Proofs for Efficient Method of Moments in Misspecified IID Models.

By

Víctor Aguirre-Torres* 
Departamento de Estadística

Manuel Domínguez Toribio**
Departamento de Estadística and
Departamento de Economía

Instituto Tecnológico Autónomo de México (ITAM)
Rio Hondo #1
México DF 01000
MEXICO

TEL (office): (+52) 5628-4000 ext 3835
FAX: (+52) 5628-4086
EMAIL: aguirre@itam.mx

*Víctor Aguirre-Torres acknowledges partial support from the Asociación Mexicana de la Cultura A. C.

**Manuel A. Domínguez acknowledges funding from CONACYT, research project J38076-D and partial support from Asociación Mexicana de la Cultura A. C.
Summary

This document contains the proofs that were omitted from

Aguirre-Torres, V., and Domínguez Manuel, “Efficient Method of Moments in
Misspecified IID Models”.

Published in *Econometric Theory*, 20, 2004, 513-534.

In addition, definitions, notation and statements of the theorems are reproduced. The following is a summary of the complete paper.

The paper presents the asymptotic theory of EMM when the model of interest is not correctly specified. The paper assumes a sequence of i.i.d. observations and a global misspecification. It is found that the limiting distribution of the estimator is still asymptotically normal, but it suffers a strong impact in the covariance matrix. A consistent estimator of this covariance matrix is provided. The large sample distribution on the estimated moment function is also obtained. These results are used to discuss the situation when the moment conditions hold but the model is misspecified. It is shown also that the overidentifying restrictions test has asymptotic power one whenever the limit moment function is different from zero. It is also proved that the bootstrap distributions converge almost surely to the above mentioned distributions and hence they could be used as an alternative to draw inferences under misspecification. Interestingly, it is also shown that bootstrap can be reliably applied even if the number of bootstrap replications is very small.
1. INTRODUCTION

The paper deals with the Efficient Method of Moments (EMM) presented in Gallant and Tauchen (1996). This method uses the scores of an auxiliary model to generate moment constraints. With the inspiration yielded from Hall and Inoue (2003), the purpose of this work is to make a qualitative analysis of the consequences of model misspecification on EMM estimators. Interestingly, we extend the procedures in Domínguez and Aguirre-Torres (2003) concerning the properties of the bootstrap when the number of the bootstrap replications is small so that the application of the bootstrap does not lead to a large increase in computation demands.

The paper deals with the iid case, while most of the applications of EMM are for time series data. Nonetheless, the results of the paper give an indication of the consequences of misspecification of the maintained model in time series models too. A referee also noticed that when the auxiliary model is a good approximation of the data then the scores are essentially serially uncorrelated and then one may expect the results from the iid case to be relevant.

The document is organized as follows. Section 2 presents the data generating process and the large sample properties of the quasi-maximum likelihood estimation for the auxiliary model; these properties are required to derive the asymptotic properties of the EMM procedure. Section 3 defines the EMM and discusses the almost sure limit of the estimator. Section 4 presents the results under global misspecification, including the distributional results for the estimator, the moment function, and the test for overidentifying constraints. Section 5 presents the results regarding bootstrap.

2. QUASI-MAXIMUM LIKELIHOOD ESTIMATION.
The following assumption gives the data generating process.

**Assumption 1**

The data \( \{X_t\}_{t=1}^n \) is a sequence of i.i.d. random vectors with joint continuous density \( h(x) \). The observed data will be denoted by \( \{\tilde{X}_t\}_{t=1}^n \).

The model of interest is

\[
p(x | \rho), \quad \rho \in \mathcal{R},
\]

where \( \mathcal{R} \) is the parameter space for \( \rho \) (notice that \( \mathcal{R} \) is not the real line). Model \( p \) is correctly specified if there exists \( \rho_* \in \mathcal{R} \) such that

\[
h(x) = p(x | \rho_*) \quad \text{for every } x. \tag{0}
\]

We are assuming that this model is not directly estimable by maximum likelihood and hence an alternative procedure must be used. In order to estimate \( \rho \) the EMM maximizes the pseudolikelihood of an auxiliary model that we denote by

\[
f(x | \theta), \quad \theta \in \Theta,
\]

and uses the first order condition of this problem to generate a set of moment restrictions. Formally, we consider the following notation.

**Notation 1.** Let

\[
l_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \log f(\tilde{x}_i | \theta),
\]

\[
\bar{I}(\theta) = \int \log f(x | \theta) h(x) dx,
\]

\[
\tilde{\theta}_n = \arg \max_{\theta} l_n(\theta),
\]

\[
\theta_0 = \arg \max_{\theta} \bar{I}(\theta),
\]
\[ \tilde{J}_n = \frac{1}{n} \sum_{i=1}^{n} \left[ (\partial^2 / \partial \theta \partial \theta^T) \log f(\tilde{x}_i | \tilde{\theta}_n) \right], \]

where

\[ I_0 = \int [((\partial / \partial \theta) \log f(x | \theta_0)] [(\partial / \partial \theta) \log f(x | \theta_0)]^T h(x) dx, \]

\[ J_0 = \int [(\partial^2 / \partial \theta \partial \theta^T) \log f(x | \theta_0)] h(x) dx \quad \text{and} \]

\[ V_0 = J_0^{-1} I_0 J_0^{-1}. \]

Recall from pseudomaximum likelihood theory that if the auxiliary model suits the data, \( \theta_0 \) identifies the data generating process. In order to obtain the large sample properties of the QMLE the following assumption is required.

**Assumption 2.**

The parameter space \( \Theta \) is closed and bounded. The function \( \log f(x | \theta) \), its first and second partial derivatives with respect to \( \theta \) are continuous in \( \theta \), they are also uniformly bounded in \( \theta \) by an integrable function \( b_j(x) \) such that \( \int b_j(x) h(x) dx < \infty \).

The vector \( \theta_0 \) is unique and it is an interior point of \( \Theta \). For every nonzero vector \( a \)

\[ 0 < \left\{ a^T (\partial / \partial \theta) \log f(x | \theta_0) \right\}^2 h(x) dx. \]

Moreover, there exist an open neighborhood of \( \theta_0 \), say \( NB_0 \) such that,

\[ \sup_{\theta \in NB_0} \left\{ a^T (\partial / \partial \theta) \log f(x | \theta) \right\}^2 h(x) dx < \infty. \]

The uniformly bounded part of Assumption 2 insures that averages converge to integrals, while asking for the integral to be greater than zero allows the use of the Cramer-Wold device to show asymptotic normality. With the above assumptions the following theorem gives the large sample properties of the QMLE. The proofs of these
theorems are a consequence of results given in Gallant (1987) Chapter 3 or similarly from White (1982). Hence these results are not proved.

THEOREM 1. Under Assumptions 1 and 2:

a) $\hat{\theta}_n \xrightarrow{w, p} \theta_0$.

THEOREM 2. Under Assumptions 1 and 2:

a) $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, V_\rho)$, where $V_\rho = J_\rho^{-1} I_\rho J_\rho^{-1}$.

b) $\tilde{I}_n \xrightarrow{w, p} I_\rho$ and $\tilde{J}_n \xrightarrow{w, p} J_\rho$.

In the case of the EMM, a common choice of the auxiliary model consists of a flexible semi-nonparametric (SNP) score generator that makes the efficiency of the estimator of $\rho$ reach that of maximum likelihood when model $p$ is correctly specified. The interested reader may consult Gallant and Nychka (1987) for a description of the SNP model suitable for cross sectional data. Although the score generator typically used for the EMM is the SNP, the discussion of the paper will be in general regarding the definition of the auxiliary model $f(x | \theta)$.

3. THE EFFICIENT METHOD OF MOMENTS.

The model of interest is $p(x | \rho)$ which we do not assume to be correctly specified, i.e. we do not assume that equation (0) holds.

