DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA 91125

CRAMER-RAO BOUNDS FOR MISSPECIFIED MODELS

Quang H. Vuong

SOCIAL SCIENCE WORKING PAPER 652

October 1986

CRAMER-RAO BOUNDS FOR MISSPECIFIED MODELS

Quang H. Vuong

ABSTRACT

In this paper, we derive some lower bounds of the Cramer-Rao type for the covariance matrix of any unbiased estimator of the pseudo-true parameters in a parametric model that may be misspecified. We obtain some lower bounds when the true distribution belongs either to a parametric model that may differ from the specified parametric model or to the class of all distributions with respect to which the model is regular. As an illustration, we apply our results to the normal linear regression model. In particular, we extend the Gauss-Markov Theorem by showing that the OLS estimator has minimum variance in the entire class of unbiased estimators of the pseudo-true parameters when the mean and the distribution of the errors are both misspecified.

CRAMER-RAO BOUNDS FOR MISSPECIFIED MODELS*

Quang H. Vuong
California Institute of Technology
Pasadena, California 91125

1. INTRODUCTION

The purpose of this paper is to derive some lower bounds of the Cramer-Rao type when the parametric model of interest may not contain the true distribution, i.e., when the model is misspecified. That this situation is frequent arises from the fact that any (finite) parametric model which is sufficiently simple to estimate is likely to be misspecified given the complexity of the economic phenomena. To avoid possible misspecification, alternatives to parametric modelling which are of increasing interest are semi-parametric modelling (see, e.g., Stein (1956), Chamberlain (1984)) and non-parametric modelling (see, e.g., Rosenblatt (1956), Ullah and Singh (1985)).

If one however retains parametric models because of their simplicity, a first and important question is what can be estimated if misspecification is present. When the distance between distributions is measured by the Kullback-Leibler (1951) information criterion, a well-known answer is that one can estimate the closest distribution in the specified parametric model to the true distribution. For, under some regularity conditions, the quasi maximum-likelihood estimator, which is obtained by maximizing the likelihood function associated with the parametric model, is known to be a strongly consistent and asymptotically normal estimator of the pseudo-true parameters characterizing the closest distribution (see, e.g., Huber (1967), White (1982)).

Another important question which naturally follows is how well can the closest distribution or equivalently the pseudo-true parameters be estimated. Indeed, its answer is a prerequisite to the study of efficiency or asymptotic efficiency of estimators in possibly misspecified models. When the parametric model is correctly specified this second question has a widely known answer, which is that the covariance matrix of any unbiased estimator of the true parameters is at least as large, in the positive definite sense, than a bound called the Cramer-Rao lower bound (Rao (1945), Cramer (1946)).[1] The main result of this paper is to obtain a similar result for unbiased estimators of the pseudo-true parameters under general misspecification and hence when nothing is known a priori about the true distribution. It turns out that this new bound reduces to the usual one when the model is known to be correctly specified.

The paper also gives some Cramer-Rao bounds for unbiased estimators of the pseudo-true parameters when the true distribution is known to belong to a parametric model which may differ from the original parametric model. Though the assumption that the true distribution belongs to a parametric model is restrictive, these bounds are not without interest. Indeed, they will be used to derive the general bound when the true distribution is unrestricted except for smoothness and regularity conditions. Second, on theoretical grounds, it is interesting to study the importance of the information that the true distribution belongs to a parametric model on the estimation accuracy that can be achieved when estimating the pseudo-true parameters. Third, on practical grounds, when interest centers on some true parameters, it is sometimes more convenient to estimate a parametric model that differs from the parametric model to which the true distribution is known to belong. For instance, this latter technique was exploited by Gourieroux, Monfort, and Trognon (1984a, b) where interest centers on the first two moments of the true distribution.

Finally, we illustrate our results by studying the homoscedastic normal linear regression model under various types of misspecification. First, we consider cases where the mean is correctly specified but the distributional assumption on the errors is violated. For these cases, we derive some lower bounds under the assumption that the true distribution of the errors belongs to some other parametric family. Because of this parametric assumption, our approach is here more restrictive than the one considered in the semi-parametric literature. Then we consider cases where misspecification arises from incorrect specification of both the mean and the distribution of the errors. For these cases, we derive some lower bounds for unbiased estimators of the pseudo-true parameters when the errors are jointly normal and when the errors are unrestricted except for smoothness and regularity conditions. In particular, we extend the well-known Gauss-Markov Theorem by establishing the optimality of the ordinary least-squares (OLS) estimator in these situations. Hence our results complement recent generalizations of the Gauss-Markov Theorem that have appeared in the semi-parametric literature when the mean is still correctly specified (Hwang (1985), Kariya (1985), and Andrews and Phillips (1985, 1986)).

The paper is organized as follows. In Section 2, we define what we mean by regular and semi-regular models. These models are considered throughout the paper. In Section 3, we derive some Cramer-Rao bounds when it is known that the true distribution belongs to a parametric model which may differ from the specified parametric model. We also study how these bounds change with the available information. In Section 4, we derive a general bound when the true distribution is restricted only to belong to the class of distribution with respect to which the given parametric model is regular. In Section 5, we illustrate our results with the simple homoscedastic normal linear regression model. In particular, we shall establish the optimality of the ordinary least-squares estimator of the pseudo-true parameters in a normal linear regression model which is misspecified with respect to the mean and the distribution of the errors.

2. REGULAR AND SEMI-REGULAR MODELS

Let Y be a $n \times 1$ observed random vector defined on an Euclidean measure space (Y,σ,ν). For instance, in the case of a continuous random vector, Y, σ, and ν are respectively $\mathbb{R}^n$, the Borel σ-algebra on $\mathbb{R}^n$, and the Lebesgue measure on $\mathbb{R}^n$. To fix ideas, Y may be the vector of n independent or dependent observations on a scalar random variables. Let H_0 be the true (joint) cumulative distribution function (cdf) of Y.

To estimate or approximate the true cdf H_0, we specify a parametric model for Y, i.e., a parametric family of joint cdf's $\mathbf{F}_\theta \equiv \{F_\theta \,;\, \theta \in \Theta \subset I\!R^k\}$. We shall not, however, require that H_0 belongs to $\mathbf{F}_\theta$. Thus the model $\mathbf{F}_\theta$ may be misspecified. On the other hand, we shall restrict ourselves to parametric models that are regular with respect to a cdf or a family of cdf's for Y. We now list a set of regularity conditions. Let G be a cdf for Y.

Assumption A1: (a) Θ is compact, and for every $\theta \in \Theta$ the cdf F_θ has a density $f(y\,;\theta)$ with respect to *nu.*[2] (b) The density $f(y;\,\theta)$ is strictly positive for all $(y;\,\theta) \in \mathbf{Y} \times \Theta$, measurable in y for every $\theta \in \Theta$, and twice continuously differentiable on Θ for every $y \in \mathbf{Y}$.

Assumption A2: (a) For every $\theta \in \Theta$, the functions $|\log f(\cdot\,;\theta)|$, $|\partial \log f(\cdot\,;\theta)/\partial\theta|$ and $|\partial^2 \log f(\cdot\,;\theta)/\partial\theta\partial\theta'|$ are dominated by a function $M(\cdot)$ independent of θ and square-integrable with respect to G.

Assumption A3: (a) The function $z\ell_G(\theta) \equiv \int \log f(y\,;\theta)\, dG(y)$ has a unique maximum on Θ at an interior point $\theta_*(G)$. (b) The $k \times k$ matrix $A\ell_G(\theta_*(G))$ is non-singular where

$$A\ell_G(\theta) \equiv \int \frac{\partial^2 \log f(y\,;\theta)}{\partial\theta\partial\theta'}\, dG(y). \tag{2.1}$$

Assumption A4: There exists a neighborhood N_* of $\theta_*(G)$ such that for every $(\theta, \tilde{\theta}) \in N_* \times N_*$ the function $[f(\cdot\,;\theta)]^{-1}\, |\partial f(\cdot\,;\tilde{\theta})/\partial\theta|$ is dominated by a function $M_\theta(\cdot)$ independent of $\tilde{\theta}$ and square-integrable with respect to G.

Let us note that, contrary to Assumptions A2 - A4, Assumption A1 does not depend on G. Assumptions A1 - A3 or similar ones are frequent in the theory of maximum-likelihood (ML) estimation of possibly misspecified parametric models (see, e.g., White (1982)).[3] The value $\theta_*(G)$ is called the pseudo-true value of θ for the model $\mathbf{F}_\theta$ when G is the true cdf H_0 (see, e.g., Sawa (1978)). Assumption A3 - (a) requires that the closest distribution in $\mathbf{F}_\theta$ to G be unique or identified under G when the distance between cdf's is measured by the Kullback- Leibler (1951) information criterion (KLIC):

$$KLIC\,(G, F_\theta) = \int \log \frac{g(y)}{f(y\,;\theta)}\, g(y) d\nu(y). \tag{2.2}$$

where $g(\cdot)$ is, when it exists, the density associated with G.

The relatively unusual Assumption A4 is a local uniform Lipschitz condition on the family $\mathbf{F}_\theta$. It clearly implies the following assumption:

Assumption $A4_*$: There exists a neighborhood N_* of $\theta_*(G)$ such that for every $\theta \in N_*$ the function $[f(\cdot\,;\theta_*(G))]^{-1} \mid \partial f(\cdot\,;\theta)/\partial\theta \mid$ is dominated by a function $M_*(\cdot)$ independent of θ and square-integrable with respect to G.

When $G = F_\theta$ for some $\theta \in \Theta$, Assumption $A4_*$ or similar ones often appear in the derivation of the Cramer-Rao bound (see, e.g., Rao (1973, pp. 324-325)). This assumption essentially allows differentiation under the integral sign of the expectation of any statistic with finite variance under F_θ.[4] Moreover, as seen below, it is used to establish the usual information matrix equivalence when the model $\mathbf{F}_\theta$ is correctly specified.

We are now in a position to define what we mean by a parametric model that is regular with respect to a cdf or a family of cdf's for Y. Let Θ^0 denote the interior of Θ.

DEFINITION 2.1 (Regular Models): A parametric model $\mathbf{F}_\theta = \{F_\theta\,;\theta \in \Theta \subset \mathbb{R}^k\}$ is regular with respect to a cdf G if Assumptions A1 - A4 hold. It is regular with respect to a family $\mathbf{G}$ of cdf's if it is regular with respect to every cdf in $\mathbf{G}$. It is regular if it is regular with respect to $\mathbf{F}_\theta^0 \equiv \{F_\theta\,;\theta \in \Theta^0\}$.

The following lemma summarizes some useful and known properties of parametric models that are regular with respect to a cdf. For any cdf G, let:

$$B_G^f(\theta) = \int \frac{\partial \log f(y;\theta)}{\partial\theta}\,\frac{\partial \log f(y;\theta)}{\partial\theta'}\,dG(y). \tag{2.3}$$

LEMMA 2.1: Let $\mathbf{F}_\theta = \{F_\theta\,;\theta \in \Theta \subset \mathbb{R}^k\}$ be a parametric model for Y which is regular with respect to a cdf G. Then

(i) $z_G^f(\theta)$ is finite and twice continuously differentiable on Θ, and for every $\theta \in \Theta$:

$$\frac{\partial z_G^f(\theta)}{\partial\theta} = \int \frac{\partial \log f(y;\theta)}{\partial\theta}\,dG(y) < \infty, \tag{2.4}$$

$$\frac{\partial^2 z_G^f(\theta)}{\partial\theta\partial\theta'} = A_G^f(\theta) < \infty, \tag{2.5}$$

$B_G^f(\theta) < \infty$, and $A_G^f(\theta_*(G))$ is negative definite (n.d.).

(ii) If $G = F_{\theta_0}$ for some $\theta_0 \in \Theta^0$, then $\theta_*(G) = \theta_0$, and

$$A_G^f(\theta_0) + B_G^f(\theta_0) = 0.^5 \tag{2.6}$$

Equation (2.6) is recognized to be the familiar information matrix equivalence under correct specification of the model $\mathbf{F}_\theta$. The framework adopted here is however quite general as the examples of Section 5 illustrate. Its generality arises from the fact that the family $\mathbf{G}$ of cdf's with respect to which $\mathbf{F}_\theta$ is assumed regular need not be equal nor included in $\mathbf{F}_\theta$. Since the family $\mathbf{G}$ will be the family to which the true cdf H_0 is thought or restricted to belong, it follows that the true data generating process characterized by H_0 may be quite different from the one implicit in the specification of the model $\mathbf{F}_\theta$. Moreover, the family $\mathbf{G}$ may be itself parametric as in Section 3, or quite broad as in Section 4 where it is taken to be the class of all cdf's for Y with respect to which the given parametric model $\mathbf{F}_\theta$ is regular.

When $\mathbf{G}$ is a parametric model $\mathbf{G}_\gamma = \{G_\gamma ; \gamma \in \Gamma \subset \mathbb{R}^p\}$, it can be for instance taken to be regular, i.e., regular with respect to $\mathbf{G}_\gamma^0$. For the result of this paper, a more useful class of parametric models G_γ is that of semi-regular models. Formally, let:

Assumption B1: (a) Γ is open, and for every $\gamma \in \Gamma$ the cdf G_γ has a density $g(y;\gamma)$ with respect to ν. (b) The density $g(y;\gamma)$ is strictly positive for all $(y,\gamma) \in \mathbf{Y} \times \Gamma$, measurable in y for every $\gamma \in \Gamma$, and once continuously differentiable on Γ for every $y \in \mathbf{Y}$.

Assumption B2: For every $\gamma \in \Gamma$, there exists a neighborhood N_γ of γ such that for every $\tilde{\gamma} \in N_\gamma$, the function $[g(\cdot\,;\gamma)]^{-1} \mid \partial g(\cdot\,;\tilde{\gamma}) / \partial\gamma \mid$ is dominated by a function $M_\gamma(\cdot)$ independent of $\tilde{\gamma}$ and square-integrable with respect to G_γ.

Assumption B2 corresponds to the local Lipschitz Assumption $A4_*$ that must hold for every G_γ in $\mathbf{G}_\gamma$. For, $\gamma_*(G_\gamma)$ can be taken to be γ which is justified since G_γ is the closest distribution in $\mathbf{G}_\gamma$ to G_γ. A semi-regular model is defined formally as follows.

DEFINITION 2.2 (Semi-Regular Models): A parametric model $\mathbf{G}_\gamma = \{G_\gamma ; \gamma \in \Gamma \subset \mathbb{R}^p\}$ is semi-regular if Assumptions B1 - B2 hold.

The essential difference between semi-regular models and regular models is that the former models need not satisfy the domination conditions of Assumption A2 and the uniqueness requirement of Assumption A3.[6] In particular, there may exist more than one solution in $\tilde{\gamma} \in \Gamma$ to the equation $G_{\tilde{\gamma}} = G_\gamma$ for every $\gamma \in \Gamma$. Nonetheless, semi-regular models enjoy some properties of regular models, as stated in the next lemma.

LEMMA 2.2: Let $\mathbf{G}_\gamma = \{G_\gamma ; \gamma \in \Gamma \subset \mathbb{R}^p\}$ be a parametric model for Y which is semi-regular. Then, for every $G \in G_\gamma$, and every $\gamma \in \Gamma$ such that $G = \mathbf{G}_\gamma$:

$$\int \frac{\partial \log g(y;\gamma)}{\partial\gamma}\, dG(y) = 0, \tag{2.7}$$

$$B_G^\gamma(\gamma) \equiv \int \frac{\partial \log g(y;\gamma)}{\partial\gamma} \frac{\partial \log g(y;\gamma)}{\partial\gamma'}\, dG(y) < \infty, \tag{2.8}$$

though $B_G^\gamma(\gamma)$ may be singular.

