi'(z) = 3z^2$, and by using moment properties of the standard normal distribution, we immediately obtain $\mathbf{Var}\{\psi(Z)\} = \mathbf{E}\{Z^6\} = (5)(3)(1) = 15$, $\mathbf{E}\{Z\psi'(Z)\} = 3\mathbf{E}\{Z^3\} = 0$, $\mathbf{E}\{\psi'(Z)\} = 3$, and $\mathbf{E}\{Z\psi(Z)\} = \mathbf{E}\{Z^4\} = 3$. Consequently, from (3.5), we obtain

$$\Xi_{\psi_3} = 15 + 0 + 3[3 - 2(3)](\nu\mathbf{V}^{-1}\nu^{t}) = 15 - 9(\nu\mathbf{V}^{-1}\nu^{t}).$$

For the linear model with intercept term so that $\hat{\Delta} = 1$ from (3.9), an estimate of this asymptotic variance will then be $\hat{\Xi}_{\psi_3} = 15 - 9 = 6$. When $p = 2$, this asymptotic result for the "skewness-based" test statistic

$$\frac{1}{\sqrt{n}}\hat{Q}_{\psi_3} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i}{\hat{\sigma}}\right)^3 \tag{3.11}$$

was first obtained in Pierce (1982). Again, notice the effect of the plug-in procedure, which for the linear model with intercept term is a 60% reduction in the asymptotic variance relative to when the parameters are known. This skewness-type statistic could be used for validating the linear model assumption, e.g., a 5%-level asymptotic test will declare the model inappropriate whenever $|\frac{1}{\sqrt{n}}\hat{Q}_{\psi_3}| > (1.96)\sqrt{6} = 4.80$.

A "kurtosis-based" test statistic is obtained by taking

$$\psi(z) = \psi_4(z) = z^4 - 3. \tag{3.12}$$

For this choice, $\psi'(z) = 4z^3$, and so $\mathbf{Var}\{\psi(Z)\} = \mathbf{E}\{Z^8\} - [\mathbf{E}(Z^4)]^2 = (7)(5)(3)(1) - 3^2 = 96$, $\mathbf{E}\{\psi'(Z)\} = 0$, $\mathbf{E}\{Z\psi'(Z)\} = 4\mathbf{E}\{Z^4\} = 12$, and $\mathbf{E}\{Z\psi(Z)\} = \mathbf{E}\{Z^5 - 3Z\} = 0$. Consequently, from (3.5),

$$\Xi = \Xi_{\psi_4} = 96 + \frac{1}{2}(12)[12 - 2(12)] + 0 = 24, \tag{3.13}$$

which, interestingly, is not affected by the behavior of the matrix $\mathbf{X}$. Thus, under the linear model, the kurtosis-flavored statistic

$$\frac{1}{\sqrt{n}}\hat{Q}_{\psi_4} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\left(\frac{Y_i - \mathbf{X}_i\hat{\beta}}{\hat{\sigma}}\right)^4 - 3\right] \tag{3.14}$$

is asymptotically normal with mean zero and variance 24. This could be used for testing the validity of the model, e.g., an asymptotic 5%-level test will be to declare the linear model inappropriate if $|\frac{1}{\sqrt{n}}\hat{Q}_{\psi_4}| > (1.96)(2)\sqrt{6} = 9.60$. Again, note the impact of the plug-in procedure, which is a four-fold decrease in the asymptotic variance relative to the asymptotic variance if the parameters were known.

More generally, for a positive integer k, we could take $\psi(z) = \psi_{2k}(z) = z^{2k} - \gamma_{2k}$, where $\gamma_{2k} \equiv \prod_{j=1}^{k}(2k - 2j + 1), k = 1, 2, \ldots$. This leads to the statistic

$$\frac{1}{\sqrt{n}}\hat{Q}_{\psi_{2k}} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\left(\frac{Y_i - \mathbf{X}_i\hat{\beta}}{\hat{\sigma}}\right)^{2k} - \gamma_{2k}\right]. \tag{3.15}$$