Once the auxiliary model has been estimated, the moment function is taken as

$$m_a(\rho) = \int \left[ f(\hat{\theta} / \hat{\theta} \theta) \log f(x | \hat{\theta}_n) / p(x | \rho) \right] dx$$

(1)

A direct application of the method of moments to (1) would require finding $\rho$ that solves the equation $m_a(\rho) = 0$. Usually, the number of parameters in the auxiliary
model exceeds that of the model of interest and, instead of finding a root for (1), the estimator of $\rho$ is defined as follows

$$\hat{\rho}_n = \arg \min_{\rho \in \mathbb{R}} m_n^T(\rho) \tilde{T}_n^{-1} m_n(\rho)$$  \hspace{1cm} (2)$$

where $\tilde{T}_n$ is given in Notation 1.

The advantage of using this approach in a situation where direct maximum likelihood is not possible is as follows; if we let $\{\hat{x}_t\}_{t=1}^N$ be a long simulated sequence from $p(x | \rho)$ then (1) could be approximated by

$$m_n(\rho) = \frac{1}{N} \sum_{t=1}^N (\partial / \partial \theta) \log f(\hat{x}_t | \tilde{\theta}_n)$$

It is assumed that $N$ is large enough that the Monte Carlo integral approximates the analytic integral to within a negligible error, of the same sort that is made in computing any mathematical expression on a digital computer.

Also, the method has a built in specification test via the following test statistic

$$L_0 = n m_n^T(\hat{\rho}_n) \tilde{T}_n^{-1} m_n(\hat{\rho}_n).$$

The discussion of the large sample properties of the EMM estimator, defined in (2), requires the following notation.

**Notation 2.** Let

$$m(\rho, \theta) = \int [(\partial / \partial \theta) \log f(x | \theta)] p(x | \rho) dx,$$

$$m_n(\rho) = m(\rho, \tilde{\theta}_n),$$

$$m_0(\rho) = m(\rho, \theta_0),$$

$$M(\rho, \theta) = (\partial / \partial \rho^T) m(\rho, \theta).$$
\[
N(\rho, \theta) = (\partial / \partial \theta^\top) m(\rho, \theta),
\]
\[
M_0 = M(\rho_0, \theta_0),
\]
\[
N_0 = N(\rho_0, \theta_0),
\]
\[
s_n(\rho) = m_n^\top(\rho) \tilde{I}_n^{-1} m_n(\rho),
\]
\[
s_s(\rho) = m_s^\top(\rho) I_0^{-1} m_s(\rho),
\]
\[
\hat{\rho}_n = \arg \min_{\rho \in \mathbb{R}} s_n(\rho),
\]
\[
\rho_0 = \arg \min_{\rho \in \mathbb{R}} s_s(\rho) \text{ and }
\]
\[
V_m^o = N_0 V_0 N_0.
\]

Notice that \(N(\rho, \theta)\) will be a square symmetric matrix under regularity conditions. A referee pointed out that it is important to observe that from the definition of \(\rho_0\), its value depends on the weighting matrix \(I_0\), and also to notice that in the case of EMM, once \(f(x | \theta)\) is defined, then \(I_0\) is completely defined too, and that \(\tilde{I}_n\) is a natural choice for its estimation. Besides Assumptions 1 and 2 we require Assumption 3.

**Assumption 3**

The parameter space \(\mathbb{R}\) is closed and bounded. The function \((\log f(x | \theta)) p(x | \rho)\), its first and second partial derivatives with respect to \(\rho\) and \(\theta\) are continuous in \((x, \theta, \rho)\) and uniformly bounded in \((\theta, \rho)\) by an integrable function \(b_2(x)\) such that \(\int b_2(x) \, dx < \infty\). The vector \(\rho_0\) is unique and it is an interior point of \(\mathbb{R}\).
The matrices $M_0$ and $N_0$ are full column rank. The Hessian of $s_0(\rho)$ evaluated at $\rho_0$ is nonsingular.

**THEOREM 3.** Under Assumptions 1 through 3, $\hat{\rho}_n \xrightarrow{up} \rho_0$.

Notice that Theorem 3 does not require that moment conditions hold true, that is, it may happen that

$$m_0(\rho_0) \neq 0, \quad (3)$$

but, since $\rho_0$ is a minimum of $s_0(\rho)$, then first-order conditions imply

$$(\partial / \partial \rho) s_0(\rho_0) = 0,$$

and then, differentiation of the quadratic $s_0$ results in

$$M_0^T T_0^{-1} m_0(\rho_0) = 0. \quad (4)$$

This last equation does not imply that $m_0(\rho_0) = 0$ either.

A random vector that plays an important role in the large sample results is the moment function $m$ when only the $\theta$ argument is random. The next lemma gives its large sample distribution under misspecification.

**LEMMA 1.** Under Assumptions 1 through 3:

$$\sqrt{n} [ m_n(\rho_0) - m_0(\rho_0) ] \xrightarrow{L} N(0, V_m^0), \text{ where } V_m^0 = N_0 V_\theta N_0.$$

The matrix $V_m^0$ represents the variance-covariance of the moment vector when only the $\theta$ argument is random. A referee suggested to point out that the matrix $V_m^0$ is different from the expression (17) of Gallant and Tauchen (1996); in fact expression (18) of Gallant and Tauchen (1996) results from Lemma 1 under correct specification.

**4. EMM UNDER GLOBAL MISSPECIFICATION.**

4.1. The Large Sample Distribution of the Parameter Estimator.
The starting point to deal with the large sample approximation of $\hat{\rho}_n$ is the first order condition for $s_n(\rho)$

$$
(\partial / \partial \rho)s_n(\hat{\rho}_n) = 2 \hat{M}_n^T (\hat{I}_n)^{-1} m_n(\hat{\rho}_n) = 0,
$$

(5)

where $\hat{M}_n = M(\hat{\rho}_n, \tilde{\theta}_n)$.

The next step is to apply a Taylor series approximation to $m_n(\hat{\rho}_n)$. Let $m_{in}(\hat{\rho}_n)$ be the i-th coordinate of $m_n(\hat{\rho}_n)$, from Theorem B page 44 of Serfling (1980), there exists a $\tilde{\rho}_n$ such that $\|\tilde{\rho}_n - \rho_0\| \leq \|\hat{\rho}_n - \rho_0\|$ and

$$
m_{in}(\hat{\rho}_n) = m_{in}(\rho_0) + (\partial / \partial \rho^T)m_{in}(\tilde{\rho}_n)(\hat{\rho}_n - \rho_0)
$$

if we let $\tilde{M}_n$ the matrix that is obtained stacking $(\partial / \partial \rho^T)m_{in}(\tilde{\rho}_n)$ then, stacking the above scalar equations gives the vector approximation

$$
m_a(\hat{\rho}_n) = m_a(\rho_0) + \tilde{M}_n(\hat{\rho}_n - \rho_0)
$$

(6)

and notice that $\tilde{M}_n$ converges almost surely to $M_0$. Now let $\hat{W}_n = \hat{M}_n^T \tilde{I}_n^{-1} \tilde{M}_n$, then substituting (6) into (5) gives

$$
\hat{W}_n(\hat{\rho}_n - \rho_0) = -\hat{M}_n^T \tilde{I}_n^{-1} m_n(\rho_0).
$$

(7)

This equation holds whether the model is correctly or incorrectly specified and relates $(\hat{\rho}_n - \rho_0)$ to $m_n(\rho_0)$. Under correct specification (7) is all that is needed to obtain the large sample distribution of $\sqrt{n}(\hat{\rho}_n - \rho_0)$ because, from Lemma 1 $m_n(\rho_0)$ is itself asymptotically normal centered around zero. When the model is incorrectly specified it may happen that $m_0(\rho_0) \neq 0$ but in this case we can write

$$
\hat{W}_n \sqrt{n}(\hat{\rho}_n - \rho_0) = -\hat{M}_n^T \tilde{I}_n^{-1} \sqrt{n}[m_n(\rho_0) - m_0(\rho_0)] - \sqrt{n} \hat{M}_n^T \tilde{I}_n^{-1} m_0(\rho_0).
$$

(8)
The first term of (8) converges in law to a normal distribution from Lemma 1. The presence of the second term will require to deal with the joint distribution of \( \hat{M}_n \) and \( \tilde{I}_n^{-1} \), which in turn leads to consider the joint distribution of \( \tilde{\theta}_n \) and \( \tilde{I}_n^{-1} \). For that reason we require the next assumption.