3. CRAMER-RAO BOUNDS UNDER PARAMETRIC INFORMATION

Given a possibly misspecified parametric model $\mathbf{F}_\theta$ for Y, the question of interest is how well can the closest distribution in $\mathbf{F}_\theta$ to the true cdf H_0 be estimated. When the parametric model $\mathbf{F}_\theta$ is regular with respect to H_0 and when the distance between cdf's is measured by the KLIC (2.2), it follows from Assumption A3 - (a) that an equivalent question is how well can the pseudo-true parameter $\theta_*(H_0)$ be estimated.

In this section, we shall derive a lower bound of the Cramer and Rao type for any unbiased estimator of the pseudo-true parameters $\theta_*(H_0)$ under the assumption that the true distribution H_0 belongs to a parametric model $\mathbf{G}_\gamma$ which may differ from the specified parametric model $\mathbf{F}_\theta$. Specifically, $\mathbf{G}_\gamma$ will be assumed to be semi-regular. Thus, its number of parameters will be finite. As the examples of Section 5 illustrate, however, this does not prevent the number of parameters of $\mathbf{G}_\gamma$ to increase with the sample size.

Following the usual derivation of the Cramer-Rao lower bound, we first define the concept of unbiasedness.[7] To allow for possible misspecification, we propose the following definition.

DEFINITION 3.1 (Unbiasedness): Let $\mathbf{G}$ be a family of cdf's for Y with respect to which the specified parametric model $\mathbf{F}_\theta = \{F_\theta\,; \theta \in \Theta \subset I\!R^k\}$ is regular. Let $\phi(\cdot)$ be a mapping from Θ to $\Phi \subset I\!R^s$, and let $T(Y)$ be a statistic taking its values in Φ. Then $T(Y)$ is an unbiased estimator of $\phi(\theta_*(G))$ under $\mathbf{G}$ if and only if:

$$\int T(y)\, dG(y) = \phi(\theta_*(G)), \ \forall\, G \in \mathbf{G}. \tag{3.1}$$

As usual, the function $\phi(\cdot)$ introduces some flexibility in the choice of the parameters of interest. Specifically, we may be interested in subsets or more generally in functions of the pseudo-true parameters $\theta_*(G)$. Let us note that the requirement that $\mathbf{F}_\theta$ be regular with respect to $\mathbf{G}$ implies that $\theta_*(G)$ is identified under any cdf G in $\mathbf{G}$ (see Section 2). If $\mathbf{G}$ is identical to $\mathbf{F}_\theta$, then it follows from Assumption A3 - (a) and Jensen inequality that $\theta_*(G) = \theta$ when $G = F_\theta$. Hence, Equation (3.1) becomes equivalent to:

$$\int T(y)\, dF_\theta(y) = \phi(\theta), \ \forall\, \theta \in \Theta. \tag{3.2}$$

Thus Definition 3.1 extends the usual definition to the case where the parametric model $\mathbf{F}_\theta$ may be misspecified.

In this section, the family $\mathbf{G}$ is a parametric model $\mathbf{G}_\gamma = \{G_\gamma\,; \gamma \in \Gamma \subset I\!R^p\}$ which is semi-regular. The following lemma is useful. It gives some additional properties of parametric models that are regular with respect to a semi-regular model. Let,

$$z^f(\theta,\gamma) \equiv \int \log f(y\,;\theta)\, dG_\gamma(y), \tag{3.3}$$

$$B_G^f(\theta,\gamma) \equiv \int \frac{\partial \log f(y\,;\theta)}{\partial\theta}\, \frac{\partial \log g(y\,;\gamma)}{\partial\gamma'}\, dG(y). \tag{3.4}$$

Note that $z^f(\theta,\gamma) = z^f_{G_\gamma}(\theta)$ as defined in Assumption A3.

LEMMA 3.1: Let $\mathbf{F}_\theta = \{F_\theta\, ; \theta \in \Theta \subset \mathbb{R}^k\}$ be a parametric model which is regular with respect to a semi-regular model $\mathbf{G}_\gamma = \{G_\gamma\, ; \gamma \in \Gamma \subset \mathbb{R}^p\}$. Then

(i) the partial derivatives $\partial z^f(\theta,\gamma)/\,\partial\theta$ exist and are continuously differentiable in both θ and γ on $\Theta \times \Gamma$ with:

$$\frac{\partial^2 z^f(\theta,\gamma)}{\partial\theta\partial\theta'} = A^f_{G_\gamma}(\theta) < \infty, \tag{3.5}$$

$$\frac{\partial^2 z^f(\theta,\gamma)}{\partial\theta\partial\gamma'} = B^f_{G_\gamma}(\theta,\gamma) < \infty, \tag{3.6}$$

(ii) the function $\theta_*(G_\gamma)$ is continuously differentiable in $\gamma \in \Gamma$ with

$$\frac{\partial\, \theta_*(G_\gamma)}{\partial\gamma'} = -[A^f_{G_\gamma}(\theta_*(G_\gamma))]^{-1} B^{fg}_{G_\gamma}(\theta_*(G_\gamma),\gamma) \tag{3.7}$$

where $A^f_{G_\gamma}(\theta_*(G_\gamma))$ is negative definite for every $\gamma \in \Gamma$.

We can now state the main result of this section. Let $Var_G\, T(Y)$ denote the variance of the statistic $T(Y)$ under the cdf G. Since $B^g_G(\gamma)$ is not insured to be non-singular (see Section 2), we shall use generalized ($g-$) inverses (see Rao and Mitra (1971)). Finally, to simplify the notation, we shall sometimes use θ_* instead of $\theta_*(G)$ when there is no ambiguity.

THEOREM 3.1 (Cramer-Rao Bound Under Parametric Information): Let $\mathbf{F}_\theta = \{F_\theta\, ; \theta \in \Theta \subset \mathbb{R}^k\}$ be a parametric model for Y which is regular with respect to the semi-regular model $\mathbf{G}_\gamma = \{G_\gamma\, ; \gamma \in \Gamma \subset \mathbb{R}^p\}$. Let $\phi(\cdot)$ be a continuously differentiable mapping from Θ to $\Phi \subset \mathbb{R}^s$, and $T(Y)$ be an unbiased estimator of $\phi(\theta_*)$ with finite variance under any cdf G in $\mathbf{G}_\gamma$. Then, for every $G \in \mathbf{G}_\gamma$, and every γ such that $G_\gamma = G$,

$$Var_G\; T(Y) \geq LB_G(\gamma) \tag{3.8}$$

in the positive semi-definite (psd) sense, where

$$LB_G(\gamma) = \frac{\partial\phi(\theta_*)}{\partial\theta'}\, [A^f_G(\theta_*)]^{-1}\, B^{fg}_G(\theta_*,\gamma)\, [B^g_G(\gamma)]^-\, B^{gf}_G(\gamma,\theta_*)\, [A^f_G(\theta_*)]^{-1}\, \frac{\partial\phi'(\theta_*)}{\partial\theta}, \tag{3.9}$$

and $[B^g_G(\gamma)]^-$ is any symmetric reflexive g-inverse of $B^g_G(\gamma)$. In addition, all the matrices exist and $LB_G(\gamma)$ is independent of any choice of symmetric reflexive g-inverses.[8,9]

Since the lower bound $LB_G(\gamma)$ depends only on $\mathbf{F}_\theta$ and $\mathbf{G}_\gamma$, Theorem 3.1 says that the covariance matrix of any unbiased estimator of $\phi(\theta_*)$ is not smaller than a quantity that is independent of any method for estimating the pseudo-true parameters θ_*. Though this may appear surprising, it is in fact a direct consequence of the well-known Cramer-Rao lower bound. Indeed, we are acting as if the true cdf H_0 belongs to the parametric model $\mathbf{G}_\gamma$ so that $H_0 = G_\gamma$ for some $\gamma \in \Gamma$. Since the pseudo-true parameters $\theta_* = \theta_*(G_\gamma)$ is a particular function of the true parameters γ, it follows from the Cramer-Rao lower bound that, under suitable regularity conditions, we have:

$$Var_{G_\gamma} T(Y) \geq \frac{\partial \phi^c(\gamma)}{\partial \gamma'} [B^g_{G_\gamma}(\gamma)]^{-1} \frac{\partial \phi^c(\gamma)'}{\partial \gamma}, \forall \gamma \in \Gamma, \tag{3.10}$$

where $\phi^c(\gamma)$ is the composite function $\phi(\theta_*(G_\gamma))$. Then, the inequality (3.8) follows from the chain rule and Equation (3.7). It only extends the inequality (3.10) to allow for non-singularity of the information matrix $B^g_G(\gamma)$ and possibly more than one parameter value γ such that $G_\gamma = G$.

Given the previous remark, it follows that all the results on minimum variance unbiased estimation apply (see, e.g., Rao (1973), Section 5a). For instance, the lower bound (3.8) is attained if and only if the parametric densities $g(\cdot\,;\gamma)$ are of the form:

$$g(y;\gamma) = \exp\{\lambda'(\gamma)T(y) + \mu(\gamma) + \tau(y)\}$$

for some vector function $\lambda(\cdot)$ and function $\mu(\cdot)$ and $\tau(\cdot)$. That is, the model $\mathbf{G}_\gamma$ is exponential and the estimator $T(Y)$ is a sufficient statistic for γ.

As another immediate consequence of Theorem 3.1, we obtain the familiar Cramer-Rao result by considering the special case $\mathbf{G}_\gamma = \mathbf{F}^0_\theta$. Since $\mathbf{G}_\gamma$ contains the true distribution H_0, this case corresponds to the usual one where $\mathbf{F}_\theta$ is assumed correctly specified. Let $Var_\theta T(Y)$ be the variance of the statistic $T(Y)$ under the cdf F_θ. Let

$$A^f(\theta) \equiv \int \frac{\partial^2 \log f(y;\theta)}{\partial\theta\partial\theta'} dF_\theta(y), \tag{3.11}$$

$$B^f(\theta) \equiv \int \frac{\partial \log f(y;\theta)}{\partial\theta} \frac{\partial \log f(y;\theta)}{\partial\theta'} dF_\theta(y). \tag{3.12}$$

COROLLARY 3.1: Let $\mathbf{F}_\theta = \{F_\theta\,;\theta \in \boldsymbol{\Theta} \subset \mathbb{R}^k\}$ be a regular model for Y. Let $\phi(\cdot)$ be a continuously differentiable mapping from $\boldsymbol{\Theta}$ to $\boldsymbol{\Phi} \subset \mathbb{R}^s$, and $T(Y)$ be an unbiased estimator of $\phi(\theta)$ with finite variance for every $\theta \in \boldsymbol{\Theta}^0$. Then, for every $\theta \in \boldsymbol{\Theta}^0$:

$$Var_\theta T(Y) \geq \frac{\partial\phi(\theta)}{\partial\theta'} [B^f(\theta)]^{-1} \frac{\partial\phi(\theta)'}{\partial\theta} \tag{3.13}$$

where $B^f(\theta)$ exists, is non-singular, and $A^f(\theta) + B^f(\theta) = 0$ for every $\theta \in \boldsymbol{\Theta}^0$.

The importance of Theorem 3.1, however, is that it applies whether the models $\mathbf{F}_\theta$ and $\mathbf{G}_\gamma$ are nested, overlapping, or strictly non-nested.[10] To obtain a better understanding of this result, one may consider the following two questions. First, one may ask how the lower bound (3.9) varies when one considers another parametric model $\mathbf{F}_\alpha = \{\tilde{F}_\alpha : \alpha \in A \subset \mathbb{R}^a\}$ instead of $\mathbf{F}_\theta$.[11] Specifically, suppose that one knows that the true distribution belongs to a parametric model $\mathbf{G}_\gamma$ and that one is interested in estimating some functions of the true parameters. But, suppose that for computational simplicity, one prefers to estimate another parametric model $\mathbf{F}_\theta$. Can this model $\mathbf{F}_\theta$ be chosen so as to maximize the estimation accuracy of the functions of interest? The answer is in fact trivial and is given in the following corollary. To indicate the possible dependence of the lower bound (3.9) on the model $\mathbf{F}_\theta$, we use the notation $LB_G(\gamma, \mathbf{F}_\theta)$.

COROLLARY 3.2: Let $\mathbf{F}_\theta = \{F_\theta ; \theta \in \Theta \subset \mathbb{R}^k\}$ and $\mathbf{F}_\alpha = \{\tilde{F}_\alpha ; \alpha \in A \subset \mathbb{R}^a\}$ be two parametric models for Y which are regular with respect to the semi-regular model $\mathbf{G}_\gamma = \{G_\gamma ; \gamma \in \Gamma \subset \mathbb{R}^p\}$. Let $\phi(\cdot)$ and $\tilde{\phi}(\cdot)$ be continuously differentiable mappings from Θ to Φ and A to $\Phi \subset \mathbb{R}^s$ respectively. Suppose that:

$$\phi(\theta_*(G)) = \tilde{\phi}(\alpha_*(G)), \ \forall\, G \in \mathbf{G}_\gamma. \tag{3.14}$$

Then, for every $G \in \mathbf{G}_\gamma$, and every γ such that $G_\gamma = G$,

$$LB_G(\gamma, \mathbf{F}_\theta) = LB_G(\gamma, \mathbf{F}_\alpha). \tag{3.15}$$

Condition (3.14) ensures that one can identify and estimate the same functions of the true parameters under either parametric models $\mathbf{F}_\theta$ and $\mathbf{F}_\alpha$. Equation (3.15) says that the lower bound is independent of the choice of the parametric model $\mathbf{F}_\theta$. In fact, this result is obvious since one is estimating the same functions of the true parameters.

A second question is how the lower bound (3.9) varies with the model $\mathbf{G}_\gamma$. Since the true distribution H_0 is assumed to belong to $\mathbf{G}_\gamma$, the issue is how the lower bound (3.9) varies with the available information. The answer is given in the following corollary. To indicate the dependence of the lower bound (3.9) on the model $\mathbf{G}_\gamma$, we use the notation $LB_G(\mathbf{G}_\gamma, \gamma)$.

COROLLARY 3.3: Let $\mathbf{F}_\theta = \{F_\theta ; \theta \in \Theta \subset \mathbb{R}^k\}$ be a parametric model for Y which is regular with respect to the semi-regular models $\mathbf{G}_\gamma = \{G_\gamma ; \gamma \in \Gamma \subset \mathbb{R}^p\}$ and $\mathbf{G}_\beta = \{\tilde{G}_\beta ; \beta \in B \subset \mathbb{R}^b\}$. Suppose that there exists a continuously differentiable mapping $\lambda(\cdot)$ from Γ to B such that:

$$G_\gamma = \tilde{G}_{\lambda(\gamma)}, \ \forall\, \gamma \in \Gamma. \tag{3.16}$$

Then, for every $G \in \mathbf{G}_\gamma$, and every γ such that $G_\gamma = G$,

$$LB_G(\mathbf{G}_\gamma, \gamma) \le LB_G(\mathbf{G}_\beta, \lambda(\gamma)). \tag{3.17}$$

Moreover, the equality holds if $p = b$ and $B\ell(\gamma)$ is non-singular.[12]

Condition (3.16) essentially requires that the model $\mathbf{G}_\gamma$ be nested in the model $\mathbf{G}_\beta$. Thus the model $\mathbf{G}_\gamma$ contains more information about the true cdf H_0 than the model $\mathbf{G}_\beta$. It is therefore expected that the estimation accuracy that can be achieved when estimating the pseudo-true parameters θ_* or the functions $\phi(\theta_*)$ is improved when it is known that the true distribution belongs to $\mathbf{G}_\gamma$ than when it belongs to $\mathbf{G}_\beta$. This intuitively explains the inequality (3.17). On the other hand, disregarding the non-singularity of $B_G^\delta(\gamma)$ which is a weak condition, the second part of Corollary 3.2 says that if $\mathbf{G}_\gamma$ and $\mathbf{G}_\beta$ are nested but with the same dimension ($p = b$) then there cannot be any improvement in the lower bound. In particular, if $\tilde{\mathbf{G}}_\gamma = \{G_\gamma ; \gamma \in \tilde{\Gamma} \subset \mathbb{R}^b\}$ where $\tilde{\Gamma}$ contains Γ so that $\tilde{\mathbf{G}}_\gamma$ contains $\mathbf{G}_\gamma$, then

$$LB_G(\mathbf{G}_\gamma, \gamma) = LB_G(\tilde{\mathbf{G}}_\gamma, \gamma) \tag{3.18}$$

for every $G \in \mathbf{G}_\gamma$, and every γ such that $G_\gamma = G$. In other words, the additional information gained by going from $\tilde{\mathbf{G}}_\gamma$ to $\mathbf{G}_\gamma$ is irrelevant. This somewhat surprising result can nonetheless be explained by the local nature of the lower bound (3.9).