Analogously to preceding calculations, it is then immediate that the asymptotic variance of (3.15) is given by $\Xi_{\psi_{2k}} = (\gamma_{4k} - \gamma_{2k}^2) - 2k\gamma_{2k}\left[\gamma_{2(k+1)} - (k-1)\gamma_{2k}\right]$. As in the case of earlier ψ choices, this more general statistic could be utilized for model validation purposes. Indeed, an interesting possibility is to consider a vector-valued ψ. For example, for a given positive even integer $K > 3$, one may take

$$\psi(z) = \left[z^3,\ z^4 - \gamma_4,\ z^5,\ z^6 - \gamma_6,\ \cdots,\ z^K - \gamma_K\right]^{\mathrm{t}} \tag{3.16}$$

which will generate a statistic of form

$$\frac{1}{\sqrt{n}}\hat{\mathbf{Q}}_K = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[R_i^3,\ R_i^4 - \gamma_4,\ R_i^5,\ R_i^6 - \gamma_6,\ \cdots,\ R_i^K - \gamma_K\right]^{\mathrm{t}}, \tag{3.17}$$

where $R_i = (Y_i - \mathbf{X}_i\hat{\beta})/\hat{\sigma}, i = 1, 2, \ldots, n$, are the estimated linear model residuals. The limiting distribution of the statistic in (3.17) could be obtained using Theorem 2.1.

For example, if $K = 6$ and under the linear model with intercept term so (3.9) obtains, the statistic

$$\frac{1}{\sqrt{n}}\hat{\mathbf{Q}}_6 = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[R_i^3, R_i^4 - 3, R_i^5, R_i^6 - 15\right]^{\mathrm{t}} \tag{3.18}$$

has, upon applying Theorem 2.1 and straightforward simplifications, an asymptotic mean equal to the zero vector and asymptotic covariance matrix

$$\mathbf{\Xi}_6 = \begin{bmatrix} 15 & 0 & 105 & 0 \\ 0 & 96 & 0 & 900 \\ 105 & 0 & 945 & 0 \\ 0 & 900 & 0 & 10170 \end{bmatrix} - \begin{bmatrix} 9 & 0 & 45 & 0 \\ 0 & 72 & 0 & 540 \\ 45 & 0 & 225 & 0 \\ 0 & 540 & 0 & 4050 \end{bmatrix} = \begin{bmatrix} 6 & 0 & 60 & 0 \\ 0 & 24 & 0 & 360 \\ 60 & 0 & 720 & 0 \\ 0 & 360 & 0 & 6120 \end{bmatrix}. \tag{3.19}$$

The first 4×4 matrix in (3.19) is the asymptotic covariance matrix of $\frac{1}{\sqrt{n}}\mathbf{Q}_6(\sigma_0^2, \beta_0)$, i.e., when the true values of the parameters are known and are not being estimated; while the second 4×4 matrix is the adjustment factor arising from the substitution of the estimators for the unknown parameters. Clearly, this adjustment term is non-negligible. Furthermore, notice that, asymptotically, the first and third components of $\hat{\mathbf{Q}}_6$ are independent from the second and fourth components, which implies that these two sets of components are detecting different features of the distribution of the estimated residuals. A possible model validation test statistic is

$$S_6^2 = \frac{1}{n}\hat{\mathbf{Q}}_6^{\mathrm{t}}\mathbf{\Xi}_6^{-1}\hat{\mathbf{Q}}_6,$$

which, if the linear model assumptions are valid, will have an asymptotic chi-squared distribution with degrees-of-freedom equal to 4. Such a test statistic may have the potential of detecting varied types of deviations from the linear model assumption, thereby generating an omnibus-type and formal model validation procedure.

For the more general test statistic $\hat{\mathbf{Q}}_K$, there is certainly the problem of determining an appropriate value of the order K. We defer though a thorough discussion of these issues in future work as the intent of the present paper is on the general asymptotic result. Instead, we now provide another application of the asymptotic theory in a popular stochastic process model arising in survival analysis and reliability.