**Assumption 4.** Let

\[
c(x, \theta) = \begin{bmatrix} (\partial / \partial \theta) \log f(x | \theta) \\ \text{vec}\{(\partial / \partial \theta) \log f(x | \theta)\} \{\partial / \partial \theta) \log f(x | \theta)^T\} \end{bmatrix}.
\]

Then, each element of \( c(x, \theta) \), \( c(x, \theta)c^T(x, \theta) \) and \( (\partial / \partial \theta^T)c(x, \theta) \) is continuous and dominated by the function \( b_1(x) \) from Assumption 2.

The following lemma establishes the asymptotic joint normal distribution of \( \tilde{\theta}_n \) and \( \tilde{I}_n^{-1} \).

**LEMMA 2.** Under Assumptions 1 thru 4

\[
\sqrt{n} \begin{bmatrix} \tilde{\theta}_n - \theta_0 \\ \text{vec}(\tilde{I}_n^{-1} - I_0^{-1}) \end{bmatrix} \xrightarrow{L} N\left(0, \begin{bmatrix} V_{\theta} & V_{\theta I} \\ V_{\theta I}^T & V_I \end{bmatrix}\right),
\]

where

\[
\begin{bmatrix} V_{\theta} & V_{\theta I} \\ V_{\theta I} & V_I \end{bmatrix} = \begin{bmatrix} -J_0^{-1} & 0 \\ (I_0^{-1} \otimes I_0^{-1})R_0J_0^{-1} & -(I_0^{-1} \otimes I_0^{-1}) \end{bmatrix} \Gamma_0 \begin{bmatrix} -J_0^{-1} & 0 \\ (I_0^{-1} \otimes I_0^{-1})R_0J_0^{-1} & -(I_0^{-1} \otimes I_0^{-1}) \end{bmatrix}^T,
\]

\[
R_0 = \int \frac{\partial}{\partial \theta^T} \text{vec}\left[ \left(\frac{\partial}{\partial \theta} \log f(x | \theta_0)\right) \left(\frac{\partial}{\partial \theta} \log f(x | \theta_0)^T\right) \right] h(x)dx,
\]

\[
\Gamma_0 = \int c(x, \theta_0)c(x, \theta_0)^T h(x)dx - \left[ \int c(x, \theta_0)h(x)dx \right] \left[ \int c(x, \theta_0)h(x)dx \right]^T,
\]

which is partitioned accordingly.
In order to find the expression of the covariance matrix for $\hat{\rho}$, additional notation is required.

**Notation 3.** Let

$k$ any natural number,

$I_k$ the identity matrix\(^{1}\) of order $k$,

$p = \text{dim}(\rho)$, $q = \text{dim}(\theta)$,

$m_i(\rho,\theta)$ $i$-th element of vector $m(\rho,\theta)$ for $i = 1, 2, ..., q$,

$H_i(\rho,\theta) = (\partial^2 / \partial \rho^T \partial \rho)m_i(\rho,\theta)$ the Hessian matrix of $m_i(\rho,\theta)$ relative to $\rho$,

$$H(\rho,\theta) = \begin{bmatrix} H_1(\rho,\theta) \\ \vdots \\ H_q(\rho,\theta) \end{bmatrix},$$

$$G_1 = [m_{0}^T(\rho_0)I_0^{-1} \otimes I_p] H(\rho_0,\theta_0),$$

$Q_i(\rho,\theta) = (\partial^2 / \partial \theta^T \partial \rho)m_i(\rho,\theta)$ the second crossed partial derivative of $m_i(\rho,\theta)$,

$$Q(\rho,\theta) = \begin{bmatrix} Q_1(\rho,\theta) \\ \vdots \\ Q_q(\rho,\theta) \end{bmatrix},$$

$$G_2 = [m_{0}^T(\rho_0)I_0^{-1} \otimes I_p] Q(\rho_0,\theta_0),$$

$$G_3 = m_{0}^T(\rho_0) \otimes M_0^T,$$

$$G_4 = M_0^T I_0^{-1} N_0 + G_2,$$

\(^{1}\) Notice that since 0 is not a natural number, we can keep the notation $I_0$ for the information matrix. We realize that the notation is confusing but we prefer to keep the usual notation for both, the identity and the information matrices.
\[
\Omega = \begin{bmatrix} G_1 & G_2 \\ V_\theta & V_{\theta l} \end{bmatrix} \begin{bmatrix} V_\theta^T & V_{\theta l}^T \\ V_{\theta l} & V_l \end{bmatrix} \begin{bmatrix} G_1^T \\ G_2^T \end{bmatrix},
\]

\[
W_0 = M_0^T I_0^{-1} M_0,
\]

and \[ V_\rho = (W_0 + G_j)^{-1} \Omega (W_0 + G_j)^{-1}. \]

THEOREM 4. Under Assumptions 1 thru 4

\[
\sqrt{n}(\hat{\rho} - \rho_0) \xrightarrow{L} N(0, V_\rho).
\]

COROLLARY 1. Under Assumptions 1 through 4, and if \( m_0(\rho_0) = 0 \), then

\[
\sqrt{n}(\hat{\rho} - \rho_0) \xrightarrow{L} N(0, W_0^{-1} M_0^T I_0^{-1} N_0 V_\rho N_0 I_0^{-1} M_0 W_0^{-1}).
\]

4.2. The Large Sample Distribution of the Moment Function.

Its distribution can be derived from the asymptotic distribution of \( \hat{\rho} \). In fact, subtracting \( m_0(\rho_0) \) from both sides of (6) gives

\[ m_n(\hat{\rho}_n) - m_0(\rho_0) = m_n(\rho_0) - m_0(\rho_0) + M_n(\hat{\rho}_n - \rho_0) \quad (9) \]

Performing a Taylor series approximation similar to the one that led to (6), there exists \( \tilde{N}_n \) such that converges with probability one to \( N_0 \) and

\[ m_n(\rho_0) - m_0(\rho_0) = \tilde{N}_n(\tilde{\theta}_n - \theta_0), \]

substitution of this last equality into (9) produces

\[ m_n(\hat{\rho}_n) - m_0(\rho_0) = \tilde{N}_n(\tilde{\theta}_n - \theta_0) + M_n(\hat{\rho}_n - \rho_0) \quad (10) \]

From formula (A.18) in the proof of Theorem 4 there exist matrices \( \tilde{G}_{3n} \), \( \tilde{G}_{2n} \), such that

\[ \hat{\rho}_n - \rho_0 = -\tilde{G}_{3n}(\tilde{M}_n^T \tilde{I}_n^{-1} \tilde{N}_n + \tilde{G}_{2n})(\tilde{\theta}_n - \theta_0) - \tilde{G}_{3n} G_j \text{vec}(\tilde{I}_n^{-1} - I_0^{-1}) \quad (11) \]
then substituting (11) into (10) produces an expression of $m_n(\hat{\rho}_n) - m_0(\rho_0)$ as a matrix times the vector $\left[\tilde{\theta}_n - \theta_0 \right]$, and this is enough to derive its large sample distribution. The following notation is useful to describe the corresponding results.