4. A GENERAL LOWER BOUND UNDER MISSPECIFICATION

In the previous section, the true distribution H_0 was restricted to belong to a parametric model $\mathbf{G}_\gamma$. In this section, we shall drop such an assumption. Specifically, we shall derive a lower bound for the covariance matrix of any unbiased estimator of the pseudo-true parameters $\theta_*(H_0)$ when H_0 is unrestricted except for smoothness and regularity conditions. More precisely, H_0 will be restricted only to belong to the class of all cdf's with respect to which the specified parametric model $\mathbf{F}_\theta$ is regular. Given the identification requirement of Assumption A2, such a class is a natural one to consider.

As in the semi-parametric literature (see, e.g., Stein (1956), Chamberlain (1984)), our approach is to consider a least favorable parametric model containing H_0.[13] This model will be constructed so that the specified parametric model $\mathbf{F}_\theta$ is regular with respect to it. In addition, it will be semi-regular so that Theorem 3.1 applies. The next lemma exhibits such a parametric model.

LEMMA 4.1: Let $\mathbf{F}_\theta = \{F_\theta ; \theta \in \Theta \subset \mathbb{R}^k\}$ be a parametric model for Y that is regular with respect to a ν-absolutely continuous cdf G. Then there exists a neighborhood N_0 of θ_* such that:

(i) the parametric model $\mathbf{G}_\theta = \{G_\theta ; \theta \in N_0 \subset \mathbb{R}^k\}$ is semi-regular, where

$$\frac{dG_\theta(y)}{d\nu} \equiv g(y;\theta) = \frac{1}{C(\theta)}\{1 + \exp[1 - \frac{f(y;\theta)}{f(y;\theta_*)}]\}\, g(y), \tag{4.1}$$

$$C(\theta) = \int \{1 + \exp[1 - \frac{f(y;\theta)}{f(y;\theta_*)}]\} dG(y), \tag{4.2}$$

(ii) $\mathbf{F}_\theta$ is regular with respect to $\mathbf{G}_\theta$.

Since the neighborhood N_0 contains θ_*, and since $G_{\theta_*} = G$, it follows that the model $\mathbf{G}_\theta$ contains G, as required.[14] We are now in a position to state the main result of this paper.

THEOREM 4.1 (A General Cramer-Rao Bound): Let $\mathbf{F}_\theta = \{F_\theta \,;\, \theta \in \Theta \subset \mathbb{R}^k\}$ be a parametric model for Y. Let $\mathbf{G}(\mathbf{F}_\theta)$ be the class of all ν-absolutely continuous cdf's for Y with respect to which $\mathbf{F}_\theta$ is regular. Suppose that $\mathbf{G}(\mathbf{F}_\theta)$ is not empty. Let $\phi(\cdot)$ be a continuously differentiable mapping from Θ to $\Phi \subset \mathbb{R}^s$, and $T(Y)$ be an unbiased estimator of $\phi(\theta_*)$ with finite variance under $\mathbf{G}(\mathbf{F}_\theta)$. Then, for every G in $\mathbf{G}(\mathbf{F}_\theta)$:

$$Var_G\; T(Y) \geq LB_G \tag{4.3}$$

where

$$LB_G = \frac{\partial\phi(\theta_*)}{\partial\theta'}\,[A_G^f(\theta_*)]^{-1}\, B_G^f(\theta_*)[A_G^f(\theta_*)]^{-1}\,\frac{\partial\phi'(\theta_*)}{\partial\theta}, \tag{4.4}$$

and all the matrices exist. Moreover, for any $G \in \mathbf{G}(\mathbf{F}_\theta)$, let $\mathbf{S}_G$ denote the set of all semi-regular models $\mathbf{G}_\gamma$ containing G and with respect to which $\mathbf{F}_\theta$ is regular. Then $\mathbf{S}_G$ is non-empty, and

$$LB_G = \max LB_G(\mathbf{G}_\gamma, \gamma) \tag{4.5}$$

where the max is taken over all models $\mathbf{G}_\gamma \in \mathbf{S}_G$ and $\gamma \in \Gamma$ such that $G_\gamma = G$.

Recall that psd matrices are note totally ordered in the psd sense. The meaning of Equation (4.5), however, is that the general lower bound LB_G is at least as large as the lower bound $LB_G(\mathbf{G}_\gamma, \gamma)$ associated with any parametric model $\mathbf{G}_\gamma$ in $\mathbf{S}_G$, and that it is attained for at least one model in $\mathbf{S}_G$. As the proof shows, the bound LB_G is attained when $\mathbf{G}_\gamma$ is the model $\mathbf{G}_\theta$ of Lemma 4.1 Thus this model $\mathbf{G}_\theta$ is indeed a least favorable semi-regular model containing G and with respect to which the specified model $\mathbf{F}_\theta$ is regular.

Let us also note that the assumption that $\mathbf{G}(\mathbf{F}_\theta)$ be non-empty makes sense. For, from Assumption A3 - (a), it requires that there exists at least one cdf G under which the closest distribution in $\mathbf{F}_\theta$ to G or equivalently $\theta_*(G)$ is identified and hence estimable. In addition, such an assumption is satisfied if $\mathbf{F}_\theta$ is regular with respect to itself, in which case the class $\mathbf{G}(\mathbf{F}_\theta)$ must contain $\mathbf{F}_\theta$.

Third, it is interesting to note that the general bound (4.4) reduces to the familiar Cramer-Rao bound when the model $\mathbf{F}_\theta$ is correctly specified. Specifically, suppose that $G = F_\theta$ for some $\theta \in \Theta^0$. Then, for any such cdf G, it follows from Lemma 2.1 - (ii) that the general lower bound becomes

$$LB_G = \frac{\partial\phi(\theta)}{\partial\theta'}\,[B_G^f(\theta)]^{-1}\,\frac{\partial\phi'(\theta)}{\partial\theta}, \tag{4.6}$$

which is the usual bound (see Corollary 3.1).

Finally, we can obtain results similar to Corollaries 3.2 and 3.3 when the true distribution is no longer restricted to belong to a parametric model. Specifically, we shall consider two parametric models $\mathbf{F}_\theta$ and $\mathbf{F}_\alpha$ for Y. To indicate the dependence of the general lower bound (4.4) on the model $\mathbf{F}_\theta$, we use the notation $LB_G(\mathbf{F}_\theta)$. We have:

COROLLARY 4.1: Let $\mathbf{F}_\theta = \{F_\theta ; \theta \in \Theta \subset \mathbb{R}^k\}$ and $\mathbf{F}_\alpha = \{\tilde{F}_\alpha ; \alpha \in A \subset \mathbb{R}^a\}$ be two parametric models for Y. Let $\mathbf{G}(\mathbf{F}_\theta)$ and $\mathbf{G}(\mathbf{F}_\alpha)$ be the classes of ν-absolutely continuous cdf's with respect to which these models are regular. Let $\phi(\cdot)$ and $\tilde{\phi}(\cdot)$ be continuously differentiable mappings from Θ to Φ and A to $\Phi \subset \mathbb{R}^s$. Suppose that

$$\phi(\theta_*(G)) = \tilde{\phi}(\alpha_*(G)), \ \forall\, G \in \mathbf{G}(\mathbf{F}_\theta) \cap \mathbf{G}(\mathbf{F}_\alpha). \tag{4.7}$$

(i) If $\mathbf{G}(\mathbf{F}_\theta) = \mathbf{G}(\mathbf{F}_\alpha)$, then for every $G \in \mathbf{G}(\mathbf{F}_\theta) \cap \mathbf{G}(\mathbf{F}_\alpha)$,

$$LB_G(\mathbf{F}_\theta) = LB_G(\mathbf{F}_\alpha). \tag{4.8}$$

(ii) If $\mathbf{G}(\mathbf{F}_\theta) \subset \mathbf{G}(\mathbf{F}_\alpha)$, then for every $G \in \mathbf{G}(\mathbf{F}_\theta) \cap \mathbf{G}(\mathbf{F}_\alpha)$,

$$LB_G(\mathbf{F}_\theta) \leq LB_G(\mathbf{F}_\alpha). \tag{4.9}$$

As in Corollary 3.2, Condition (4.7) ensures that one can identify the same functions of the true distribution $H_0 = G$ when considering either model $\mathbf{F}_\theta$ or $\mathbf{F}_\alpha$. Part (i) says that if both models are regular with respect to the same class of cdf's, then there cannot be any improvement in the lower bound associated with either one of the models. Part (ii) says that if the class of cdf's with respect to which one model is regular is larger than for the other model, then the lower bound associated with the former must be at least as large as for the latter.

5. EXAMPLES

In this section, we illustrate our results with the simple normal linear regression model. Specifically, let $Y = (Y_1, \ldots, Y_n)'$ where Y_i is the i-th observation on a scalar random variable. Suppose that one postulates the following normal linear regression model:

$$Y_i = z_i'\beta + \varepsilon_i, i = 1, \ldots, n \tag{5.1}$$

where $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_1)'$ is normally distributed $N(0,\sigma^2 I)$, and z_i and β are k_0-dimensional vectors. Throughout, Z is treated as a non-random matrix of full-column rank.[15] Then, the specified model $\mathbf{F}_\theta$ for Y is defined by the (joint) densities:

$$f(y; \theta) = (2\pi\sigma^2)^{-n/2} \exp[-\frac{1}{2\sigma^2}(y - Z\beta)'(y - Z\beta)], \tag{5.2}$$

where θ is the k-dimensional vector $(\beta', \sigma^2)'$. We shall take the parameter space Θ to be of the form $\mathbf{B} \times [a, b]$ where $\mathbf{B}$ is a compact subset of $\mathbb{R}^{k_*}$ and $0 < a < b$. Thus Assumption A1 is satisfied, and we have:

$$\frac{\partial \log f(y;\theta)}{\partial\theta} = \begin{bmatrix} \frac{1}{\sigma^2} Z'(y - Z\beta) \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(y - Z\beta)'(y - Z\beta) \end{bmatrix}, \tag{5.3}$$

$$\frac{\partial^2 \log f(y;\theta)}{\partial\theta\partial\theta'} = \begin{bmatrix} -\frac{1}{\sigma^2} Z'Z & ; -\frac{1}{\sigma^4} Z'(y - Z\beta) \\ -\frac{1}{\sigma^4}(y - Z\beta)'Z\ ; & \frac{n}{2}\sigma^4 - \frac{1}{\sigma^6}(y - Z\beta)'(y - Z\beta) \end{bmatrix}. \tag{5.4}$$

For various reasons (see below), the joint density of Y may not, however, be of the normal form (5.2) with some θ in Θ. It follows that the normal linear regression model $\mathbf{F}_\theta$ may be misspecified. Then it is useful to characterize the joint distributions for Y with respect to which the normal linear regression model is regular. This is the purpose of the next lemma. For any cdf G, let $\mu_i \equiv E_G(Y_i)$ and $\sigma_i^2 = Var_G(Y_i)$, where we have omitted the dependence on G for notational simplicity. Let $\mu = (\mu_1, \ldots, \mu_n)'$ and $M_Z = I - Z(Z'Z)^{-1}Z'$.

LEMMA 5.1: Let $\mathbf{F}_\theta$ be the normal linear regression model (5.1) with $\Theta = \mathbf{B} \times [a, b]$, and G be a cdf for Y.

(i) Assumptions A2 - A3 hold if and only if $E_G(Y_i^4) < \infty$ for every i, and $[\mu' Z(Z'Z)^{-1}, \frac{1}{n}(\sum_{i=1}^{n} \sigma_i^2 + \mu' M_Z \mu)]' \in \Theta^0$, in which case

$$\beta_*(G) = (Z'Z)^{-1} Z'\mu, \tag{5.5}$$

$$\sigma_*^2(G) = \frac{1}{n}(\sum_{i=1}^{n} \sigma_i^2 + \mu' M_Z \mu). \tag{5.6}$$

(ii) Assumption A4 holds if the moment generating function of $Y'Y$ exists, i.e., if there exists a $t_0 > 0$, $\forall\, t \in (-t_0, t_0)$, $E_G[e^{tY'Y}] < \infty$.

The condition of Part (ii) is relatively strong since it implies that the moment generating function of the random vector Y exists, and hence that all the moments of Y exist (see, e.g., Lehmann (1983, p. 30), Monfort (1980, p. 149)). Hopefully, this condition is only sufficient unlike the condition of Part (i). In addition, many distributions for Y do satisfy that condition. For instance, it is clearly satisfied if Y is multivariate normal or multivariate logistic. Thus, the class $\mathbf{G}(\mathbf{F}_\theta)$ with respect to which the normal linear regression model $\mathbf{F}_\theta$ is regular is in fact quite large.

As mentioned earlier, there are many reasons why the normal linear regression model (5.1) may be misspecified, some of which are now studied. In the first two cases, misspecification arises because the errors are not homoscedastic independent normally distributed but the mean of Y is correctly specified. Our approach is here more restrictive than the one considered in the semi-parametric literature since it still assumes that the true joint distribution of the errors belongs to some parametric family of distributions. For these cases, we readily obtain from Theorem 3.1 some lower bounds for unbiased estimators of the true coefficient β of the normal linear regression model. On the other hand, in the other two cases, the mean of Y is also incorrectly specified. The pseudo-true parameters β_* are therefore of interest. In this situation, we extend the Gauss-Markov optimality property of the OLS estimator when the errors are and are not normal. Our results complement recent generalizations of the Gauss-Markov Theorem that have appeared in the semi-parametric literature where the mean is still correct (see Hwang (1985), Kariya (1985), and Andrews and Phillips (1985, 1986)).

CASE A: Suppose that the true distribution of Y is given by Equation (5.1) where the ε_i's are independent and identically distributed with some common distribution that is different from the normal distribution. For instance, suppose that the true distribution of ε_i is logistic. Then, $H_0 = G_{1\gamma}$ for some $\gamma \in \Gamma$ where

$$\frac{dG_{1\gamma}}{d\nu} \equiv g_1(y;\gamma) = \prod_{i=1}^{n} \frac{\sqrt{3}}{\pi s} \frac{\exp[-\pi(y_i - z_i'b)/(s\sqrt{3})]}{(1+\exp[-\pi(y_i - z_i'b)/(s\sqrt{3})])^2} \tag{5.7}$$

$\gamma = (b', s)'$, and $\Gamma_1 = \mathbf{B}^0 \times (\sqrt{a}, \sqrt{c})$.[16]

Since H_0 is known to belong to the parametric model $\mathbf{G}_{1\gamma} = \{G_{1\gamma}; \gamma \in \Gamma_1\}$, we can use Theorem 3.1 to obtain a lower bound for the variance of any unbiased estimator $T(Y)$ under $\mathbf{G}_{1\gamma}$ of the pseudo-true parameters $\theta_*(G)$ associated with the closest distribution in $\mathbf{F}_\theta$ to G.[17] As a matter of fact, for this simple case, it is easier to use Equation (3.10) since one can explicitly obtain the function $\theta_*(G_\gamma)$ and hence the derivatives $\partial\theta_*/\partial\gamma'$. For, from Lemma 5.1 and footnote 16, it follows that

$$(\beta_*'(G_{1\gamma}), \sigma_*^2(G_{1\gamma})) = (b', s^2) \tag{5.8}$$

where $\gamma = (b', s^2)$. Therefore, letting $\phi(\cdot)$ be the identity mapping, we obtain for every $\gamma \in \Gamma_1$:

$$Var_{G_{1\gamma}} T(Y) \geq [B_{G_{1\gamma}}^{g_1}(\gamma)]^{-1}, \tag{5.9}$$

where $B_{G_{1\gamma}}^{g_1}(\gamma)$ is non-singular, and

$$B_{G_{1\gamma}}^{g_1}(\gamma) \equiv \int \frac{\partial \log g_1(y;\gamma)}{\partial\gamma} \frac{\partial \log g_1(y;\gamma)}{\partial\gamma'} g_1(y;\gamma)\, d\nu(y)$$

$$= -\int \frac{\partial^2 \log g_1(y;\gamma)}{\partial\gamma\partial\gamma'} g_1(y;\gamma)\, d\nu(y), \tag{5.10}$$

as can readily be checked. Let us note that the lower bound (5.9) is not equal to the lower bound that would be obtained if the normal linear regression model (5.1) is correctly specified.