3.2 Goodness-of-Fit Testing in a Counting Process Model

Let $N = \{(N_1(t), \ldots, N_n(t)) : t \in [0, \tau]\}$ be an observable multivariate counting process with respect to a filtration $\mathbf{F} = \{\mathcal{F}_t : t \in [0, \tau]\}$. Let the compensator of N be $A = \{(A_1(t), \ldots, A_n(t)) : t \in [0, \tau]\}$ with

$$A_j(t) = \int_0^t Y_j(s)\lambda(s)\exp\{\beta^{\mathrm{t}} Z_j(s)\}\mathrm{d}s, \tag{3.20}$$

where $Y = \{(Y_1(t), \ldots, Y_n(t)) : t \in [0, \tau]\}$ is an observable nonnegative predictable process, $Z = \{(Z_1(t), \ldots, Z_n(t)) : t \in [0, \tau]\}$ is an observable bounded predictable matrix of processes consisting of the $s \times 1$ covariate processes Z_j's, $\lambda(\cdot)$ is an unknown hazard rate function, and $\beta = (\beta_1, \ldots, \beta_s)^{\mathrm{t}}$ is an unknown $s \times 1$ regression coefficient vector. This is the multiplicative intensity model considered by Andersen and Gill (1982), which includes as a special case the Cox proportional hazards model (Cox, 1972).

Suppose it is of interest in this multiplicative intensity model to test the composite null hypothesis that the hazard rate function $\lambda(\cdot)$ belongs to a parametric class of hazard rate function given by $\mathcal{C}_0 = \{\lambda_0(\cdot; \eta) : \eta \in \mathcal{N} \subseteq \Re^r\}$, where the functional form of $\lambda_0(\cdot, \eta)$ is known except for the parameter η. The alternative hypothesis is that $\lambda(\cdot) \notin \mathcal{C}_0$. To develop a formal goodness-of-fit test, let $\{(\psi_1(t), \psi_2(t), \ldots) : t \in [0, \tau]\}$ be a set of basis functions for functions defined on $[0, \tau]$, e.g., trigonometric, polynomial, wavelets, etc., such that if $\lambda(\cdot)$ is the true hazard rate function, then for any $\lambda_0(\cdot, \eta) \in \mathcal{C}_0$,

$$\log\left\{\frac{\lambda(t)}{\lambda_0(t, \eta)}\right\} = \sum_{k=1}^{\infty} \theta_k \psi_k(t), \quad t \in [0, \tau]. \tag{3.21}$$

See Peña (1998) for an application of this hazard-based smooth goodness-of-fit formulation in a simpler model and for other aspects of this approach, such as the choice of the ψ_j's. The idea behind the 'Neyman truncation' (cf., Neyman (1937); Rayner and Best (1989); Fan (1996)) is that it is usually sufficient to truncate the infinite sum on the right-hand side of (3.21) to obtain an acceptable approximation as later coefficients of the expansion will be small in magnitude owing to Parseval's Theorem. Therefore, choose a smoothing order K such that we may expand the left-hand side of (3.21) via

$$\log\left\{\frac{\lambda(t)}{\lambda_0(t, \eta)}\right\} \approx \sum_{k=1}^{K} \theta_k \psi_k(t) = \theta^{\mathrm{t}} \Psi(t), \quad t \in [0, \tau], \tag{3.22}$$

where $\theta = (\theta_1, \ldots, \theta_K)^{\mathrm{t}}$ and $\Psi = (\psi_1, \ldots, \psi_K)^{\mathrm{t}}$. If the null hypothesis holds, then for some $\eta_0 \in \mathcal{N}$, $\theta = 0$. This amounts to embedding the class of possible $\lambda(\cdot)$'s in the wider class

$$\mathcal{C}_K = \left\{\lambda(\cdot, \theta, \eta) = \lambda_0(\cdot, \eta)\exp\{\theta^{\mathrm{t}} \Psi(\cdot)\} : \theta \in \Re^K, \eta \in \mathcal{N}\right\}. \tag{3.23}$$

With this embedding the testing problem is reduced to testing the composite hypothesis $H_0 : \theta = 0, (\eta, \beta) \in \mathcal{N} \times \mathcal{B}$ versus the alternative hypothesis $H_1 : \theta \neq 0, (\eta, \beta) \in \mathcal{N} \times \mathcal{B}$. Note in this testing problem that the parameters η and β are nuisance parameters.