**Notation 4.** Let

$$G_5 = (W_0 + G_1)^{-1},$$

$$A = \left[ N_0 - M_0 G_5 (M_0^T \lambda_0^{-1} N_0 + G_2) - M_0^T G_5 G_3 \right],$$

$$V_m = A \begin{bmatrix} V_\theta & V_\theta I \\ V_\lambda T & V_I \end{bmatrix} A^T,$$

$$U_0 = I_q - M_0 W_0^{-1} M_0^T I_0^{-1},$$

$$\Sigma_m = U_0 N_0 V_\theta N_0 U_0^T.$$ 

**Theorem 5.** Under Assumptions 1 thru 4

$$\sqrt{n} \left[ m_n(\hat{\rho}_n) - m_0(\rho_0) \right] \xrightarrow{d} N(0,V_m)$$

**4.3. Behavior of the Test for Overidentifying Restrictions.**

The following results characterize the large sample behavior of $L_0 = n m_n^T(\hat{\rho}_n) \tilde{I}_n^{-1} m_n(\hat{\rho}_n)$.

**Theorem 6.** Under Assumptions 1 thru 4.

a) If $m_0(\rho_0) \neq 0$, then for every $c > 0$

$$P(L_0 > c) \to 1 \text{ when } n \to \infty$$

b) If $m_0(\rho_0) = 0$, then
where \( X \approx N(0, \Sigma_m^\theta) \).

**LEMMA 3.** Let \( U_0 \) be defined as in Notation 4. Then

a) \( U_0 \) is an idempotent matrix.

b) \( U_0^T I_0^{-1} U_0 = I_0^{-1} U_0 = U_0^T I_0^{-1} \)

The following theorem establishes a necessary and sufficient condition for the limiting distribution of Theorem 6, b) to be chi-squared.

**THEOREM 7.** Under assumptions 1 thru 4, if \( m_0(\rho_0) = 0 \), then \( L_0 \) has a central chi-squared distribution if and only if

\[
\Sigma_m^0 I_0^{-1} N_0 V_0 N_0^T \Sigma_m^0 = \Sigma_m^0 I_0^{-1} N_0^T V_0 N_0 U_0^T
\]

in which case the degrees of freedom is \( \text{trace}(I_0^{-1} \Sigma_m^0) \).

It is not clear whether (12) is true in general, but the following corollary states a sufficient condition for the quadratic form to have a chi-squared distribution. Condition (13) is obtained by “canceling” the first five matrices on both sides of (12).

**COROLLARY 2.** Under assumptions 1 thru 4. If \( m_0(\rho_0) = 0 \), \( L_0 \) has a central chi-squared distribution if

\[
I_0^{-1} \Sigma_m^0 = U_0^T
\]

in which case the degrees of freedom is \( \text{dim}(\theta) - \text{dim}(\rho) \).

**5. BOOTSTRAPPING EMM**

In order to design a suitable resampling scheme that is able to do the job, two features of the problem in hand deserve our attention, namely, that the model is misspecified and that the sample is random. Hence, in first place, if the researcher wants
to prevent from the effects of misspecification, she should use a resampling scheme that does not use the specified model. And in second place, the resampling plan needs to generate a bootstrap sample that is random. Therefore, the natural resampling plan for obtaining the artificial sample is by sampling with replacement from the original sample because first, the bootstrap sample will replicate the behavior of the data generating process without imposing any additional structure and second, the observations of the artificial sample are conditionally independent. Formally, we consider that \( \{X^*_t\}_{t=1}^n \) is a bootstrap sample drawn with replacement from \( \{x_t\}_{t=1}^n \).

5.1. Bootstrapping QMLE.

We first obtain the result for the auxiliary model.

**Notation 5.** Let

\[
I^*_n(\theta) = \frac{1}{n} \sum_{t=1}^n \log f(X^*_t | \theta),
\]

\[
\tilde{\theta}^*_n = \arg \max_{\theta \in \Theta} I^*_n(\theta),
\]

\[
m^*_n(\rho) = m(\rho, \tilde{\theta}^*_n),
\]

\[
\tilde{I}^*_n = \frac{1}{n} \sum_{t=1}^n \left[ \frac{\partial}{\partial \theta} f(X^*_t, \tilde{\theta}^*_n) \right] \left[ \frac{\partial}{\partial \theta} f(X^*_t, \tilde{\theta}^*_n) \right]^T,
\]

\[
s^*_n(\rho) = m^*_n(\rho)^T (\tilde{I}^*_n)^{-1} m^*_n(\rho),
\]

\[
\hat{\rho}^*_n = \arg \min_{\rho \in \mathbb{R}} s^*_n(\rho),
\]

\[
\Rightarrow^{\rightarrow} \text{a.s. or } \Leftrightarrow^{\rightarrow} \text{a.s. convergence in probability or in distribution of the bootstrap random variable for almost every realization of the original sample.}
\]
LEMMA 4. Under Assumptions 1 and 2, \( |l_n^*(\theta) - l_n(\theta)| \overset{p}{\to} 0 \) a.s.

As a consequence of the triangular inequality, we can state the following corollary.

COROLLARY 3. Under Assumptions 1 and 2, \( |l_n^*(\theta) - l(\theta)| \overset{p}{\to} 0 \) a.s.

Now, we extend the pointwise convergence results above to uniform convergence.

LEMMA 5. Under Assumptions 1 and 2, \( \sup_{\theta \in \Theta} |l_n^*(\theta) - l(\theta)| \overset{p}{\to} 0 \) a.s.

Given the uniform convergence of the criterion functions, the next is a standard result in the extremum estimator theory.

COROLLARY 4. Then, under Assumptions 1 and 2

a) \( \tilde{\theta}^*_n - \hat{\theta}_n \overset{p}{\to} 0 \) a.s.

b) \( \tilde{\theta}^*_n - \theta_0 \overset{p}{\to} 0 \) a.s.

We now state the consistency of the bootstrap distribution.

THEOREM 8. Under Assumption 1 and 2 then

\[
\sqrt{n}(\tilde{\theta}^*_n - \tilde{\theta}_n) \overset{d}{\to} N(0, V_0) \text{ a.s.}
\]

5.2 Bootstrapping EMM.

Once the asymptotic normality result of the bootstrap for the auxiliary model is available, we focus on the properties of the bootstrap for the parameters of interest. Again, we state first the convergence properties of the minimizing function. Henceforth, let \( \| \| \) denote the Euclidean norm in the corresponding finite dimensional vector space.

LEMMA 6. Under Assumptions 1, 2 and 3.

a) \( \sup_{\rho \in \mathbb{R}} \| m_n^*(\rho) - m_n(\rho) \| \overset{p}{\to} 0 \) a.s.
b) \( \sup_{\rho \in \mathbb{R}} \| m_n(\rho) - m_0(\rho) \| \to 0 \) a.s.

COROLLARY 5. Under Assumption 1, 2 and 3.

a) \( \sup_{\rho \in \mathbb{R}} \| s_n^*(\rho) - s_n(\rho) \| \to 0 \) a.s.

b) \( \sup_{\rho \in \mathbb{R}} \| s_n(\rho) - m(\rho)^T (\overline{I}_n)^{-1} m(\rho) \| \to 0 \) a.s.

COROLLARY 6. Under Assumptions 1, 2 and 3.

\[ \hat{\rho}_n^* - \hat{\rho}_n \to 0 \) a.s.

To show that the proposed bootstrap is able to fit the distribution of interest, we will mimic the steps given in sections 3 and 4 above, using the same arguments but conditionally on the sample.

LEMMA 7: Under assumptions 1 to 4

\[
\sqrt{n} \left[ \tilde{\theta}_n^* - \tilde{\theta}_n \right] \to^d N \left( 0, \begin{bmatrix} V_{\theta} & V_{\theta_1} \\ V_{\theta_1}^T & V_I \end{bmatrix} \right) \text{ a.s.}
\]

Finally, we can state the distributional result for \( \hat{\rho}_n^* \).