CASE B: Suppose that the true distribution of Y is still given by Equation (5.1) but that the error ε are jointly normally distributed with zero means and covariance matrix V which is not necessarily of the form $\sigma^2 I$. This occurs if the errors are no longer homoscedastic or independent.

Thus, let $\mathbf{G}_{2\gamma} = \{G_{2\gamma}\,;\gamma \in \Gamma_2\}$ where

$$\frac{dG_{2\gamma}}{d\nu} \equiv g_2(y\,;\gamma) = (2\pi)^{-n/2}\,|V|^{-1/2}\exp\{-\frac{1}{2}(y-Zb)'\,V^{-1}(y-Zb)\}, \tag{5.11}$$

$\gamma = (b', (vech\ V)')'$, $\gamma \in \mathbf{B}^0 \times \mathbf{V}$, and $\mathbf{V}$ is the set of positive definite matrices of which the mean of the diagonal elements belongs to (a, c).[18] Let us note that the number of parameters of the model $\mathbf{G}_\gamma$ is equal to $k_0 + n(n+1)/2$ where n is the sample size. Hence the number of parameters in $\mathbf{G}_\gamma$ is increasing with the sample size. The results of Section 3 nonetheless apply.

Since $E_{G_{2\gamma}}(Y) = Zb$, it follows from Lemma 5.1 that the pseudo-true parameters are for every $\gamma \in \Gamma_2$:

$$(\beta_*'(G_{2\gamma}), \sigma_*^2(G_{2\gamma})) = (b', \frac{1}{n}\sum_{i=1}^{n} v_{ii}) \in \mathbf{B}^0 \times (a, c) \tag{5.12}$$

where v_{ii} is the i-th diagonal element of V. It follows that:

$$\frac{\partial \theta_*(G_{2\gamma})}{\partial \gamma'} = \begin{bmatrix} I & 0 \\ 0 & e'J/n \end{bmatrix} \tag{5.13}$$

where I is the $(k_0 \times k_0)$ identity matrix, e is the $(N \times 1)$ vector of ones, and J is the $n \times n(n+1)/2$ matrix such that $(v_{11}, \ldots, v_{nn})' = J\ vech\ V$. On the other hand, it follows from Richard (1975) that for every $\gamma \in \Gamma_2$:

$$B_{G_{2\gamma}}^{g_2}(\gamma) = \int \frac{\partial^2 \log g_2(y\,;\gamma)}{\partial\gamma\partial\gamma'}\, g_2(y\,;\gamma)\, d\nu(y)$$

$$= \begin{bmatrix} Z'V^{-1}Z & 0 \\ 0 & \frac{1}{2}R(V^{-1}\otimes V^{-1})R' \end{bmatrix} \tag{5.14}$$

as can be verified, where R is the $n(n+1)/2 \times n^2$ matrix such that $vech\ V = R\ vec\ V$. Let us note that $B_{G_{2\gamma}}^{g_2}(\gamma)$ is non-singular even though the sample size n is strictly less than the number of parameters of $\mathbf{G}_{2\gamma}$.

From Theorem 3.1, it follows that if $T(Y)$ is an unbiased estimator of the pseudo-true parameters $\theta_*(G_{2\gamma}) = (b', \frac{1}{n}\sum_{i=1}^{n} v_{ii})'$ under $\mathbf{G}_{2\gamma}$, then for every $\gamma \in \Gamma_2$:

$$Var_{G_{2\gamma}} T(Y) \geq \begin{bmatrix} (Z'V^{-1}Z)^{-1} & 0 \\ 0 & \frac{2}{n^2} e'J[R(V^{-1} \otimes V^{-1})R']^{-1} J'e \end{bmatrix}. \tag{5.15}$$

When interest centers on the pseudo-true parameters $\beta_*(G_{2\gamma})$, which are equal to the true parameters b by Equation (5.12), it follows that a lower bound is $(Z'V^{-1}Z)^{-1}$. Let us note that the GLS estimator $\tilde{\beta} = (Z'V^{-1}Z)^{-1}Z'V^{-1}Y$ is not an estimator since V is unknown, though its variance is equal to the lower bound. One can think using instead a feasible GLS estimator where V is consistently estimated. Consistent estimation of V is however difficult in this case since V is unrestricted and hence characterized by $n(n+1)/2$ parameters where n is the sample size. On the other hand, the OLS estimator $\hat{\beta} = (Z'Z)^{-1}Z'Y$ is always unbiased for $\beta_*(G_{2\gamma}) = b$, but not necessarily optimal since we must have:

$$Var_{G_{2\gamma}} \hat{\beta} = (Z'Z)^{-1}Z'VZ(Z'Z)^{-1} \geq (Z'V^{-1}Z)^{-1}, \tag{5.16}$$

for every $G_{2\gamma} \in \mathbf{G}_{2\gamma}$, i.e., for every $(b, V) \in \mathbf{B}^0 \times \mathbf{V}$, as can be directly verified using the Cauchy-Schwarz inequality (Lemma A in Appendix).

CASE C: In the previous case, the normal linear regression model (5.1) was misspecified only because the true covariance matrix of Y (or equivalently ε) was not necessarily of the postulated form $\sigma^2 I$. On the other hand, the mean of Y (the functional form) was specified correctly as being of the linear form $Z\beta$. In many circumstances, this may also be violated. To study this situation, we may want to consider the following flexible model in which we have only retained the normality assumption. Specifically, let $\mathbf{G}_{3\gamma} = \{G_{3\gamma}; \gamma \in \Gamma_3\}$ where

$$\frac{dG_{3\gamma}}{d\nu} \equiv g_3(y;\gamma) = (2\pi)^{-n/2} \, |V|^{-1} \exp\{-\frac{1}{2}(y-\mu)'V^{-1}(y-\mu)\}, \tag{5.17}$$

$\gamma = (\mu', (vech\ V)')'$, $\Gamma_3 = \mathbf{M} \times \mathbf{V}$, $\mathbf{M}$ is the set of vectors of $\mathbb{R}^n$ of which the mean of the components belong to $\mathbf{B}^0$, and $\mathbf{V}$ is defined as before. Let us note that, for every $G_{3\gamma}$ in $\mathbf{G}_{3\gamma}$, we have:

$$E_{G_{3\gamma}}(Y) = \mu, \; Var_{G_{3\gamma}}(Y) = V, \tag{5.18}$$

where the true mean μ is now unrestricted.

It is easy to see that the assumptions of Lemma 5.1 are satisfied. Thus it follows that the pseudo-true parameters for every $\gamma \in \Gamma_3$ are:

$$\beta_*(G_{3\gamma}) = (Z'Z)^{-1}Z'\mu, \tag{5.19}$$

$$\sigma_*^2(G_{3\gamma}) = \frac{1}{n}\left(\sum_{i=1}^{n} v_{ii} + \mu' M_Z \mu\right), \tag{5.20}$$

where $\gamma = (\mu', (vech\ V)')'$. Therefore the closest distribution in the specified normal linear model F_θ to the true distribution when this latter one belongs to $G_{3\gamma}$ is the multivariate normal distribution $N(Z\beta_*, \sigma_*^2 I)$, where $\beta_* = \beta_*(G_{3\gamma})$, $\sigma_*^2 = \sigma_*^2(G_{3\gamma})$. We note that the "approximate" mean $Z\beta_*$ is the orthogonal projection of the true mean on the column space of Z (see also White (1980a)).

From Theorem 3.1, we can readily obtain a lower bound for the covariance matrix of any estimator $T(Y)$ of the pseudo-true parameters θ_* that is unbiased under $G_{3\gamma}$. For, letting Z be the identity matrix in Equation (5.14), we have for every $\gamma \in \Gamma_3$:

$$B_{G_{3\gamma}}^{g_3}(\gamma) = \begin{bmatrix} V^{-1} & 0 \\ 0 & \frac{1}{2} R(V^{-1} \otimes V^{-1})R' \end{bmatrix}, \tag{5.21}$$

which is non-singular. Since

$$\frac{\partial \theta_*(G_{3\gamma})}{\partial \gamma'} = \begin{bmatrix} (Z'Z)^{-1}Z' & ; & 0 \\ 2\mu' M_Z / n & ; & e'J/n \end{bmatrix}, \tag{5.22}$$

we obtain for every $\gamma = (\mu', (vech\ V)')' \in \Gamma_3$:

$$Var_{G_{3\gamma}} T(Y)) \geq \begin{bmatrix} (Z'Z)^{-1} Z'VZ(Z'Z)^{-1} & ; & \frac{2}{n}(Z'Z)^{-1}Z'VM_Z\mu \\ \frac{2}{n}\mu' M_Z VZ(Z'Z)^{-1} & ; & \frac{4}{n^2}\mu' M_Z VM_Z\mu + \frac{2}{n^2}e'J(R(V^{-1} \otimes V^{-1})R')^{-1}J'e \end{bmatrix} \tag{5.23}$$

One may ask if this lower bound is attained. A positive answer is given by the OLS estimator $\hat\beta = (Z'Z)^{-1}Z'Y$. Indeed,

$$E_G(\hat\beta) = (Z'Z)^{-1}Z'\mu = \beta_*(G), \; Var_G(\hat\beta) = (Z'Z)^{-1}Z'VZ(Z'Z)^{-1} \tag{5.24}$$

for every $G \in G_{3\gamma}$. Then, from Equations (5.23) and (5.24), we obtain the following optimal property: *The OLS estimator is a uniformly minimum variance unbiased estimator* (UMVUE) *under* $G_{2\gamma}$ *of the pseudo-true parameters* β_*. This result extends (to the case where the normal linear regression model is misspecified but the errors ε are still normal) the well-known result that the OLS estimator is best unbiased (BUE), i.e., has minimum variance in the entire class of unbiased estimators of the true parameters $b = \beta_*$ when the normal linear regression model is correctly specified (see, e.g., Rao (1973, p.319)). It is also worthnoting that $Z\hat\beta$ is an UMVUE of $Z\beta_*$ under $G_{3\gamma}$. As noticed earlier $Z\beta_*$ is the closest vector in the column space of Z to the true mean of Y.

CASE D: We shall now relax the last assumption, which is the normality of Y or equivalently ε. Specifically, we consider the (non-parametric) model $G(F_\theta)$ which is the class of all cdf's for Y with respect to which the normal linear regression model F_θ is regular. Then, if $\mu \equiv E_G(Y)$ and $V \equiv Var_G(Y)$, then it follows from Lemma 5.1 that the pseudo-true parameters are given by

$$\beta_*(G) = (Z'Z)^{-1}Z\mu, \tag{5.25}$$

$$\sigma_*^2(G) = \frac{1}{n}(\sum_{i=1}^{n} v_{ii} + \mu' M_Z \mu), \tag{5.26}$$

for every G as in Equations (5.19) - (5.20).

From Theorem 4.1, we obtain a lower bound for unbiased estimators under $\mathbf{G(F_\theta)}$ of the pseudo-true parameters $\beta_*(G)$.

LEMMA 5.2: Let $\mathbf{F_\theta}$ be the normal linear regression model (5.1) with $\Theta = \mathbf{B} \times [a, b]$. Let $\mathbf{G(F)}$ be the class of all cdf's with respect to which $\mathbf{F_\theta}$ is regular. If $T(Y)$ is an unbiased estimator of $\beta_*(G)$ with finite variance under every $G \in \mathbf{G(F_\theta)}$, then:

$$Var_G T(Y) \geq LB_G = (Z'Z)^{-1} Z'VZ(Z'Z)^{-1}.\text{[19]} \tag{5.27}$$

From Equation (5.24), we know that the OLS estimator $\hat{\beta}$ is an unbiased estimator of the pseudo-true parameters $\beta_*(G)$ under every G, and that its covariance matrix is equal to the lower bound (5.27). Hence, we obtain the following general optimal property when the mean and variances of Y may be incorrectly specified and the errors are not necessarily normal: *The OLS estimator is an* UMVUE *under* $\mathbf{G(F_\theta)}$ *of the pseudo-true parameters* β_*. This result extends, to the general misspecified case, the celebrated Gauss-Markov Theorem, which states that the OLS estimator is BLUE, i.e., has minimum variance in the class of *linear* unbiased estimator of the true parameters $b = \beta_*$ when the *mean* of the linear regression model $\mathbf{F_\theta}$ is correctly specified (see, e.g., Rao (1973, p.223)).

An important special case of the normal linear regression model is one in which there are no explanatory variables so that $Z = e$ (the vector on ones). The model (5.1) becomes:

$$Y_i = \beta + \varepsilon_i, i = 1, \ldots, n, \tag{5.28}$$

where the ε_i's are iid $N(0,\sigma^2)$. The pseudo-true parameters are:

$$\beta_*(G) = \frac{1}{n}\sum_{i=1}^{n} \mu_i, \; \sigma_*^2(G) = \frac{1}{n}\sum_{i=1}^{n} v_{ii} \tag{5.29}$$

where G is any cdf for Y with respect to which the model (5.28) is regular, and $\mu_i \equiv E_G(Y_i) < \infty$, $v_{ii} = Var_G(Y_i) < \infty$ (Lemma 5.1). Then, we obtain the following optimal property of the sample mean: *The sample mean* $\bar{Y}_n \equiv \frac{1}{n}\sum_{i=1}^{n} Y_i$ *is an UMVUE of the mean of the true means*, i.e., of $\frac{1}{n}\sum_{i=1}^{n}\mu_i$, *under the class of all joint cdf's for Y with respect to which the model (5.28) is regular.*

6. CONCLUSION

In this paper, we have derived some lower bounds of the Cramer-Rao type for the covariance matrix of any unbiased estimator of the pseudo-true parameters in a parametric model that may be misspecified. Specifically, we have obtained some lower bounds when the true distribution generating the observations belongs either to a parametric model which may differ from the specified parametric model or to the class of all distributions with respect to which the model is regular. These two extreme situations contrast with those considered in the semi-parametric literature where the true distribution is unrestricted except for the true mean being of a known parametric form and for some additional non-parametric restrictions on higher moments such as symmetry.

As an illustration, we have applied our results to the homoscedastic normal linear regression model with fixed regressors. In particular, we have generalized the widely known Gauss-Markov Theorem, and established that the OLS estimator is best in the sense of minimizing the covariance matrix of any unbiased estimator of the pseudo-true parameters when misspecification of the model arises from misspecification of the functional form and non-homoscedasticity and non-normality of the errors.

Some final remarks are in order. First, it is clear that our results apply to parametric models that are much more complicated than the normal linear regression model. For instance, one can obtain some similar lower bounds for unbiased estimators of the pseudo-true parameters in a normal linear regression model with lagged dependent variables and correlated errors.[20] Other interesting examples are limited-dependent variables models and simultaneous equations models.