The score function associated with θ based on a realization of $\{(N(t), Y(t)) : t \in [0, \tau]\}$, when evaluated at $\theta = 0$, could be easily shown to be (cf., Borgan (1984); Andersen, Borgan, Gill and Keiding (1993); Peña (1998))

$$Q(\eta, \beta) = \sum_{j=1}^{n} \int_0^{\tau} \Psi \left\{ \mathrm{d}N_j - Y_j \lambda_0(\eta) \exp\{\beta^{\mathrm{t}} Z_j\} \mathrm{d}t \right\}, \tag{3.24}$$

where, for economy of notation, we suppress the argument t in $\Psi(t)$, $N_j(t)$, $Y_j(t)$, $\lambda_0(t, \eta)$, and $Z_j(t)$. However, this is not a statistic since η and β are unknown, so the need to substitute estimators for η and β in $Q(\eta, \beta)$. It is usual to estimate β under this model by solving the estimating equation $S(\beta) = 0$, where

$$S(\beta) = \sum_{j=1}^{n} \int_0^{\tau} [Z_j - E(\beta)] \mathrm{d}N_j \tag{3.25}$$

with $E(t, \beta) = S^{(1)}(t, \beta)/S^{(0)}(t, \beta)$ and $S^{(m)}(t, \beta) = \sum_{j=1}^{n} Z_j^{\otimes m} Y_j \exp\{\beta^{\mathrm{t}} Z_j\}, m = 0, 1, 2$. This is the estimation procedure for β arising from the partial likelihood function, and the resulting estimator is denoted by $\hat{\beta}$. Upon obtaining $\hat{\beta}$, we estimate η by solving in η the profile estimating equation $R(\eta, \hat{\beta}) = 0$, where

$$R(\eta, \beta) = \sum_{j=1}^{n} \int_0^{\tau} \rho(\eta) \left\{ \mathrm{d}N_j - Y_j \lambda_0(\eta) \exp\{\beta^{\mathrm{t}} Z_j\} \mathrm{d}t \right\}, \tag{3.26}$$

where $\rho(t, \eta) = \{\partial/\partial\eta\} \log \lambda_0(t, \eta)$. The resulting estimator is denoted by $\hat{\eta} \equiv \hat{\eta}(\hat{\beta})$. The test statistic for H_0 versus H_1 is some function, e.g., a quadratic form, of the estimated score statistic

$$\hat{Q} = Q(\hat{\eta}, \hat{\beta}) = \sum_{j=1}^{n} \int_0^{\tau} \Psi \left\{ \mathrm{d}N_j - Y_j \lambda_0(\hat{\eta}) \exp\{\hat{\beta}^{\mathrm{t}} Z_j\} \mathrm{d}t \right\}. \tag{3.27}$$

Thus, the asymptotic distribution of $\hat{Q}$ under H_0 becomes of interest.

Under certain regularity conditions, if η_0 is the true parameter value when $\lambda(\cdot)$ is in $\mathcal{C}_0$ and β_0 is the true regression parameter vector, it can be shown that (see, for instance, the technical report Agustin and Peña (2000))

$$\frac{1}{\sqrt{n}} \begin{bmatrix} Q(\eta_0, \beta_0) \\ R(\eta_0, \beta_0) \\ S(\beta_0) \end{bmatrix} \xrightarrow{\mathrm{d}} N \left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & 0 \\ \Sigma_{21} & \Sigma_{22} & 0 \\ 0 & 0 & \Sigma_{33} \end{pmatrix} \right], \tag{3.28}$$

where, with $s^{(m)}(\eta, \beta)$ being the limit in probability of $\frac{1}{n} S^{(m)}(\eta, \beta)$, we have

$$\Sigma_{11} = \int_0^{\tau} \Psi^{\otimes 2} s^{(0)}(\eta_0, \beta_0) \lambda_0(\eta_0) \mathrm{d}s;$$

$$\Sigma_{12} = \Sigma_{21}^{\mathrm{t}} = \int_0^{\tau} \Psi \rho(\eta_0)^{\mathrm{t}} s^{(0)}(\eta_0, \beta_0) \lambda_0(\eta_0) \mathrm{d}t;$$

$$\Sigma_{22} = \int_0^{\tau} \rho(\eta_0)^{\otimes 2} s^{(0)}(\eta_0, \beta_0) \lambda_0(\eta_0) \mathrm{d}t;$$

$$\Sigma_{33} = \int_0^{\tau} v(\beta_0) s^{(0)}(\eta_0, \beta_0) \lambda_0(\eta_0) \mathrm{d}t,$$