THEOREM 9, Under Assumptions 1 to 5.

\[
\sqrt{n} (\hat{\rho}_n^* - \hat{\rho}_n) \to^d N(0, V_\rho) \text{ a.s.}
\]

5.3 Applications of Bootstrap to EMM.

The first and most obvious form is using the bootstrap to obtain the percentiles of the distribution of \( \sqrt{n}(\hat{\rho}_n - \rho) \). This is set via the continuous mapping theorem, if \( w \) is a continuous function from the parameter space \( \mathbb{R} \) into the real line, then using a Polya theorem type result, Serfling (1980) p. 18, we can write
Since inference for $\rho_0$ is usually related with functions $w$, we can state the following corollary.

**COROLLARY 7.** Under Assumptions 1 to 4.

$$
\lim_{n\to\infty} P \left[ w\left( \sqrt{n}(\hat{\rho}_n - \rho_0) \right) \right] = \Phi(\alpha),
$$

where, in general, $F_Z$ denotes the distribution function of the random variable $Z$. Hence, we can use the percentiles of the $w(\sqrt{n}(\hat{\rho}_n - \hat{\rho}_b))$ to make inference about $\rho_0$. If, as usual, the percentiles of $w$ are estimated using Monte-Carlo experiments, a large number of bootstrap replications will be needed and since in each bootstrap replication a further Monte-Carlo experiment is need in order to compute $\hat{\rho}_n^*$, the computation demands of the method can become enormous.

In order to reduce the computation requirements, we also suggest a different application of the Bootstrap method to make inference on the misspecified model. This method is based on the results in Domínguez and Aguirre (2003), where it is shown that if $\hat{\rho}_n^*, \hat{\rho}_n^*, ..., \hat{\rho}_n^*$ are $B$ bootstrap replications of $\hat{\rho}_n$, then, if $\rho(i)$ is the i-th entry of vector $\rho$

$$
\frac{\hat{\rho}_n(i) - \rho_0(i)}{\left\{ \frac{1}{B} \sum_{b=1}^{B} [\hat{\rho}_n^*(i) - \hat{\rho}_n(i)]^2 \right\}^{1/2}} \overset{L}{\rightarrow} t_B,
$$

where $t_B$ is a $t$-student distribution with $B$ degrees of freedom. Notice that $B$ in (14) remains fixed, and it may even be equal to one. A thorough discussion on the choice of


\( B \) is given in Domínguez and Aguirre (2003). Equation (14) could easily be used as a pivotal quantity to get confidence intervals or to make hypothesis tests. The procedure can be straightforwardly generalized for inference on whole vector, that is, if \( A \) is a full row rank \( r \times p \) matrix, then, as long as \( B > r \)

\[
\left[ A(\hat{\rho}_n - \rho_0) \right]^T \left\{ \frac{1}{B} \sum_{b=1}^{B} \left[ A(\hat{\rho}_{n,b} - \hat{\rho}_n) \right][A(\hat{\rho}_{n,b} - \hat{\rho}_n)]^T \right\}^{-1} \left[ A(\hat{\rho}_n - \rho_0) \right] \xrightarrow{L} T^2, \tag{15}
\]

where \( T^2 \) is Hotelling’s central \( T^2 \) distribution. Hence (15) could be used to get confidence regions or to test hypotheses on \( A\rho_0 \).

**APPENDIX**

**Proof of Theorem 3.** This theorem follows from THEOREM 7, page 208 of Gallant (1987) making the following correspondences: “\( m_n(\lambda) \)” is ”\( m_n(\rho) \)” ; “\( m^*(\lambda) \)” is “\( m_o(\rho) \)” ; “\( s_n(\lambda) \)” is “\( s_n(\rho) \)” ; “\( s^*(\lambda) \)” is “\( s_o(\rho) \)” ; “\( \lambda_n \)” is “\( \hat{\rho}_n \)” ; and “\( \lambda^* \)” is “\( \lambda_0 \)”.

**Proof of Lemma 1.** From Theorem 2 \( \sqrt{n}(\tilde{\theta}_n - \theta_0) \xrightarrow{L} N(0, V_0) \). On the other hand \( m_n(\rho_0) = m(\rho_0, \tilde{\theta}_n) \) is a continuous differentiable function of \( \theta \), and the partial derivative of \( m \) at \( \theta_0 \) is different from zero, hence from Theorem A, Section 3.3 of Serfling (1980) it follows that

\[
\sqrt{n}(m(\rho_0, \tilde{\theta}_n) - m(\rho_0, \theta_0)) \xrightarrow{L} N(0, N_0 V_0 N_0^T)
\]

and hence the result follows.
Proof of Lemma 2. From the proof of THEOREM 5, Chapter 3, Gallant (1987) it is found that

\[ \sqrt{n}(\tilde{\theta} - \theta) = -\frac{J^{-1}}{\sqrt{n}} \sum_{i=1}^{n} (\partial / \partial \theta) \log f(x_i | \theta) + O_n, \]  

(A.1)

where \( O_n \) converges to zero with probability one.

Define the \( q^2 \times 1 \) vector

\[ a(x, \theta) = \text{vec}\{[(\partial / \partial \theta) \log f(x | \theta)](\partial / \partial \theta) \log f(x | \theta)^T \} \]

and let \( a_i(x, \theta) \) be its \( i \)-th coordinate. Then

\[ \text{vec}(\tilde{I}_n - I_0) = \frac{1}{n} \sum_{i=1}^{n} \left[ \begin{array}{c} a(x_i, \tilde{\theta}) - a(x_i, \theta) h(x) dx \end{array} \right] \]  

(A.2)

and the mean value theorem allows us to write that \( \sqrt{n} \left[ \begin{array}{c} \tilde{\theta} - \theta \\ \text{vec}(\tilde{I}_n - I_0) \end{array} \right] \) is tail equivalent to

\[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ \begin{array}{c} -J^{-1} \partial / \partial \theta) \log f(x_i | \theta) \\ a(x_i, \theta) - a(x_i, \theta) h(x) dx + R_o \sqrt{n}(\tilde{\theta} - \theta) \end{array} \right] \]

\[ = \left[ \begin{array}{cc} -J^{-1} & 0 \\ -R_o J^{-1} \end{array} \right] \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ \begin{array}{c} c(x_i, \theta) - c(x_i, \theta) h(x) dx \end{array} \right]. \]  

(A.3)

where \( R_o = \int (\partial / \partial \theta) a(x, \theta) h(x) dx \), \( c(x, \theta) \) is partitioned as in the definition and \( I_{q^2} \) is the identity matrix of dimension \( q^2 \).

From THEOREM 2, Chapter 3 of Gallant (1987) (A.3) has a joint normal distribution and hence \( \sqrt{n} \left[ \begin{array}{c} \tilde{\theta} - \theta \\ \text{vec}(\tilde{I}_n - I_0) \end{array} \right] \) has a joint \( N(0, \Pi_o) \) singular distribution, with
\[ \Pi_0 = \begin{bmatrix} -J_0^{-1} & 0 \\ -R_0J_0^{-1} & I_{q^2} \end{bmatrix} \Gamma_0 \begin{bmatrix} -J_0^{-1} & 0 \\ -R_0J_0^{-1} & I_{q^2} \end{bmatrix}^T \]

and

\[ \Gamma_0 = \int c(x,\theta_0) c(x,\theta_0)^T h(x) dx - \int \int c(x,\theta_0) h(x) dx \int \int c(x,\theta_0) h(x) dx \]

The symmetry of the information matrix produces the singularity of \( \Gamma_0 \), but this fact is covered in Gallant’s result. Now, from Assumption 2, \( I_0 \) is positive definite, hence the inverse transformation is continuous and differentiable on a neighborhood around \( I_0 \).