Second, our lower bounds can be used to evaluate the efficiency of an unbiased estimator of the pseudo-true parameters. For instance, in the normal linear regression model, the efficiency of the OLS estimator was established by showing that its covariance matrix is equal to our lower bounds. As for the usual Cramer-Rao bound obtained under correct specification, our bounds may not be sharp in the sense that, for some other models, there may not exist an unbiased estimator of the pseudo-true parameters of which the covariance matrix is equal to these bounds. Following Bhattacharya (1946), one can however obtain some new and sharper lower bounds in those situations.

Finally, one may use our lower bounds to evaluate the asymptotic efficiency of consistent and asymptotically normal estimators of the pseudo-true parameters. Our general lower bound obtained in Section 4 suggests that the quasi-maximum likelihood estimator of the pseudo-true parameters is asymptotically efficient since its asymptotic covariance matrix is equal to that bound (see, e.g., White (1982)). As in the correct specification case, however, careful definitions of asymptotic efficiency must be proposed so as to avoid super efficiency (see, e.g., LeCam (1953), Rao (1963)). For instance, the concept of uniform convergence must be appropriately defined in the misspecification context. These are clearly topics for further research.

APPENDIX

Throughout, the norm of a real vector is the euclidean norm, and the norm of a matrix is the one defined in, e.g., Rudin (1976, pp. 208-211). This is denoted by $|A|$. We shall also use the following generalization of the Cauchy-Schwarz inequality.

LEMMA A: Let U and V be two real random vectors on $\mathbb{R}^m$ and $\mathbb{R}^n$ with finite covariance matrix $Var\ U$ and $Var\ V$. Let $Cov(U,V)$ be the covariance matrix of U and V, and $Cov(V,U) = Cov(U,V)'$. Then, $Cov(U,V)$ is finite, and

$$Var\ U \geq Cov(U,V)\,[Var\ V]^{-}\,Cov(V,U) \tag{A.1}$$

where $[Var\ V]^{-}$ is any symmetric reflexive g-inverse of $Var\ V$. Moreover, the right-hand side is independent of any choice of symmetric reflexive g-inverses.

PROOF OF LEMMA A: By the usual Cauchy-Schwarz inequality, $Cov(U,V)$ is finite. Let Λ by any $m \times n$ real matrix, and consider the covariance matrix of $U + \Lambda V$:

$$Var(U + \Lambda V) = Var\ U + \Lambda\ Var\ V\ \Lambda' + \Lambda\ Cov(V,U) + Cov(U,V)\,\Lambda'. \tag{A.2}$$

Choose $\Lambda_* = Cov(U,V)\,[Var\ V]^{-}$ where $[Var\ V]^{-}$ is any symmetric reflexive g-inverse of $Var\ V$. Since $([Var\ V]^{-})' = [Var\ V]^{-}$ and $[Var\ V]^{-}\,[Var\ V]\,Var\ V]^{-} = [Var\ V]^{-}$ by definition, we have:

$$Var(U + \Lambda_* V) = Var\ U - Cov(U,V)\,[Var\ V]^{-}\,Cov(V,U), \tag{A.3}$$

which establishes the desired inequality.

To prove the second part, we note that since $Var\ V$ is psd, there exists an $n \times n$ matrix P such that $PP' = P'P = I$ and $P'(Var\ V)P = D$ where

$$D = \begin{bmatrix} D_r & 0 \\ 0 & 0 \end{bmatrix} \tag{A.4}$$

and D_r is a positive diagonal matrix of dimension $r = \text{rank}\ Var\ V$. Let $W \equiv P'V$. Then $Var\ W = D$ so that the last $n - r$ components of the random vector W are (almost surely) equal to zero. Let W_r be the first r components of W. Then

$$Cov(U,W) = (Cov(U,W_r),0),\ \ Cov(U,V) = (Cov(U,W_r),0)P', \tag{A.5}$$

since $V' = W'P' = (W_r',0)P'$. From Vuong (1986b, Lemma 3), any symmetric reflexive g-inverse of $Var\ V$ is of the form $[Var\ V]^{-} = PHP'$ where the top-left block of H is equal to $D_r^{-1} = [Var\ W_r]^{-1}$. Hence

$$Cov(U,V)\,[Var\ V]^{-}\,Cov(V,U) = Cov(U,W_r)\,[Var\ W_r]^{-1}\,Cov(W_r,U). \tag{A.6}$$

This establishes the second part.

PROOF OF LEMMA 2.1: Part (i) is straightforward from Assumptions A1 - A3 (see, e.g., White (1982)). That $\theta_*(G) = \theta_0$ if $G = F_{\theta_0}$ follows from Jensen inequality and Assumption A3 - (a). The proof of Equation (2.6) is slightly different since Assumption A4 is relatively unusual. We shall prove that Assumptions A2 - $A4_*$ imply the more usual assumption that $\int f(y;\theta)\, d\nu(y)$ can be differentiated twice under the integral in a neighborhood of θ_0 so that in that neighborhood:

$$\int \frac{\partial^2 f(y;\theta)}{\partial\theta\partial\theta'}\, d\nu(y) = 0 \tag{A.7}$$

(see, e.g., Silvey (1959, Assumption 13)).

From Assumption $A4_*$, for every $\theta \in N_*$,

$$\left| \frac{\partial f(y;\theta)}{\partial\theta} \right| \le f(y;\theta_*(G))M_*(y), \ \forall\, y \in \mathbf{Y}, \tag{A.8}$$

where $M_*(\cdot)$ is (square) integrable with respect to G. Since $G = F_{\theta_*}$, it follows that $\int M_*(y) f(Y;\theta_0)\, d\nu(y) < \infty$. Hence the left-hand side of (A.8) is dominated by a ν-integrable function independent of θ because $\theta_*(G) = \theta_0$. Thus by the Lebesgue Dominated Convergence (LDC) Theorem (see, e.g., Rudin (1976, p.321)), we can differentiate $\int f(y;\theta)\, d\nu(y) = 1$ on N_*:

$$\int \frac{\partial f(y;\theta)}{\partial\theta}\, d\nu(y) = 0, \ \forall\, \theta \in N_*. \tag{A.9}$$

Then we note the identity for every $(y,\theta) \in \mathbf{Y} \times \boldsymbol{\Theta}$:

$$\frac{\partial^2 f(y;\theta)}{\partial\theta\partial\theta'} = \frac{\partial^2 \log f(y;\theta)}{\partial\theta\partial\theta'} f(y;\theta) + \frac{\partial \log f(y;\theta)}{\partial\theta} \frac{\partial \log f(y;\theta)}{\partial\theta'} f(y;\theta). \tag{A.10}$$

But, from Assumptions A2 and $A4_*$, for every $(y;\theta) \in \mathbf{Y} \times N_*$:

$$\left| \frac{\partial^2 \log f(y;\theta)}{\partial\theta\partial\theta'} \right| f(y;\theta) \le \left| \frac{\partial^2 \log f(y;\theta)}{\partial\theta\partial\theta'} \right| \left| f(y;\theta_*) + \frac{\partial f(y;\bar{\theta}_y)}{\partial\theta'} (\theta - \theta_*) \right|$$

$$\le M(y) f(y;\theta_*) + M(y) M_*(y) f(y;\theta_*) \mid \theta - \theta_* \mid \tag{A.11}$$

where $\theta_* = \theta_*(G)$, $\bar{\theta}_y \in (\theta, \theta_*)$, and $M(\cdot)$ and $M_*(\cdot)$ are square-integrable with respect to $G = F_{\theta_*} = F_{\theta}$. Hence, the right-hand side of (A.11) is ν-integrable. For θ sufficiently close to θ_*, $\mid \theta - \theta_* \mid$ is bounded above. Hence the left-hand side of (A.11) is dominated in a neighborhood of $\theta_* = \theta_0$ by a ν-integrable function independent of θ. On the other hand from (A.8) and Assumption A2, for every $\theta \in N_*$:

$$\left| \frac{\partial \log f(y;\theta)}{\partial \theta} \frac{\partial \log f(y;\theta)}{\partial \theta} \right| f(y;\theta) \le \left| \frac{\partial \log f(y;\theta)}{\partial \theta} \right| \left| \frac{\partial f(y;\theta)}{\partial \theta} \right|$$

$$\le M(Y) M_*(y) f(y;\theta_*), \tag{A.12}$$

which is also ν-integrable. Using now (A.10) - (A.12), it follows that in a neighborhood of $\theta_* = \theta_0$, $| \partial^2 f(y;\theta)/\partial\theta\partial\theta' |$ is dominated by a ν-integrable function independent of θ, Hence, by the LDC Theorem, we can differentiate (A.9) under the integral sign in a neighborhood of $\theta_* = \theta_0$ so as to get (A.7). Equation (2.6) follows by letting $\theta = \theta_0$ in (A.10) and integrating it with respect to ν.

PROOF OF LEMMA 2.2: To prove Equation (2.7), it suffices to show that we can differentiate $\int g(y;\gamma)\, d\nu(y) = 1$ under the integral. This is done as in Lemma 2.1 since for every $\gamma \in \Gamma$ and every $\tilde{\gamma} \in N_{\gamma}$:

$$\left| \frac{\partial g(y;\tilde{\gamma})}{\partial \gamma} \right| \le M_{\gamma}(y) g(y;\gamma), \tag{A.13}$$

(Assumption B2.) To prove the second part, we note that for every $\gamma \in \Gamma$,

$$\left| \frac{\partial \log g(y;\gamma)}{\partial \gamma} \frac{\partial \log g(y;\gamma)}{\partial \gamma'} \right| g(y;\gamma) \le \left| \frac{1}{g(y;\gamma)} \frac{\partial g(y;\gamma)}{\partial \gamma} \right|^2 g(y;\gamma),$$

$$\le M_{\gamma}^2(\gamma) g(y;\gamma), \tag{A.14}$$

where the right-hand side is ν-integrable. Hence $B_{\gamma}^{g}(\gamma)$ is finite for every $\gamma \in \Gamma$.

PROOF OF LEMMA 3.1: (i) From Lemma 2.1, it follows that for every $\gamma \in \Gamma$, and $\theta \in \Theta$,

$$\frac{\partial z^f(\theta,\gamma)}{\partial \theta} = \int \frac{\partial \log f(y;\theta)}{\partial \theta}\, dG_{\gamma}(y) < \infty, \tag{A.15}$$

$$\frac{\partial^2 z^f(\theta,\gamma)}{\partial \theta \partial \theta'} = \int \frac{\partial^2 \log f(y;\theta)}{\partial \theta \partial \theta'}\, dG_{\gamma}(y) = A^f_{G_\gamma}(\theta) < \infty. \tag{A.16}$$

To prove that the second partial derivatives are continuous in both θ and γ, we note from Assumptions A2 and B2 that for every $(\tilde{\theta}, \tilde{\gamma}) \in \Theta \times N_{\gamma}$, we have for $\bar{\gamma}_y \in (\tilde{\gamma}, \gamma)$:

$$\left| \frac{\partial^2 \log f(y;\tilde{\theta})}{\partial \theta \partial \theta'} \right| g(y;\tilde{\gamma}) \le \left| \frac{\partial^2 \log f(y;\tilde{\theta})}{\partial \theta \partial \theta'} \right| \left| g(y;\gamma) + \frac{\partial g(y;\bar{\gamma}_y)}{\partial \gamma'} (\tilde{\gamma} - \gamma) \right|,$$

$$\le M(y) g(y;\gamma) + M(y) M_{\gamma}(y) g(y,\gamma)\, | \tilde{\gamma} - \gamma | \tag{A.17}$$

where $M(\cdot)$ and $M_\gamma(\cdot)$ are square-integrable with respect to G_γ. Hence, for $\tilde{\gamma}$ sufficiently close to γ, the left-hand side of (A.17) is dominated by a ν-integrable function independent of $(\tilde{\theta}, \tilde{\gamma})$. From the LDC Theorem and (A.16), it follows that $\partial^2 z^f(\theta,\gamma)/\partial\theta\partial\theta'$ is continuous on $\Theta \times \Gamma$.

To prove that $\partial^2 z^f(\theta,\gamma)/\partial\theta\partial\gamma'$ exists and is continuous on $\Theta \times \Gamma$, we note from Assumptions A2 and B2 that for every $(\tilde{\theta}, \tilde{\gamma}) \in \Theta \times N_\gamma$:

$$\left| \frac{\partial \log f(y;\tilde{\theta})}{\partial\theta} \frac{\partial g(y;\tilde{\gamma})}{\partial\gamma'} \right| \leq M(y) M_\gamma(y) g(y;\gamma), \tag{A.18}$$

where $M(\cdot)$ and $M_\gamma(\cdot)$ are square-integrable with respect to G_γ. Hence, the left-hand side of (A.18) is in a neighborhood of γ dominated by a ν-integrable function independent of $(\tilde{\theta}, \tilde{\gamma})$. By the LDC Theorem, we can differentiate (A.15) under the integral sign with respect to γ so that for every $(\theta, \gamma) \in \Theta \times \Gamma$:

$$\frac{\partial^2 z^f(\theta,\gamma)}{\partial\theta\partial\gamma'} = \int \frac{\partial \log f(y;\theta)}{\partial\theta} \frac{\partial g(y;\gamma)}{\partial\gamma'} d\nu(y) < \infty, \tag{A.19}$$

which is equal to $B^{fg}_{G_\gamma}(\theta,\gamma)$. Moreover, using again the LDC Theorem, the domination condition (A.18) implies that $\partial^2 z^f(\theta,\gamma)/\partial\theta\partial\theta'$ is continuous in both θ and γ.

(ii) Since for every $\gamma \in \Gamma$, $\theta_*(G_\gamma)$ maximizes the function $z^f(\theta,\gamma) \equiv z^f_{G_\gamma}(\theta)$, and since $\theta_*(G_\gamma) \in \Theta^0$ (Assumption A3), it follows that we must have,

$$\frac{\partial z^f(\theta_*(G_\gamma),\gamma)}{\partial\gamma} = 0, \ \forall \gamma \in \Gamma. \tag{A.20}$$

Since $\theta_*(G_\gamma)$ is unique, (A.20) defines $\theta_* \equiv \theta_*(G_\gamma)$ as a function of γ over Γ. From part (i), the function $\partial z^f(\theta,\gamma)/\partial\theta$ is continuously differentiable on $\Theta \times \Gamma$. Hence, by the Implicit Function Theorem (see, e.g., Rudin (1976, p.224)), θ_* as a function of γ is continuously differentiable on Γ if $\partial^2 z^f(\theta_*,\gamma)/\partial\theta\partial\theta'$ is non-singular. But $\partial^2 z^f(\theta_*,\gamma)/\partial\theta\partial\theta' = A^f_{G_\gamma}(\theta_*)$ (see Equation (3.5)) which is non-singular by Assumption A3 - (b). Thus, by differentiating (A.20) with respect to γ, we obtain Equation (3.7) using Equations (3.5) and (3.6).