where $v(\eta,\beta) = s^{(2)}(\eta,\beta)/s^{(0)}(\eta,\beta) - e(\eta,\beta)^{\otimes 2}$ and $e(\eta,\beta) = s^{(1)}(\eta,\beta)/s^{(0)}(\eta,\beta)$. It is also easy to verify that

$$\frac{1}{n}Q_\eta(\eta_0,\beta_0) \xrightarrow{\text{pr}} A_1 \equiv -\Sigma_{12} \quad \text{and} \quad \frac{1}{n}Q_\beta(\eta_0,\beta_0) \xrightarrow{\text{pr}} A_2 \equiv -\Delta_1,$$

with $\Delta_1 = \int_0^\tau \Psi e(\eta_0,\beta_0)^{\text{t}} s^{(0)}(\eta_0,\beta_0)\lambda_0(\eta_0)\mathrm{d}t$. Furthermore,

$$\frac{1}{n}R_\eta(\eta_0,\beta_0) \xrightarrow{\text{pr}} B_1 \equiv -\Sigma_{22}; \quad \frac{1}{n}R_\beta(\eta_0,\beta_0) \xrightarrow{\text{pr}} B_2 \equiv -\Delta_2, \quad \text{and} \quad \frac{1}{n}S_\beta(\eta_0,\beta_0) \xrightarrow{\text{pr}} C \equiv -\Sigma_{33},$$

with $\Delta_2 = \int_0^\tau \rho(\eta_0)e(\eta_0,\beta_0)^{\text{t}} s^{(0)}(\eta_0,\beta_0)\lambda_0(\eta_0)\mathrm{d}t$.

By applying Theorem 2.1, we therefore obtain that, under H_0, the statistic $\frac{1}{\sqrt{n}}\hat{Q}$ converges in distribution to a zero-mean normal random vector whose covariance matrix is given by

$$\Xi = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} + (\Delta_1 - \Sigma_{12}\Sigma_{22}^{-1}\Delta_2)\Sigma_{33}^{-1}(\Delta_1 - \Sigma_{12}\Sigma_{22}^{-1}\Delta_2)^{\text{t}}. \tag{3.29}$$

A more general result is obtained by considering the distribution under the sequence of alternative hypotheses of form $H_1^n : \theta = \gamma/\sqrt{n} + o(1)$, where γ is a $q \times 1$ direction vector. Under this situation, the asymptotic mean of $\frac{1}{\sqrt{n}}\begin{pmatrix} Q(\eta_0,\beta_0)^{\text{t}} & R(\eta_0,\beta_0)^{\text{t}} & S(\beta_0)^{\text{t}} \end{pmatrix}^{\text{t}}$, under H_1^n, is given by

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} \Sigma_{11} \\ \Sigma_{12} \\ 0 \end{bmatrix} \gamma. \tag{3.30}$$

Consequently, under H_1^n, $\frac{1}{\sqrt{n}}\hat{Q}$ converges in distribution to a normal random vector whose covariance matrix is given in (3.29) and whose mean vector is equal to

$$\nu = \left[\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}\right]\gamma. \tag{3.31}$$

Several things are worth observing from these results. From the covariance matrix in (3.29), we see that plugging-in of $(\hat{\eta},\hat{\beta})$ for (η,β) to obtain the statistic $\hat{Q}$ has no asymptotic effect if $\Sigma_{12} = 0$ and $\Delta_1 = 0$. Under these conditions, we would say that "adaptiveness" (see for instance Bickel, Ritov, Klaasen and Wellner (1993)) obtains in the sense that it does not matter that the nuisance parameters (η,β) are unknown in $Q(\eta,\beta)$ since they can be replaced by their estimators and still have the asymptotic distributions of $Q(\eta_0,\beta_0)$ and $Q(\hat{\eta},\hat{\beta})$ to be identical. The conditions $\Sigma_{12} = 0$ and $\Delta_1 = 0$ are orthogonality conditions between Ψ and $e(\eta_0,\beta_0)$ and between $\rho(\eta_0))$ and $e(\eta_0,\beta_0)$, respectively, with the orthogonality defined with respect to the inner product on the space of square-integrable functions $L^2\{[0,\tau],\nu_0\}$. The inner product is defined, for $f,g \in L^2\{[0,\tau],\nu_0\}$, by $\langle f,g\rangle = \int_0^\tau fg\nu_0(\mathrm{d}t)$, where for a Borel set $A \subseteq [0,\tau]$, $\nu_0(A) = \int_A s^{(0)}(\eta_0,\beta_0)\lambda_0(\eta_0)\mathrm{d}t$.