Also the derivative of the inverse is different from zero, since from Proposition 106 of Dhrymes (1984), for any invertible \( p \times p \) invertible matrix \( A \),

\[ \frac{\partial \text{vec}(A^{-1})}{\partial \text{vec}(A)} \bigg|_n = -(I_0^{-1} \otimes I_0^{-1}). \]

Hence, \( \sqrt{n}[\tilde{\theta}_n - \theta_0] \)

\[ \text{vec}(\tilde{I}_n^{-1} - I_0^{-1}) \]

is a differentiable transformation of a vector that is asymptotically normal, and the partial derivative of the transformation evaluated at the limit is different from zero. Thus, from Theorem A, Chapter 3 of Serfling (1980) it is asymptotically normal too. The expression for the covariance matrix is:

\[ \begin{bmatrix} I_p & 0 \\ 0 & -(I_0^{-1} \otimes I_0^{-1}) \end{bmatrix} \Pi_0 \begin{bmatrix} I_p & 0 \\ 0 & -(I_0^{-1} \otimes I_0^{-1}) \end{bmatrix}. \]

The above matrix could be partitioned into the blocks \( V_\theta \), \( V_{\theta t} \), and \( V_t \) mentioned in lemma 2 and then

\[ \begin{bmatrix} V_\theta & V_{\theta t} \\ V_{\theta t} & V_t \end{bmatrix} = \begin{bmatrix} -J_0^{-1} & 0 \\ -R_0J_0^{-1} & (I_0^{-1} \otimes I_0^{-1}) \end{bmatrix} \Gamma_0 \begin{bmatrix} -J_0^{-1} & 0 \\ -R_0J_0^{-1} & (I_0^{-1} \otimes I_0^{-1}) \end{bmatrix}^T. \]
**Proof of Theorem 4.** Consider (8). The first term in this equation is
\[ -\hat{M}_n^T \tilde{I}_n^{-1} \sqrt{n} \left[ m_n(\rho_0) - m_0(\rho_0) \right]. \] (A.4)

Performing a Taylor series approximation similar to the one that led to (6), there exists \( \tilde{N}_n \) such that converges with probability one to \( N_0 \) and \( m_n(\rho_0) - m_0(\rho_0) = \tilde{N}_n(\tilde{\theta}_n - \theta_0) \). Thus, we can write (A.4) as
\[ -\hat{M}_n^T \tilde{I}_n^{-1} \sqrt{n} \left[ m_n(\rho_0) - m_0(\rho_0) \right] = -\hat{M}_n^T \tilde{I}_n^{-1} \tilde{N}_n \sqrt{n} (\tilde{\theta}_n - \theta_0). \] (A.5)

On the other hand, the second term of (8) is
\[ -\sqrt{n} \hat{M}_n^T \tilde{I}_n^{-1} m_0(\rho_0). \] (A.6)

Recall that \( \hat{M}_n = M^T(\hat{\rho}_n, \hat{\theta}_n) \). Then
\[ \hat{M}_n^T \tilde{I}_n^{-1} m_0(\rho_0) = \left[ M^T(\hat{\rho}_n, \hat{\theta}_n) - M^T(\rho_0, \tilde{\theta}_n) \right] \tilde{I}_n^{-1} m_0(\rho_0) + M^T(\rho_0, \tilde{\theta}_n) \tilde{I}_n^{-1} m_0(\rho_0) \] (A.7)
and from Corollary 25, page 103 of Dhrymes (1984)
\[ \left[ M^T(\hat{\rho}_n, \tilde{\theta}_n) - M^T(\rho_0, \tilde{\theta}_n) \right] \tilde{I}_n^{-1} m_0(\rho_0) = \left[ m_0^T(\rho_0) \tilde{I}_n^{-1} \otimes I_p \right] \text{vec} \left[ M^T(\hat{\rho}_n, \tilde{\theta}_n) - M^T(\rho_0, \tilde{\theta}_n) \right] \] (A.8)

Let
\[ m(\rho, \theta) = \begin{bmatrix} m_1(\rho, \theta) \\ \vdots \\ m_n(\rho, \theta) \end{bmatrix}. \]
Then \( M^T(\rho, \theta) = \left[ (\partial / \partial \rho) m_1(\rho, \theta), \ldots, (\partial / \partial \rho) m_n(\rho, \theta) \right] \). Applying a Taylor series approximation similar to the one that led to (6), there exists \( \tilde{H}_n^i \) such that
converges with probability one to \( H_0^i = (\partial^2 / \partial \rho^T \partial \rho) m_i(\rho, \theta_0) \), the Hessian matrix of the \( i \)-th moment relative to \( \rho \), and

\[
(\partial / \partial \rho) m_i(\hat{\rho}_n, \tilde{\theta}_n) = (\partial / \partial \rho) m_i(\rho_0, \tilde{\theta}_n) + \tilde{H}_n^i(\hat{\rho}_n - \rho_0).
\]  \hspace{1cm} (A.9)

Substitution of (A.9) into the matrix \( M^T \) in equation (A.8) produces

\[
M^T(\hat{\rho}_n, \tilde{\theta}_n) - M^T(\rho_0, \tilde{\theta}_n) = \begin{bmatrix} \tilde{H}_n^1(\hat{\rho}_n - \rho_0), & \ldots, & \tilde{H}_n^q(\hat{\rho}_n - \rho_0) \end{bmatrix}
\]

and hence

\[
\text{vec}\begin{bmatrix} M^T(\hat{\rho}_n, \tilde{\theta}_n) - M^T(\rho_0, \tilde{\theta}_n) \end{bmatrix} = \tilde{H}_n(\hat{\rho}_n - \rho_0)
\]

where \( \tilde{H}_n = \begin{bmatrix} \tilde{H}_n^T \end{bmatrix} \) is a \( pq \) by \( p \) matrix. Therefore (A.8) becomes

\[
\begin{bmatrix} M^T(\hat{\rho}_n, \tilde{\theta}_n) - M^T(\rho_0, \tilde{\theta}_n) \end{bmatrix} \tilde{I}_n^{-1} m_0(\rho) = \begin{bmatrix} m_0^T(\rho_0) \tilde{I}_n^{-1} \otimes I_p \end{bmatrix} \tilde{H}_n(\hat{\rho}_n - \rho_0).
\]  \hspace{1cm} (A.10)

Let

\[
\tilde{G}_j = \begin{bmatrix} m_0^T(\rho_0) \tilde{I}_n^{-1} \otimes I_p \end{bmatrix} \tilde{H}_n
\]

which converges with probability one to \( G_j \) given in Notation 3. Substituting (A.10) into (A.7) results in

\[
\tilde{M}_n^T \tilde{I}_n^{-1} m_0(\rho_0) = \tilde{G}_j(\hat{\rho}_n - \rho_0) + M^T(\rho_0, \tilde{\theta}_n) \tilde{I}_n^{-1} m_0(\rho_0).
\]  \hspace{1cm} (A.11)

Now substitution of (A.5) and (A.11) into (8) yields

\[
\hat{W}_n \sqrt{n}(\hat{\rho}_n - \rho_0) = -\tilde{M}_n^T \tilde{I}_n^{-1} \tilde{N}_n \sqrt{n}(\hat{\theta}_n - \theta_0) - \tilde{G}_j(\hat{\theta}_n - \theta_0) - \sqrt{n} M^T(\rho_0, \tilde{\theta}_n) \tilde{I}_n^{-1} m_0(\rho_0)
\]

24
and hence
\[
(W_n + G_{in})\sqrt{n}(\hat{\rho}_n - \rho_0) = -\hat{M}_n^{T} \tilde{I}_n^{-1} \hat{N}_n \sqrt{n}(\tilde{\theta}_n - \theta_0) - \sqrt{n}M_T(\rho_0, \tilde{\theta}_n)\tilde{I}_n^{-1}m_0(\rho_0).
\]

(A.12)