PROOF OF THEOREM 3.1: Since the model $\mathbf{G}_\gamma$ is parametric, the unbiasedness condition (3.1) becomes:

$$\int T(y) g(y;\gamma)\, d\nu(y) = \phi(\theta_*(G_\gamma)), \ \forall \gamma \in \Gamma. \tag{A.21}$$

By Assumption B2, for $\gamma \in \Gamma$ and $\tilde{\gamma} \in N_\gamma$:

$$\left| T(y) \frac{\partial g(y;\tilde{\gamma})}{\partial\gamma} \right| \leq | T(y) | M_\gamma(y) g(y;\gamma), \ \forall \gamma \in \Gamma, \tag{A.22}$$

where $M_{\gamma}(\cdot)$ is square-integrable with respect to G_{γ}. On the other hand $T(\cdot)$ is square-integrable with respect to G_{γ} since $Var_{G_{\gamma}} T(Y) < \infty$. Hence, the left-hand side of (A.22) is dominated by a function independent of $\tilde{\gamma}$ that is ν-integrable. From the LDC Theorem. we can differentiate (A.21) under the integral signs and get:

$$\int T(y) \frac{\partial \log g(y;\gamma)}{\partial \gamma'} dG_{\gamma}(y) = \frac{\partial \phi(\theta_*)}{\partial \theta'} \frac{\partial \theta_*}{\partial \gamma'}, \ \forall \gamma \in \Gamma, \tag{A.23}$$

where $\theta_* = \theta_*(G_{\gamma})$. Using Equation (2.7), this is equivalent to:

$$Cov_{G_{\gamma}} \left[T(Y), \frac{\partial \log g(Y;\gamma)}{\partial \gamma} \right] = \frac{\partial \phi(\theta_*)}{\partial \theta'} \frac{\partial \theta_*}{\partial \gamma'}, \ \forall \gamma \in \Gamma. \tag{A.24}$$

In addition, from Lemma 2.2,

$$Var_{G_{\gamma}} \left[\frac{\partial \log g(Y;\gamma)}{\partial \gamma} \right] = B^{g}_{G_{\gamma}}(\gamma) < \infty, \ \forall \gamma \in \Gamma. \tag{A.25}$$

Hence, by Lemma A, $\forall \gamma \in \Gamma$:

$$Var_{G_{\gamma}} T(Y) \geq \frac{\partial \phi(\theta_*)}{\partial \theta'} \frac{\partial \theta_*}{\partial \gamma'} [B^{g}_{G_{\gamma}}(\gamma)]^{-} \frac{\partial \theta_*'}{\partial \gamma} \frac{\partial \phi'(\theta_*)}{\partial \theta}, \tag{A.26}$$

where $[B^{g}_{G_{\gamma}}(\gamma)]^{-}$ is any symmetric reflexive g-inverse of $B^{g}_{G_{\gamma}}(\gamma)$. The desired result follows from Equation (3.7).

PROOF OF COROLLARY 3.1: Recalling Definition 2.1, it follows from Jensen inequality and Assumption A3 - (a), that $\theta_*(F_{\theta}) = \theta$ for every $\theta \in \Theta^0$. Since Assumption $A4_*$ holds for every $G = F_{\theta}$, $\theta \in \Theta^0$, it follows that the parametric model $\mathbf{F}^0_{\theta} \equiv \{F_{\theta};\, \theta \in \Theta^0\}$ is semi-regular (see Definition 2.2). It now suffices to apply Theorem 3.1 with $\mathbf{G} = \mathbf{F}^0_{\theta}$, and to note that for every $G \in \mathbf{F}^0_{\theta}$ and $\theta \in \Theta^0$, such that $G = F_{\theta}$:

$$A^{f}_{G}(\theta_*) = A^{f}(\theta), \tag{A.27}$$

$$B^{fg}_{G}(\theta_*, \theta) = B^{g}_{G}(\theta) = B^{f}(\theta), \tag{A.28}$$

using Equations (2.1), (2.8), (3.4), (3.11), and (3.12). Moreover, from the information matrix equivalence (2.6) and Assumption A3 - (b), the matrix $B^{f}(\theta) = B^{f}_{G}(\theta)$ is non-singular. The desired result follows from Equation (3.9).

PROOF OF COROLLARY 3.2: Since $\theta_*(G_\gamma)$ and $\alpha_*(G_\gamma)$ are differentiable on Γ (Lemma 3.1), it follows from Equation (3.14) that

$$\partial\phi(\theta_*) \tag{A.29}$$

where $\theta_* = \theta_*(G_\gamma)$ and $\alpha_* = \alpha_*(G_\gamma)$. The result immediately follows from Equations (3.7) and (3.9).

PROOF OF COROLLARY 3.3: Let $\theta_* \equiv \theta_*(G_\gamma)$ and $\tilde\theta_* \equiv \tilde\theta_*(\tilde G_\beta)$ be the pseudo-true parameters for the model $\mathbf{F}_\theta$ when $G = G_\gamma = \tilde G_\beta$ with $\beta = \lambda(\gamma)$. From the assumptions, we have:

$$\theta_*(G_\gamma) = \tilde\theta_*(\tilde G_{\lambda(\gamma)}),\ \forall\, \gamma \in \Gamma. \tag{A.30}$$

Hence,

$$\frac{\partial\phi(\theta_*)}{\partial\theta'}\,[A_G^f(\theta_*)]^{-1} = \frac{\partial\phi(\tilde\theta_*)}{\partial\theta'}\,[A_G^f(\tilde\theta_*)]^{-1}. \tag{A.31}$$

In view of Equation (3.9), it suffices to show that

$$B_G^{fg}(\theta_*,\gamma)\,[B_G^{g}(\gamma)]^{-}\,B_G^{gf}(\gamma,\theta_*) \le B_G^{f\tilde g}(\tilde\theta_*,\beta)\,[B_G^{\tilde g}(\beta)]^{-}\,B_G^{\tilde g f}(\beta,\tilde\theta_*), \tag{A.32}$$

for any choice of symmetric reflexive g-inverses. We note that

$$\frac{\partial \log g(y;\gamma)}{\partial\gamma} = \frac{\partial\lambda'}{\partial\gamma}\,\frac{\partial \log \tilde g(y;\beta)}{\partial\beta}$$

where $\partial\lambda'/\partial\gamma$ is evaluated at γ, and $\beta = \lambda(\gamma)$. Since $\theta_* = \tilde\theta_*$, (A.32) becomes

$$B_G^{f\tilde g}(\theta_*,\beta)\,\frac{\partial\lambda}{\partial\gamma'}\left[\frac{\partial\lambda'}{\partial\gamma}\,B_G^{\tilde g}(\beta)\,\frac{\partial\lambda}{\partial\gamma}\right]^{-}\frac{\partial\lambda'}{\partial\gamma}\,B_G^{\tilde g f}(\beta,\theta_*) \le B_G^{f\tilde g}(\theta_*,\beta)\,[B_G^{\tilde g}(\beta)]^{-}\,B_G^{\tilde g f}(\beta,\theta_*). \tag{A.33}$$

As in the proof of Lemma A, let P be an orthogonal matrix that diagonalizes $B_G^{\tilde g}(\beta)$ into a diagonal matrix of which the first r diagonal elements are strictly positive with $r = rank\ B_G^{\tilde g}(\beta)$. Define

$$W = P'\,\frac{\partial \log \tilde g(Y;\beta)}{\partial\beta} = \begin{bmatrix} W_r \\ 0 \end{bmatrix}, \tag{A.34}$$

Since $E_G(W) = 0$ when $G = \tilde G_\beta$ (Lemma 2.2), it follows from (A.6) that the right-hand side (RHS) of (A.33) satisfies:

$$RHS = Cov_G\left[\frac{\partial \log f(Y;\theta_*)}{\partial\theta},\, W_r\right]\left[Var_G W_r\right]^{-1} Cov_G\left[W_r,\, \frac{\partial \log f(Y;\theta_*)}{\partial\theta}\right]. \tag{A.35}$$

On the other hand, let P_r be the $b \times r$ matrix of which the columns are the first r columns of P. From (A.34), it follows that the left-hand side (LHS) of (A.33) becomes

$$LHS = Cov_G\left[\frac{\partial \log f(Y;\theta_*)}{\partial\theta}, W_r \right] P_\gamma' \, [P_\gamma (Var_G W_r) P_\gamma']^- P_\gamma Cov_G\left[W_r, \frac{\partial \log f(Y;\theta_*)}{\partial\theta} \right] \qquad (A.36)$$

where $P_\gamma' = P_r' \partial\lambda' / \partial\gamma'$. Thus, from (A.35) - (A.36), the inequality (A.33) holds if

$$P_\gamma' [P_\gamma (Var_G W_r) P_\gamma']^- P_\gamma \le [Var_G W_r]^{-1} \qquad (A.37)$$

for any choice of symmetric reflexive g-inverses, i.e., if

$$Var_G W_r \ge (Var_G W_r) P_\gamma' \, [P_\gamma (Var_g W_r) P_\gamma']^- P_\gamma Var_G W_r. \qquad (A.38)$$

Defining $V = W_r$ and $V = P_\gamma W_r$, the inequality (A.38) follows from Lemma A.

To prove the second part, we note that for $\beta = \lambda(\gamma)$:

$$B_G^g(\gamma) = \frac{\partial\lambda'}{\partial\gamma} B_G^{\tilde g}(\beta) \frac{\partial\lambda}{\partial\gamma'}, \qquad (A.39)$$

so that $p \le b$ if $B_G^g(\gamma)$ is non-singular. Moreover $rank\ \partial\lambda'/\partial\gamma = p$ and $p \le rank\ B_G^{\tilde g}(\beta) \le b$. Hence, if $p = b$, all the matrices in (A.39) are non-singular. It follows that (A.33) holds with equality, and the desired result follows.

PROOF OF LEMMA 4.1: Since $\mathbf{F}_\theta$ satisfies Assumption A1, we need to show that $\mathbf{G}_\theta$ satisfies Assumptions B1 - B2, and that $\mathbf{F}_\theta$ satisfies Assumptions A2 - A4 for every cdf in $\mathbf{G}_\theta$. The proof is done in five steps, one step for each condition.

Step 1: First we show that the parametric model $\mathbf{G}_\theta^g \equiv \{G_\theta ; \theta \in N_*\}$ satisfies Assumption B1, where N_* is the neighborhood of θ_* of Assumption A4. In view of Equation (4.1), it suffices to show that $C(\theta)$ exists and is continuously differentiable on N_*. Since $f(Y;\theta) > 0$ for every $(y, \theta) \in \mathbf{Y} \times \boldsymbol{\Theta}$, we have:

$$1 < 1 + \exp\left[1 - \frac{f(y;\theta)}{f(y;\theta_*)} \right] \le 1 + e, \ \forall y \in \mathbf{Y}, \ \forall \theta \in \boldsymbol{\Theta}. \qquad (A.40)$$

Integrating (A.40) with respect to G, we obtain

$$1 < C(\theta) \le 1 + e, \ \forall \theta \in \boldsymbol{\Theta}. \qquad (A.41)$$

So $C(\theta) < \infty$. Moreover, from Assumption $A4_*$, we have for every $\theta \in N_*$:

$$\frac{1}{f(y;\theta_*)} \left| \frac{\partial f(y;\theta)}{\partial\theta} \right| \exp\left[1 - \frac{f(y;\theta)}{f(y;\theta_*)} \right] g(y) \le e\, M_*(y) g(y), \ \forall y \in \mathbf{Y}, \qquad (A.42)$$

where $M_*(\cdot)$ is square and hence simply G-integrable. Thus the derivatives of the integrand of $C(\theta)$ with respect to θ are dominated by a function independent of θ that is ν-integrable. By the LDC Theorem, we obtain that for every $\theta \in N_*$:

$$\frac{\partial C(\theta)}{\partial \theta} = -\int \frac{1}{f(y;\theta_*)} \frac{\partial f(y;\theta)}{\partial \theta} \exp\left[1 - \frac{f(y;\theta)}{f(y;\theta_*)}\right] dG(y) < \infty. \tag{A.43}$$

Moreover, since the integrand of $\partial C(\theta)/\partial\theta$ is continuous in θ and since (A.42) holds for every $\theta \in N_*$, then $\partial C(\theta)/\partial\theta$ is continuous on N_*.

Step 2: A useful result is that if $M(y)$ is a square-integrable function with respect to G, then it is also square-integrable with respect to G_θ for every $\theta \in N_*$. Indeed, from (A.40) we have:

$$0 \le M^2(y) g(y;\theta) \le \frac{1+e}{C(\theta)} M^2(y) g(y), \ \forall\, y \in \mathbf{Y},\ \forall\, \theta \in N_*. \tag{A.44}$$

Hence $\int M^2(y) dG(y) < \infty$ implies that $\int M^2(y) dG_\theta(y) < \infty$ for every $\theta \in N_*$.

From this result, it follows that Assumption A2 is satisfied for every $G_{\tilde\theta}$, $\tilde\theta \in N_*$, since the functions $|\log f(\cdot\,;\theta)|$, $|\partial \log f(\cdot\,;\theta)/\partial\theta|$, and $|\partial^2 \log f(\cdot\,;\theta/\partial\theta\partial\theta'|$ are dominated by a function square-integrable with respect to G and hence with respect to $G_{\tilde\theta}$ also.

Step 3: We shall show that the model $\mathbf{G}_{\theta}$ satisfies Assumption B2. For every $(y,\theta) \in \mathbf{Y} \times N_*$, we have:

$$\frac{\partial \log g(y;\theta)}{\partial\theta} = -\frac{1}{C(\theta)} \frac{\partial C(\theta)}{\partial\theta} - \frac{\exp\left[1 - \frac{f(y;\theta)}{f(y;\theta_*)}\right]}{1 + \exp\left[1 - \frac{f(y;\theta)}{f(y;\theta_*)}\right]} \frac{1}{f(y;\theta_*)} \frac{\partial f(y;\theta)}{\partial\theta}. \tag{A.45}$$

Thus, for every $(y,\theta,\tilde\theta) \in \mathbf{Y} \times N_* \times N_*$:

$$\frac{1}{g(y;\theta)} \frac{\partial g(y;\tilde\theta)}{\partial\theta} = -\frac{C(\theta)}{C(\tilde\theta)} \frac{1 + \exp\left[1 - \frac{f(y;\tilde\theta)}{f(y;\theta_*)}\right]}{1 + \exp\left[1 - \frac{f(y;\theta)}{f(y;\theta_*)}\right]}$$

$$\times \left\{ \frac{1}{C(\tilde\theta)} \frac{\partial C(\tilde\theta)}{\partial\theta} + \frac{\exp\left[1 - \frac{f(y;\tilde\theta)}{f(y;\theta_*)}\right]}{1 + \exp\left[1 - \frac{f(y;\tilde\theta)}{f(y;\theta_*)}\right]} \frac{1}{f(y;\theta_*)} \frac{\partial f(y;\tilde\theta)}{\partial\theta} \right\}$$

Using Assumption $A4_*$ and the inequalities (A.40) - (A.41), we obtain for every $(\theta, \tilde{\theta}) \in N_* \times N_*$:

$$\frac{1}{g(y;\theta)} \left| \frac{\partial g(y;\tilde{\theta})}{\partial \theta} \right| \leq (1+e)^2 \left\{ \left| \frac{\partial C(\tilde{\theta})}{\partial \theta} \right| + eM_*(y) \right\}, \; \forall\, y \in \mathbf{Y}, \tag{A.46}$$

where $M_*(\cdot)$ is square integrable with respect to G and hence with respect to G_θ (see Step 2). Moreover, since $\partial C(\tilde{\theta})/\partial\theta$ is continuous on N_* (see Step 1), for every $\theta \in N_*$, there exists a neighborhood N_θ such that

$$\left| \frac{\partial C(\tilde{\theta})}{\partial \theta} \right| \leq a(\theta), \; \forall\, \tilde{\theta} \in N_\theta, \tag{A.47}$$

where $a(\theta)$ may depend on θ but not on $\tilde{\theta}$. Hence, for every $\theta \in N_*$, and every $\tilde{\theta} \in N_\theta$:

$$\frac{1}{g(y;\theta)} \left| \frac{\partial g(y;\tilde{\theta})}{\partial \theta} \right| \leq (1+e)^2 \{a(\theta) + eM_*(y)\}, \; \forall\, y \in \mathbf{Y}, \tag{A.48}$$

where the right-hand side is independent of $\tilde{\theta}$ and square-integrable with respect to G_θ. Thus Assumption B2 is satisfied. In view of Step 1, the parametric model $\mathbf{G}_{\mathcal{G}}$ is therefore semi-regular as are all the models of the form $\{G_\theta \,;\, \theta \in N_0\}$ where N_0 is an open subset of N_*.