On the other hand, if these orthogonality conditions are not satisfied, the process of substituting the estimators $(\hat{\eta},\hat{\beta})$ for (η_0,β_0) in $Q(\eta_0,\beta_0)$ have an impact on the resulting asymptotic distribution of $\hat{Q}$. The effect is contained in the last two terms in the expression for Ξ. In particular, if the net effect of these two terms is negative, then quantitatively, the effect is to decrease the variance of Q after the process of plugging-in the estimators. Clearly, ignoring such variance reduction could have dire consequences in the resulting testing procedure. For instance, ignoring such reductions may result in a highly conservative test and thus lead into concluding model appropriateness when in fact the model is inappropriate. Looking at the asymptotic mean in (3.31) associated with the sequence of local alternatives, one notices that the substitution of the estimator has a dampening effect on the power of the test, as the true $\sqrt{n}$-difference between the null and the true model, which is $\Sigma_{11}\gamma$ when (η_0, β_0) are known, gets decreased to $[\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}]\gamma$ after the substitution of the estimators, unless $\Sigma_{12} = 0$ holds. For more elaborate and detailed discussions of some of the consequences of plugging-in estimators of nuisance parameters in a model without covariates, see Peña (1998).

References

Aban, I. and Peña, E. (1999). Properties of test statistics applied to residuals in failure time models. *Journal of Statistical Planning and Inference*, **75**, 181–209.

Agustin, Z. and Peña, E. (2000). Testing Whether the Baseline Hazard in Cox's PH Model Belongs to a Parametric Class, with Applications in Recurrent Events Models. Currently being revised for *Journal of the Royal Statistical Society, B.*

Andersen, P. and Gill, R. (1982). Cox's regression model for counting processes: A large sample study. *The Annals of Statistics*, **10**, 1100–1120.

Andersen, P., Borgan, O., Gill, R. and Keiding, N. (1993). *Statistical Models Based on Counting Processes.* Springer: New York.

Baltazar-Aban, I. and Peña, E. (1995). Properties of Hazard-Based Residuals and Implications in Model Diagnostics. *Journal of the American Statistical Association*, **90**, 185–197.

Bickel, P., Klaasen, C., Ritov, Y. and Wellner, J. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* The Johns Hopkins University Press: Baltimore.

Borgan, O. (1984). Maximum likelihood estimation in parametric counting process models, with applications to censored failure time data. *Scandinavian Journal of Statistics*, **11**, 1–16.

Cox, D. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society B*, **34**, 187–220.

Cox, D. and Snell, E. (1971). On test statistics calculated from residuals. *Journal of the Royal Statistical Society, B*, **30**, 248–275.

Durbin, J. (1973). Weak convergence of the sample distribution function where parameters are estimated. *The Annals of Statistics*, **1**, 279–290.

Fan, J. (1996). Test of significance based on wavelet thresholding and Neyman's truncation. *Journal of the American Statistical Association*, **91**, 674–688.

Loynes, R. (1980). The empirical distribution function of residuals from generalized regression. *The Annals of Statistics*, **8**, 285–298.

Neyman, J. (1937). "Smooth test" for goodness of fit. *Skand. Aktuarietidskrift*, **20**, 149–199.

Peña, E. (1998). Smooth goodness-of-fit tests for composite hypothesis in hazard-based models. *The Annals of Statistics*, **26**, 1935–1971.

Pierce, D. (1982). The asymptotic effect of substituting estimators for parameters in certain types of statistics. *The Annals of Statististics*, **10**, 475–478.

Pierce, D. and Kopecky, K. (1979). Testing goodness of fit for the distribution of errors in regression models. *Biometrika*, **66**, 1–5.

Randles, R. (1982). On the asymptotic normality of statistics with estimated parameters. *The Annals of Statistics*, **10**, 462–474.

Randles, R. (1984). On tests applied to residuals. *Journal of the American Statistical Association*, **79**, 349–354.

Rayner, J. and Best, D. (1989). *Smooth Tests of Goodness of Fit.* Oxford: New York.