Now consider that \(M_T(\rho_0, \tilde{\theta}_n)\tilde{I}_n^{-1}m_0(\rho_0)\) may be written as
\[
M_T(\rho_0, \tilde{\theta}_n)\tilde{I}_n^{-1}m_0(\rho_0) = [M_T(\rho_0, \tilde{\theta}_n) - M_T(\rho_0, \theta_0)]\tilde{I}_n^{-1}m_0(\rho_0) + M_T^{T}\tilde{I}_n^{-1}m_0(\rho_0)
\]

(A.13)

Arguments similar to those already used in (A.10) lead to
\[
[M_T(\rho_0, \tilde{\theta}_n) - M_T(\rho_0, \theta_0)]\tilde{I}_n^{-1}m_0(\rho_0) = \tilde{G}_{2n}(\tilde{\theta}_n - \theta_0),
\]

(A.14)

where \(\tilde{G}_{2n} = [m_T^{T}(\rho_0)\tilde{I}_n^{-1} \otimes I_p] \tilde{Q}_n\),
\[
\tilde{Q}_n = \begin{bmatrix} \tilde{Q}_n^1 \\ \vdots \\ \tilde{Q}_n^q \end{bmatrix}
\]

and for each \(j = 1, 2, ..., q\), \(\tilde{Q}_n^j\) converges with probability one to \((\partial^2 / \partial \theta^T \partial \rho)m_j(\rho_0, \theta_0)\). The almost sure limit of \(\tilde{G}_{2n}\) is \(G_2\) given in Notation 3.

Finally the term \(M_T^{T}\tilde{I}_n^{-1}m_0(\rho_0)\) from (A.13) can be expressed as
\[
M_T^{T}\tilde{I}_n^{-1}m_0(\rho_0) = M_T^{T}(\tilde{I}_n^{-1} - I_0^{-1})m_0(\rho_0) + M_T^{T}I_0^{-1}m_0(\rho_0).
\]

But from (4), \(M_T^{T}I_0^{-1}m_0(\rho_0) = 0\) and hence
\[
M_T^{T}\tilde{I}_n^{-1}m_0(\rho_0) = M_T^{T}(\tilde{I}_n^{-1} - I_0^{-1})m_0(\rho_0)
\]

and also, from corollary 25 in Dhrymes (1984),
Recall from Notation 3 that $G_3 = m_0^T(\rho_0) \otimes M_0^T$. Then, combining (A.14), (A.15), into (A.13) results in

$$M^T(\rho_0, \tilde{\theta}_n)\tilde{I}_n^{-1}m_0(\rho_0) = \tilde{G}_{2n}(\tilde{\theta}_n - \theta_0) + G_3 \text{vec}(\tilde{I}_n^{-1} - I_0^{-1}) \quad (A.16)$$

substituting (A.16) in (A.12) produces

$$(\hat{W}_n + \tilde{G}_{jn})\sqrt{n}(\hat{\rho}_n - \rho_0) = -(\hat{M}_n^T\hat{I}_n^{-1}\hat{N}_n + \tilde{G}_{2n})\sqrt{n}(\hat{\theta}_n - \theta_0) - G_3 \sqrt{n}\text{vec}(\tilde{I}_n^{-1} - I_0^{-1}) \quad (A.17)$$

the almost sure limit of $\hat{M}_n^T\hat{I}_n^{-1}\hat{N}_n + \tilde{G}_{2n}$ is $G_4$ given in Notation 3. Hence, from Lemma 2 the right hand side of (A.17) converges in law to a $N(0, \Omega)$ distribution, $\Omega$ is also defined in Notation 3.

It is now necessary to consider the behavior of $\hat{W}_n + \tilde{G}_{jn}$. This sequence of matrices converges with probability one to $W_0 + G_1$, which, by virtue of Corollary 30, Chapter 4 of Dhrymes (1984) is equal to $\frac{1}{2} (\frac{\partial^2}{\partial \rho^T \partial \rho})s_0(\rho_0)$. This is the Hessian of $s_0$ which, from Assumption 3, is positive definite at $\rho_0$ since it is a minimum. Then from (A.17) and Slutsky’s Theorem $\sqrt{n}(\hat{\rho}_n - \rho_0)$ converges in law to $N(0, V_{\rho})$, where $V_{\rho}$ is given in Notation 3.

It will be useful to notice that if we let $\tilde{G}_{sn} = (\hat{W}_n + \tilde{G}_{jn})^{-1}$, then from (A.17) the difference $\hat{\rho}_n - \rho_0$ may be expressed as

$$\hat{\rho}_n - \rho_0 = -\tilde{G}_{sn}(\hat{M}_n^T\hat{I}_n^{-1}\hat{N}_n + \tilde{G}_{2n})(\hat{\theta}_n - \theta_0) - \tilde{G}_{sn}G_3 \text{vec}(\tilde{I}_n^{-1} - I_0^{-1}) \quad (A.18)$$

Proof of Theorem 5. Substitution of (11) into (10) results in
\[ m_n(\hat{\rho}_n) - m_0(\rho_0) = \tilde{N}_n(\tilde{\theta}_n - \theta_0) - M_n \tilde{G}_{5n}(\tilde{M}_n^T \tilde{I}_n^{-1} \tilde{N}_n + \tilde{G}_2) (\tilde{\theta}_n - \theta_0) - M_n \tilde{G}_{5n} G_3 \text{vec}(\tilde{I}_n^{-1} - I_0^{-1}) \] (A.19)

In matrix notation (A.19) becomes

\[ m_n(\hat{\rho}_n) - m_0(\rho_0) = \left[ \tilde{N}_n - M_n \tilde{G}_{5n}(\tilde{M}_n^T \tilde{I}_n^{-1} \tilde{N}_n + \tilde{G}_2) \right] - M_n \tilde{G}_{5n} G_3 \left[ (\tilde{\theta}_n - \theta_0) \right] \text{vec}(\tilde{I}_n^{-1} - I_0^{-1}) \] (A.20)

does the matrix

\[ \left[ \tilde{N}_n - M_n \tilde{G}_{5n}(\tilde{M}_n^T \tilde{I}_n^{-1} \tilde{N}_n + \tilde{G}_2) \right] - M_n \tilde{G}_{5n} G_3 \]

converges with probability one to matrix \( A \) given in Notation 4, and the result follows from Lemma 2 immediately.

**Proof of Theorem 6.**

a) Consider

\[ \sqrt{n} m_n(\hat{\rho}_n) = \sqrt{n} \left[ m_n(\hat{\rho}_n) - m_0(\rho_0) \right] + \sqrt{n} m_0(\rho_0) \] (A.21)

the first term of (A.21) is bounded in probability, and \( \left\| \sqrt{n} m_0(\rho_0) \right\| \to \infty \), hence

\[ \left\| \sqrt{n} m_n(\hat{\rho}_n) \right\| \overset{p}{\to} \infty . \]

b) Let \( Z_n = \sqrt{n} m_n(\hat{\rho}_n) \), then from Theorem 5 and since \( m_0(\rho_0) = 0 \)

\[ Z_n = \sqrt{n} m_n(\hat{\rho}_n) \overset{L}{\to} X \approx N(0, \Sigma_m) \]

But \( \tilde{I}_n^{-1} \overset{a.s.}{\to} I_0^{-1} \), then

\[ L_0 = Z_n^T \tilde{I}_n^{-1} Z_n \overset{L}{\to} X^T I_0^{-1} X \]
Proof of Lemma 3.

a) Consider \( A_0 = M_0 W_0^{-1} M_0^T I_0^{-1} \), then \( U_0 = I - A_0 \) and

\[
A_0 A_0 = M_0 W_0^{-1} M_0^T I_0^{-1} M_0 W_0^{-1} M_0^T I_0^{-1} = M_0 W_0^{-1} W_0^{-1} M_0^T I_0^{-1} = A_0
\]