Step 4: We shall show that there exists a neighborhood N_0 of θ_* such that Assumption A3 is satisfied for every cdf G_θ, $\theta \in N_0$. Define

$$z^f(\tilde{\theta},\theta) \equiv \int \log f(y;\tilde{\theta})\, dG_\theta(y), \tag{A.49}$$

which exists and is twice continuously differentiable in its first argument $\tilde{\theta}$ because of Step 2. In addition, a look at the proof of part (i) of Lemma 3.1 reveals that only Assumptions A1, A2, B1, and B2 were used. These assumptions are satisfied here (see Steps 1, 2, 3). Hence, for every $(\tilde{\theta}, \theta) \in \boldsymbol{\Theta} \times N_*$, the derivatives

$$\frac{\partial z^f(\tilde{\theta},\theta)}{\partial \theta} = \int \frac{\partial \log f(y;\tilde{\theta})}{\partial \theta}\, dG_\theta(y) \tag{A.50}$$

are continuously differentiable in both $\tilde{\theta}$ and θ on $\boldsymbol{\Theta} \times N_*$, with

$$\frac{\partial^2 z^f(\tilde{\theta},\theta)}{\partial \tilde{\theta}\, \partial \tilde{\theta}'} = \int \frac{\partial^2 \log f(y;\tilde{\theta})}{\partial\theta\partial\theta'}\, dG_\theta(y) = A^f_{\mathcal{G}_*}(\tilde{\theta}) < \infty, \tag{A.51}$$

$$\frac{\partial^2 z^f(\tilde{\theta},\theta)}{\partial \tilde{\theta}\, \partial \theta} = \int \frac{\partial \log f(y;\tilde{\theta})}{\partial \theta} \frac{\partial \log g(y;\theta)}{\partial \theta'}\, dG_\theta(y) = B^{fg}_{\mathcal{G}_*}(\tilde{\theta},\theta) < \infty. \tag{A.52}$$

Then, we note that

$$g(y;\theta_*) = g(y), \; \forall\, y \in \mathbf{Y}, \tag{A.53}$$

so that

$$z^f(\tilde{\theta},\theta_*) = \int \log f(y;\tilde{\theta})dG(y) = z_G^f(\tilde{\theta}),\ \forall\,\tilde{\theta}\in\Theta.$$

Since $\mathbf{F}_\theta$ is regular with respect to G, it follows that $\tilde{\theta} = \theta_*$ is the unique maximizer of $z(\tilde{\theta},\theta_*)$ over Θ, and being in Θ^0, we have:

$$\frac{\partial z^f(\theta_*,\theta_*)}{\partial\tilde{\theta}} = 0. \tag{A.54}$$

Moreover $\partial^2 z^f(\theta_*,\theta_*)/\,\partial\tilde{\theta}\,\partial\tilde{\theta}' = A_G^f(\theta_*)$ which is negative definite.

Thus, since $\partial^2 z^f(\tilde{\theta},\theta)/\,\partial\tilde{\theta}\,\partial\tilde{\theta}'$ is continuous in $(\tilde{\theta},\theta)$ (see above), there exists a neighborhood of (θ_*,θ_*) of the form $N_1\times N_2$ included in $N_*\times N_*$ over which $\partial^2 z^f(\tilde{\theta},\theta)/\,\partial\tilde{\theta}\,\partial\tilde{\theta}'$ is negative definite. Therefore for every $\theta\in N_2$, the function $z^f(\tilde{\theta},\theta)$ is strictly concave in $\tilde{\theta}\in N_1$.

For any $\theta\in N_2$, let us now consider the equation in $\tilde{\theta}\in N_1$:

$$\frac{\partial z^f(\tilde{\theta},\theta)}{\partial\tilde{\theta}} = 0. \tag{A.55}$$

Since $\partial z^f(\tilde{\theta},\theta)/\,\partial\tilde{\theta}$ is continuously differentiable in both $\tilde{\theta}$ and θ on $N_1\times N_2$, it follows from (A.54) and the Implicit Function Theorem that there exists a neighborhood N_3 of θ_* (included in N_2) and a continuously differentiable function $\psi(\cdot)$ on N_3 with values in N_1 such that

$$\frac{\partial z^f(\psi(\theta),\theta)}{\partial\tilde{\theta}} = 0,\ \forall\,\theta\in N_3. \tag{A.56}$$

Since for every $\theta\in N_3\subset N_2$, the function $z^f(\tilde{\theta},\theta)$ is strictly concave in $\tilde{\theta}\in N_1$, then $\psi(\theta)$ is the unique maximizer of $z^f(\tilde{\theta},\theta)$ over $\tilde{\theta}\in N_1$. Moreover, by construction of $N_1\times N_2$, $\partial^2 z^f(\psi(\theta),\theta)/\,\partial\theta\partial\theta'$ is negative definite and hence non-singular for every $\theta\in N_3$ as required by Assumption A3 - (b).

We now show that there exists a neighborhood N_0 of θ_* (included in N_3) such that for every $\theta\in N_0$, $\psi(\theta)$ is also the unique maximizer of $z^f(\tilde{\theta},\theta)$ over $\tilde{\theta}\in\mathbf{\Theta}$. First, we note that the function $z^f(\tilde{\theta},\theta)$ is continuous in both $\tilde{\theta}$ and θ on $\mathbf{\Theta}\times N_3$. Indeed, from (A.40) - (A.41), we have for every $(\tilde{\theta},\theta)\in\mathbf{\Theta}\times N_0$:

$$|\log f(y;\tilde{\theta})\,g(y;\theta)| \le (1+e)\,|\log f(y;\tilde{\theta})|\,g(y),\ \forall\,y\in\mathbf{Y},$$

$$\le (1+e)M(y)g(y),\ \forall\,y\in\mathbf{Y}, \tag{A.57}$$

using Assumption A2 . Thus the continuity of $z^f(\tilde{\theta},\theta)$ follows from the LDC Theorem. Then, suppose that there does not exist such a neighborhood N_0. This means that

$$\forall\,n,\exists\,\theta_n\in N_3,\exists\,\tilde{\theta}_n\in\mathbf{\Theta}-N_1,\ |\theta_n-\theta_*| < \frac{1}{n},\ z^f(\tilde{\theta}_n,\theta_n)\ge z^f(\psi(\theta_n),\theta_n). \tag{A.58}$$

(Recall that for every $\theta \in N_3$, $\psi(\theta)$ is the unique maximizer of $z^f(\tilde{\theta},\theta)$ over $\tilde{\theta} \in N_1$.) Now, the sequence $\{\theta_n\}$ converges to θ_*. In addition, $\Theta - N_1$ is compact because Θ is compact and N_1 is open. Since $\{\tilde{\theta}_n\} \in \Theta - N_1$, there exists a subsequence $\{\tilde{\theta}_m\}$ converging to $\tilde{\theta}_*$ (say), and $\tilde{\theta}_* \in \Theta - N_1$. Since $z^f(\tilde{\theta},\theta)$ is continuous on $\Theta \times N_3$ and $\psi(\theta)$ is continuous on N_3, we obtain by taking the limit of (A.58) as $m \to \infty$:

$$z^f(\tilde{\theta}_*,\theta_*) \geq z^f(\theta_*,\theta_*). \tag{A.59}$$

Since $\tilde{\theta}_* \in \Theta - N_1$, and θ_* is the unique maximizer of $z^f(\tilde{\theta},\theta_*)$ over $\tilde{\theta} \in \Theta$, we obtain a contradiction. Hence, for some neighborhood N_0 of θ_*, Assumption A3 is satisfied for every cdf G_θ, $\theta \in N_0$.

Step 5: It remains to show that Assumption A4 is satisfied with respect to every G_θ, $\theta \in N_0$. From Step 3, this means that for every $\theta \in N_0$, there exists a neighborhood $N_*(\theta)$ of $\psi(\theta)$ such that for every $(\tilde{\theta},\tilde{\tilde{\theta}}) \in N_*(\theta) \times N_*(\theta)$, the function $[f(\cdot:\tilde{\theta})]^{-1} \mid \partial f(\cdot:\tilde{\tilde{\theta}})/\partial\theta)$ is dominated by a function independent of $\tilde{\tilde{\theta}}$ and G_θ-square-integrable. Recall that $\psi(\cdot)$ has its value in $N_1 \subset N_*$. Hence N_* is also a neighborhood of $\psi(\theta)$ for every $\theta \in N_0$. Let $N_*(\theta) = N_*$. The result follows from Assumption A4 because any function that is G-square-integrable is also G_θ-square-integrable (Step 2).

PROOF OF THEOREM 4.1: To prove the first part, we use Lemma 4.1 and Theorem 3.1. Let $G \in \mathbf{G}(\mathbf{F}_\theta)$. From Lemma 4.1, there exists a neighborhood $N_0(G)$ of $\theta_* \equiv \theta_*(G)$ such that $\mathbf{F}_\theta$ is regular with respect to $\mathbf{G}_\theta(G)$ where $\mathbf{G}_\theta(G) = \{G_{\theta,G}; \theta \in N_0(G)\}$, $dG_{\theta,G}/\partial\nu \equiv g(y;\theta)$, and $g(y;\theta)$ is defined by Equations (4.1) - (4.2). Since $\mathbf{G}_\theta(G) \subset \mathbf{G}(\mathbf{F}_\theta)$, $T(Y)$ is also an unbiased estimator of $\phi(\theta_*(\tilde{G}))$ with finite variance under every $\tilde{G} \in \mathbf{G}_\theta(G)$. Since $\mathbf{G}_\theta(G)$ is semi-regular and $\mathbf{F}_\theta$ is regular with respect to it, Theorem 3.1 applies. Thus, we have that for every $\tilde{G} \in \mathbf{G}(G)$, and every θ such that $G_{\theta,G} = \tilde{G}$:

$$Var_{\tilde{G}}\, T(Y) \geq LB_{\tilde{G}}(\theta). \tag{A.60}$$

But G is an element of $\mathbf{G}_\theta(G)$ by construction of $\mathbf{G}_\theta(G)$, and $G_{\theta_*,G} = G$. Thus, we have

$$Var_G\, T(Y) \geq LB_G(\theta_*). \tag{A.61}$$

Let us now compute $LB_G(\theta_*)$. From (A.43), we note that $\partial C(\theta_*)/\partial\theta = 0$ so that we obtain from (A.45):

$$\frac{\partial \log g(y,\theta_*)}{\partial\theta} = -\frac{1}{2}\, \frac{\partial \log f(y;\theta_*)}{\partial\theta}. \tag{A.62}$$

Therefore, from Equations (2.3), (2.8), and (3.4), we have:

$$B_G^g(\theta_*) = \frac{1}{4}\, B_G^f(\theta_*), \tag{A.63}$$

$$B_G^{fg}(\theta_*,\theta_*) = -\frac{1}{2}\, B_G^f(\theta_*). \tag{A.64}$$

Hence, from Equation (3.9), we obtain:

$$LB_G(\theta_*) = \frac{\partial\phi(\theta_*)}{\partial\theta'}\,[A_G^f(\theta_*)]^{-1}\,B_G^f(\theta_*)\,[A_G^f(\theta_*)]^{-1}\,\frac{\partial\phi'(\theta_*)}{\partial\theta}, \tag{A.65}$$

where we have used the definition of a g-inverse. The desired result follows from (A.61) and (A.65) by putting $LB_G \equiv LB_G(\theta_*)$.

To prove the second part, we note that $\mathbf{S}_G$ must be non-empty since it must contain the semi-regular model $\mathbf{G}_\theta(G)$. Let $\mathbf{G}_\gamma = \{G_\gamma ; \gamma \in \Gamma \subset I\!R^p\}$ be another element of $\mathbf{S}_G$, i.e., another semi-regular model containing G and with respect to which $\mathbf{F}_\theta$ is regular. Let γ belong to Γ such that $G_\gamma = G$. Putting $U = \partial \log f(Y;\theta_*)/\partial\theta$ and $V = \partial \log g(y;\gamma)/\partial\gamma$, it follows from Lemma 2.2 and Lemma A that

$$B_G^f(\theta_*) \ge B_G^{fg}(\theta_*,\gamma)\,[B_G^g(\gamma)]^-\,B_G^{gf}(\gamma,\theta_*) \tag{A.66}$$

for any choice of symmetric reflexive g-inverses. Then, the desired result follows from (A.65) and Equation (3.9).

PROOF OF COROLLARY 4.1: Part (i) follows from part (ii) by reversing the roles of $\mathbf{F}_\theta$ and $\mathbf{F}_\alpha$. To prove part (ii), we use Equation (4.5) and Lemma 4.1. For, let $G \in \mathbf{G}(\mathbf{F}_\theta)$, and $\mathbf{G}_\theta$ be the semi-regular model defined in Lemma 4.1 with respect to which $\mathbf{F}_\theta$ is regular. Since, by assumption, $\mathbf{G}(\mathbf{F}_\theta) \subset \mathbf{G}(\mathbf{F}_\alpha)$ it follows that $\mathbf{G}_\theta \subset \mathbf{G}(\mathbf{F}_\alpha)$. Since $\mathbf{G}_\theta$ is semi-regular and contains G by construction, then $\mathbf{G}_\theta \in \mathbf{S}_G(\mathbf{F}_\alpha)$ where $\mathbf{S}_G(\mathbf{F}_\alpha)$ is the set of all semi-regular models containing G and with respect to which $\mathbf{F}_\alpha$ is regular. Now, from the proof of Theorem 4.1, we have that $LB_G(\mathbf{F}_\theta) = LB_G(\mathbf{G}_\theta, \theta_*)$ where $\theta_* = \theta_*(G)$ and $G_{\theta_*} = G$. On the other hand, from Equation (4.5), $LB_G(\mathbf{F}_\alpha) = \max LB_G(G_\gamma, \gamma)$ where the max is taken over all $G_\gamma \in \mathbf{S}_G(\mathbf{F}_\alpha)$ and γ is such that $G_\gamma = G$. Since $\mathbf{G}_\theta \in \mathbf{S}(F_\alpha)$, the desired result follows.

PROOF OF LEMMA 5.1: To prove the "if" part of (i), let M be such that $|\beta| < M$ for every $\beta \in \mathbf{B}$. Since $\theta \in \mathbf{B} \times [a,b]$ and

$$|y - Z\beta|^2 \le (|y| + M|Z|)^2, \tag{A.67}$$

it is not difficult to show using Equations (5.2) - (5.4) that:

$$|\log f(y;\theta)| \le \frac{n}{2}\log 2\pi b + \frac{1}{2a}(|y| + M|Z|)^2, \tag{A.68}$$

$$\left|\frac{\partial \log f(y;\theta)}{\partial\theta}\right| \le \frac{1}{a}|Z|(|y| + M|Z|) + \frac{n}{2a} + \frac{1}{2a^2}(|y| + M|Z|)^2, \tag{A.69}$$

$$\left|\frac{\partial^2 \log f(y;\theta)}{\partial\theta\partial\theta'}\right| \le \frac{n}{2a^2} + \frac{1}{a^3}(|y| + (a+M)|Z|)^2. \tag{A.70}$$

Since $E_G(Y_i^4) < \infty$, it follows that $|y|^2$ and therefore the right-hand sides of (A.68) - (A.70) are G-square integrable. Thus Assumptions A2 holds. Moreover, it immediately follows from Equation (5.2) that

$$z\ell_G(\theta) = -\frac{n}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}\sigma_i^2 + (\mu - Z\beta)'(\mu - Z\beta)\right], \tag{A.71}$$

where $\sigma_i^2 = Var_G Y_i < \infty$ and $\mu = E_G(Y) < \infty$. Then, it is easy to show that $z\ell_G(\theta)$ attains a unique maximum over Θ at $\theta_*(G) = (\beta_*'(G), \sigma_*^2(G))'$ as defined by Equations (5.5) - (5.6), which belongs to Θ^0 by assumption. Finally, from Equation (5.4), we obtain

$$A\ell_G(\theta_*(G)) = \begin{bmatrix} -\frac{1}{\sigma_*^2} Z'Z \; ; & 0 \\ 0 & ; \; -\frac{n}{2\sigma_*^4} \end{bmatrix}, \tag{A.72}$$

which shows that $A\ell_G(\theta_*(G))$ is nd. Therefore Assumption A3 is satisfied.

To prove the "only if" part of (i), we note that if Assumption A2 holds, then $\log f(y;\theta)$ and its first and second partial derivatives with respect to θ must be G-square integrable for every $\theta \in \Theta$. From Equation (5.3), it follows that $Z'y$ and $|y - Z\beta|^2$ must be G-square integrable. This implies that $|y|^2$ must be G-square integrable so that $E_G(Y_i^4) < \infty$ for every i. It is also clear that Assumption A3 - (a) implies that $\theta_*(G)$ must belong to Θ^0.

To prove (ii), we note that if $Y'Y \equiv |Y|^2$ has a moment generating function, then for every $\beta \in I\!R^{k}$, $|Y - Z\beta|^2$ has also a moment generating function, i.e., there exists $t_1 > 0$, $\forall\, t \in (-t_1, t_1)$, $E_G(e^{t|Y - Z\beta|^2}) < \infty$. Then, we note that

$$\left|\frac{1}{2\sigma^2}|y - Z\beta|^2 - \frac{1}{2\tilde\sigma^2}|y - Z\tilde\beta|^2\right| = \left|\left(\frac{1}{2\sigma^2} - \frac{1}{2\tilde\sigma^2}\right)|y - Z\beta|^2 + \frac{1}{2\tilde\sigma^2}\left(|y - Z\beta|^2 - |y - Z\tilde\beta|^2\right)\right|,$$

$$\le \frac{1}{2\sigma^2\tilde\sigma^2}|\tilde\sigma^2 - \sigma^2|\,|y - Z\beta|^2 + \frac{1}{2\tilde\sigma^2}|Z|^2\,|\beta - \tilde\beta|^2,$$

$$\le \frac{1}{2a^2}|\tilde\sigma^2 - \sigma^2|\,|y - Z\beta|^2 + \frac{M^2}{a}|Z|^2, \tag{A.73}$$

where we have used that $|\beta| < M$ and $\sigma^2 > a$, $\tilde\sigma^2 > a$. Hence:

$$\frac{f(y;\tilde\theta)}{f(y;\theta)} = \left[\frac{\sigma^2}{\tilde\sigma^2}\right]^{n/2} \exp\left\{\frac{1}{2a^2}|y - Z\beta|^2 - \frac{1}{2\tilde\sigma^2}|y - Z\tilde\beta|^2\right\},$$

$$\le K\exp\left\{\frac{1}{2a^2}|\tilde\sigma^2 - \sigma^2|\,|y - Z\beta|^2\right\}, \tag{A.74}$$

where $K = (b/a)^{n/2} \exp(M^2 |Z|^2/a)$. Let N_ε be a neighborhood of $\sigma_*^2(G)$ with radius $\varepsilon > 0$. It follows that if $(\tilde{\sigma}^2; \sigma^2) \in N_\varepsilon \times N_\varepsilon$, then $|\tilde{\sigma}^2 - \sigma^2| < 2\varepsilon$ so that, using (A.69) and (A.74), we obtain:

$$\frac{1}{f(y;\theta)} \left|\frac{\partial f(y;\tilde{\theta})}{\partial \theta}\right| \le K \left[\frac{n}{2a} - \frac{|Z|}{2} + \frac{1}{2a^2}(|y| + (a+M)|Z|)^2\right] \exp\left[\frac{\varepsilon}{a^2}|y - Z\beta|^2\right] \tag{A.75}$$

Since the moment generating function $|y - Z\beta|^2$ exists, it follows that $E_G(|y - Z\beta|^{2m} e^{t|y - Z\beta|^2}) < \infty$ for every integer m and any $t \in (-t_1, t_1)$ (see, e.g., Monfort (1980, p. 148)). This implies that $E_G(|y - Z\beta|^m e^{t|y - |\beta|^2}) < \infty$ for every integer m and every $t \in (-t_1, t_1)$. Using this latter property, and letting ε be less than $t_1/(2a^2)$, it follows that the right-hand side of (A.75), which is independent of $\tilde{\theta}$, is G-square integrable. Thus Assumption A4 is satisfied.

PROOF OF LEMMA 5.2: The result follows from Theorem 4.1 where the mapping $\phi(\cdot)$ is such that $\phi(\theta) = \beta$. Thus, using (A.72), we have:

$$\frac{\partial \phi(\theta_*)}{\partial \theta'} [A\mathcal{E}(\theta_*)]^{-1} = [-\sigma_*^2(Z'Z)^{-1}; 0]. \tag{A.76}$$

On the other hand, using Equation (5.3), the submatrix of $B\mathcal{E}(\theta_*)$ corresponding to the parameters β is:

$$[B\mathcal{E}(\theta_*)]_{\beta\beta} = E_G\left[\frac{\partial \log f(Y;\theta_*)}{\partial \beta} \cdot \frac{\partial \log f(Y;\theta_*)}{\partial \beta'}\right]$$

$$= \sigma_*^4 Z' E_G(y - Z\beta_*)(Y - Z\beta_*)' Z. \tag{A.77}$$

Since we have:

$$E_G(Y - Z\beta_*)(Y - Z\beta_*) = V + M_Z \mu\mu' M_Z, \tag{A.78}$$

as is readily established using Equation (5.25), we obtain:

$$[B\mathcal{E}(\theta_*)]_{\beta\beta} = \sigma_*^4 Z' V Z. \tag{A.79}$$

The desired result follows from (A.76), (A.79), and Equation (4.4).

FOOTNOTES

* This research was supported by National Science Foundation Grant SES-8410593. I am indebted to D. Rivers for helpful discussions. This paper is dedicated to those who have made this past year enjoyable. Remaining errors are mine.

1. This lower bound is also sometimes credited to the french statisticians Frechet (1943) and Darmois (1945).
2. The compactness assumption is used only in Section 4. Otherwise Θ can be open.
3. For a detailed discussion of these three assumptions, see White (1982). Note that we assume here square-integrability instead of simple integrability of the dominating function $M(\cdot)$ in Assumption A2. Also, the non-singularity of the matrix $A_G^f(\theta_*(G))$ can be replaced by the weaker assumption that $\theta_*(G)$ is a regular point of $A_G^f(\theta)$ (see White (1982, Theorem 3.1)).
4. It appears more convenient to impose the additional regularity condition A4 on the model $\mathbf{F}_\theta$ rather than to impose some regularity conditions on the unbiased estimators, as this is sometimes done in the correctly specified case (see, e.g., Lehmann (1983, p. 122)). As a matter of fact, the full force of Assumption A4 relative to Assumption $A4_*$ is only used in Section 4.
5. As can be seen from the proof, Assumption $A4_*$ is used only to establish part (ii). In addition, we have the stronger result that Equation (2.6) holds in the neighborhood N_* of $\theta_* = \theta_0$, i.e., $\forall\, \theta \in N_*$,

$$\int \frac{\partial^2 \log f(y;\theta)}{\partial\theta\partial\theta'} f(y;\theta)\, d\nu(y) + \int \frac{\partial \log f(y;\theta)}{\partial\theta'} f(y;\theta)\, d\nu(y) = 0.$$

6. Had Θ been open instead of compact in Assumption A1, then it would follow that a parametric model $\mathbf{F}_\theta$ which is regular is necessarily semi-regular. On the other hand the regularity of $\mathbf{F}_\theta$ implies the semi-regularity of $\mathbf{F}_\theta^0$ as seen in Corollary 3.1.
7. It is not necessary to restrict ourselves to unbiased estimators. As usual, our results can be adapted to biased estimators under the mean squared error loss function (see, e.g., Lehmann (1983, p. 128)).
8. As can be seen from the proof, the local Lipschitz assumption A4 is not used. On the other hand, Assumption B2 is important.
9. Note that we cannot in general replace the inequality (3.8) by

$$Var_G T(Y) \geq sup\, LB_G(\gamma)$$

where the sup is taken over γ such that $G = G_\gamma$. Indeed the psd matrices $LB_G(\gamma)$ are not necessarily ordered.

10. For definitions of these concepts, see Vuong (1986a).
11. I owe this question to D. Rivers.
12. As the proof shows, if $B_G^g(\gamma)$ is non-singular, then one must have $p \leq b$ as expected since $\mathbf{G}_\gamma$ is nested in $\mathbf{G}_\beta$.

13. The only difference is that the model considered here will be k-dimensional, while in the semi-parametric literature, it is in general one-dimensional.
14. Another model that contains G is one in which the densities are of the form $f(y;\theta)g(y)/f(y;\theta_*)C(\theta)$ where $C(\theta)$ is a normalizing constant. Unfortunately this model is not necessarily semi-regular. In addition $\mathbf{F}_\theta$ is not necessarily regular with respect to it.
15. A more general framework is one in which Z is random. This corresponds to the case of stochastic regressors. Then, an appropriate framework is that considered in Lien and Vuong (1986) which builds on conditional specification (see Vuong (1983, 1984)). More general frameworks with stochastic regressors are studied in White (1980b, 1984) and White and Domowitz (1984) among others. Note that in the non-random case, misspecification arising from possible correlation between the regressors and the error terms (see, e.g., Hausman, (1978)) cannot be properly handled.
16. Note that $E_{G_{1\gamma}}(Y_i) = z_i'b$ and $Var_{G_{1\gamma}}(Y_i) = s^2$. The parameter space Γ is equal to $\mathbf{B}^0 \times (\sqrt{a}, \sqrt{b})$ so that Assumptions A2 - A3 are satisfied (see Lemma 5.1 - (i)).
17. One can readily check that the model $\mathbf{G}_{1\gamma}$ and all the models considered below are semi-regular so that our results apply.
18. In what follows vec and vech are the operators that stack the columns of a matrix and a symmetric matrix (see, e.g., Henderson and Searle (1979)). For simplicity, the covariance matrix V is restricted to be non-singular, though our results holds even if V is singular.
19. Note that in this case, LB_G is equal to the lower bound (5.23) for the pseudo-true parameters $\beta_*(G)$ where G belongs to $\mathbf{G}_{3\gamma}$ which is clearly nested in $\mathbf{G}(\mathbf{F}_\theta)$. Note also that the lower bound (5.23) is at least as large as the lower bound (5.15) which is obtained when G belongs to $\mathbf{G}_{2\gamma}$. From Corollary 3.3, this is expected since $\mathbf{G}_{2\gamma} \subset \mathbf{G}_{3\gamma}$.
20. In this case, we treat the initial conditions as fixed. See also footnote 15.

REFERENCES

Andrews, D. W. K., and P. C. B. Phillips: "An Extension of Kariya's Version of the Gauss-Markov Theorem for Nonlinear Estimators," Cowles Foundation Discussion Paper, No. 761. New Haven: Yale University, 1985.

__________: "Best Median Unbiased Estimation in Linear Regression with Bounded Asymmetric Loss Functions," Cowles Foundation Discussion Paper, No. 786. New Haven: Yale University, 1986.

Bhattacharya, A.: "On Some Analogues of the Amount of Information and Their Uses in Statistical Estimation," *Sankhya*, 8(1946), 1-14.

Chamberlain, G.: "Asymptotic Efficiency in Semiparametric Models with Censoring," mimeo, University of Wisconsin, Madison, 1984.

Cramer, H.: *Mathematical Methods of Statistics*. Princeton: Princeton University Press, 1946.

Darmois, G.: "Sur les Lois Limites de la Dispersion de Certaines Estimations," *Rev. Inst. int. Statist.*, 13(1945), 9-15.

Frechet, M.: Sur l' Extension de Certaines Evaluations Statistiques de Petits Echantillons," *Rev. Inst. Int. Statist.*, 11(1943), 182-205.

Gourieroux, C., A. Monfort, and A. Trognon: "Pseudo Maximum Likelihood Methods: Theory," *Econometrica*, 52(1984), 681-700.

__________: "Pseudo Maximum Likelihood Methods: Applications to Poisson Models," *Econometrica*, 52(1984), 701-720.

Hausman, J.: "Specification Tests in Econometrics," *Econometrica*, 46(1978), 1251-1272.

Henderson, H. V., and S. R. Searle: "Vec and Vech Operators for Matrices, with Some Uses in Jacobians and Multivariate Statistics," *Canadian Journal of Statistic*, 7(1979), 65-81.

Huber, P. J.: "The Behavior of Maximum Likelihood Estimates Under Non-standard Conditions," *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability*. Berkeley: University of California Press, 1967.

Hwang, J. T.: "Universal Domination and Stochastic Domination: Estimation Simultaneously Under a Broad Class of Loss Functions," *Annals of Statistics*, 13(1985), 295-314.

Kariya, T.: "A Nonlinear Version of the Gauss-Markov Theorem," *Journal of the American Statistical Association,* 80(1985), 476-477.

Kullback, S., and R. A. Leibler: "On Information and Sufficiency," *Annals of Mathematical Statistics,* 22(1951), 79-86.

Lecam, L.: "On Some Asymptotic Properties of Maximum Likelihood Estimates and Related Bayes' Estimates," *University of California Publications in Statistics,* 1(1953), 277-330.

Lehmann, E. L.: *Theory of Point Estimation.* New York: John Wiley and Sons, 1983.

Lien, D., and Q. H. Vuong: "Selecting the Best Linear Regression Model: A Classical Approach," Social Science Working Paper, No. 606. Pasadena: California Institute of Technology, 1986.

Monfort, A.: *Cours de Probabilites.* Paris: Economica, 1980.

Rao, C. R.: "Information and Accuracy Attainable in the Estimation of Statistical Parameters," *Sankhya,* 7(1945), 16-19.

__________: "Criteria of Estimation in Large Sample," *Sankhya,* 25(1963), 189-206.

__________: *Linear Statistical Inference and Its Applications.* New York: John Wiley and Sons, 1973.

Rao, C. R., and S. K. Mitra: *Generalized Inverse of Matrices and Its Applications.* New York: John Wiley and Sons, 1971.

Richard, J. F.: "A Note on the Information Matrix of the Multivariate Normal Distribution," *Journal of Econometric,* 3(1975), 57-60.

Rosenblatt, M.: "Remarks on Some Nonparametric Estimators of Density Function," *Annals of Mathematical Statistics,* 27(1956), 832-837.

Rudin, W.: *Principles of Mathematical Analysis.* New York: McGraw-Hill, 1976.

Sawa, T.: "Information Criteria for Discriminating Among Alternative Regression Model," *Econometrica,* 46(1978), 1273-1291.

Silvey, S. D.: "The Lagrangian Multiplier Test," *Annals of Mathematical Statistics,* 30(1959), 389-407.

Stein, C.: "Efficient Nonparametric Testing and Estimation," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* 1(1956), 187-195.

Ullah, A., and R. S. Singh: "The Estimation of Probability Density Functions and its Applications in Econometrics," Technical Report, No.6. London: The University of Western Ontario, 1985.

Vuong, Q. H.: "Misspecification and Conditional Maximum Likelihood Estimation," Social Science Working Paper, No. 503. Pasadena: California Institute of Technology, 1983.

__________: "Two-Stage Conditional Maximum Likelihood Estimation of Econometric Models," Social Science Working Paper, No. 538. Pasadena: California Institute of Technology, 1984.

__________: "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," Social Science Working Paper, No. 605. Pasadena: California Institute of Technology, 1986.

__________: Generalized Inverses and Asymptotic Properties of Wald Tests," Social Science Working Paper, No. 607. Pasadena: California Institute of Technology, 1986.

White, H.: "Using Least Squares to Approximate Unknown Regression Functions," *International Economic Review*, 21(1981), 149-170.

__________: "Nonlinear Regression on Cross-Section Data," *Econometrica*, 48(1980), 721-746.

__________: "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50(1982), 1-25.

__________: "Maximum Likelihood Estimation of Misspecified Dynamic Models," in *Misspecification Analysis*, ed. T. K. Dijkstra. Berlin: Springer Verlag, 1984, 1-19.

White, H. and I. Domowitz: "Nonlinear Regression with Dependent Observations," *Econometrica*, 52(1984), 143-161.