Hence \( A_0 \) is idempotent and it is a well known fact that in that case \( U_0 \) is also idempotent.

b) Write \( A_0 = B_0 I_0^{-1} \) with \( B_0 = M_0 W_0^{-1} M_0^T \). Observe that \( B_0 \) is symmetric and

\[
A_0^T = I_0^{-1} B_0. \text{ Then}
\]

\[
U_0^T I_0^{-1} U_0 = (I - I_0^{-1} B_0)(I_0^{-1} - I_0^{-1} B_0 I_0^{-1}) = I_0^{-1} - 2I_0^{-1} B_0 I_0^{-1} + I_0^{-1} B_0 I_0^{-1} = (A.22)
\]

From (A.22)

\[
U_0^T I_0^{-1} U_0 = I_0^{-1} - 2I_0^{-1} A_0 + I_0^{-1} A_0 A_0
\]

\[
= I_0^{-1} - I_0^{-1} A_0 = I_0^{-1} U_0.
\]

Also from (A.22)

\[
U_0^T I_0^{-1} U_0 = I_0^{-1} - 2A_0^T I_0^{-1} + A_0^T A_0 I_0^{-1}
\]

\[
= I_0^{-1} - A_0^T I_0^{-1} = U_0^T I_0^{-1}.
\]

Proof of Theorem 7. Consider \( X \approx N(\theta, \Sigma_0^m) \) as in Theorem 8, b) and \( Y = X^T I_0^{-1} X \). From the Theorem in section 3.5 of Serfling (1980) \( Y \) has a central chi-squared distribution if and only if

\[
\Sigma_0^m I_0^{-1} \Sigma_0^m I_0^{-1} \Sigma_0^m = \Sigma_0^m I_0^{-1} \Sigma_0^m
\]

(A.23)
in which case the degrees of freedom is \( \text{trace}(I_0^{-1} \Sigma_0^0) \), (A.23) is equivalent to
\[
U_0 N_0 V_0 U_0^T I_0^{-1} U_0 N_0 V_0 U_0^T I_0^{-1} U_0 N_0 V_0 U_0^T = U_0 N_0 V_0 U_0^T I_0^{-1} U_0 N_0 V_0 U_0^T
\]
applying Lemma 3 paragraph b) simplifies the above expression to (12).

\[\square\]

**Proof of Corollary 1.** If (13) is true then pre-multiplication on both sides by \( \Sigma_0^0 I_0^{-1} N_0 V_0 N_0 \) gives (12). Recall \( A_0 \) from the proof of Lemma 3, a), and since (13) holds true then
\[
\text{trace}(I_0^{-1} \Sigma_0^0) = \text{trace}(U_0^T) = \text{trace}(I - A_0^T) = \text{dim}(\theta) - \text{trace}(A_0)
\]
from Lemma 3 \( A_0 \) is idempotent and hence
\[
\text{trace}(A_0) = \text{rank}(A_0) = \text{rank}(M_0) = \text{dim}(\rho),
\]
since from Assumption 3 \( M_0 \) is full column rank.

\[\square\]

**Proof of Lemma 4.** Let \( \epsilon, \delta > 0 \) be arbitrarily small quantities and let \( Z_i = \log f(X_i | \theta) \) and \( Z_i^* = \log f(X_i^* | \theta) \). Finally, let \( (M_{1,n}, M_{2,n}, \ldots, M_{n,n}) \) be a multinomial distribution such that \( M_{t,n} \) equals the number of times that \( Z_t \) is selected in the bootstrap sample. Then
\[
\frac{1}{n} \sum_{i=1}^{n} Z_i^* = \frac{1}{n} \sum_{i=1}^{n} Z_i M_{t,n}.
\]
Let \( k \) be a fixed quantity (depending on \( \epsilon, \delta \) and eventually on \( \theta \)) such that
\[
E(\mid Z_i \mid 1(\mid Z_i \mid > k)) < \frac{\epsilon \delta}{2},
\]
and define \( Z_{i,1} = Z_i I(\mid Z_i \mid > k) \) and \( Z_{i,2} = Z_i I(\mid Z_i \mid \leq k) \).

Then

29
\[ \frac{1}{n} \sum_{i=1}^{n} Z_i^* - \frac{1}{n} \sum_{i=1}^{n} Z_i = \frac{1}{n} \sum_{i=1}^{n} Z_{i,1} [M_{t,n} - 1] + \frac{1}{n} \sum_{i=1}^{n} Z_{i,2} [M_{t,n} - 1]. \]

Thus, if we show that both addends on the right hand side vanish in probability for almost all sample sequences \( \{Z_i\} \), the result follows. Note that when considered isolated, \( M_{t,n} \) is a binomial random variable with parameters \((n, 1/n)\). Therefore, \( E(M_{t,n}) = 1 \) and \( V(M_{t,n}) = \frac{n-1}{n} \). Hence concerning the second addend,

\[ E \left( \frac{1}{n} \sum_{i=1}^{n} Z_{i,2} [M_{t,n} - 1] \Bigg| Z_1, ..., Z_n \right) = 0 \]

and

\[ V \left( \frac{1}{n} \sum_{i=1}^{n} Z_{i,2} [M_{t,n} - 1] \Bigg| Z_1, ..., Z_n \right) \]

\[ = \frac{1}{n^2} \sum_{i=1}^{n} Z_{i,2}^2 V[M_{t,n} - 1] + \frac{2}{n^2} \sum_{1 \leq i < j \leq n} Z_{i,2} Z_{j,2} E\left( [M_{t,n} - 1][M_{t,n} - 1]\right) \]

\[ \leq \frac{k^2}{n} + \frac{2k^2}{n^2} \sum_{1 \leq i < j \leq n} E\left( [M_{t,n} - 1][M_{t,n} - 1]\right) \]

\[ \leq \frac{k^2}{n} + \frac{2k^2}{n^2} \left( \frac{n(n-1)}{2} \right) \to 0. \]

Therefore, by Tchebychev inequality

\[ \lim_{n \to \infty} P \left[ \left| \frac{1}{n} \sum_{i=1}^{n} Z_{i,2} [M_{t,n} - 1] \right| > \delta \Bigg| Z_1, ..., Z_n \right] = 0 \]

for all sequences \( \{Z_i\} \). Concerning the first term, we have that

\[ E \left( \frac{1}{n} \sum_{i=1}^{n} Z_{i,1} [M_{t,n} - 1] \Bigg| Z_1, ..., Z_n \right) \leq \frac{1}{n} \sum_{i=1}^{n} |Z_{i,1}| [E(M_{t,n}) + 1] \]
\[ = 2 \frac{1}{n} \sum_{i=1}^{n} |Z_{t,i}| \]

and again, by Tchebychev inequality

\[ P \left[ \frac{1}{n} \sum_{i=1}^{n} |Z_{t,i} (M_{i,n} - 1)| > \delta \right] \leq 2 \frac{1}{n \delta} \sum_{i=1}^{n} |Z_{t,i}| < \varepsilon \text{ a.s.} \]

for \( n \) large enough because \( \frac{1}{n} \sum_{i=1}^{n} |Z_{t,i}| \xrightarrow{a.s.} E(|Z| \mid I(|Z| > k)) < \frac{\varepsilon \delta}{2}. \]

\[ \blacksquare \]

**Proof of Lemma 5.** Since \( E(\sup_{\theta \in \Theta} \log f(X \mid \theta)) < \infty \), the dominated convergence theorem allow us to write that, \( \forall \theta \in \Theta \)

\[ \lim_{k \to 0} E \left( \sup_{(\theta, \theta') \in \Theta} \left| \log f(X \mid \theta) - \log f(X \mid \theta') \right| \right) = 0 \]

and therefore, \( \forall \varepsilon > 0, \theta \in \Theta \exists k_{\varepsilon, \theta} > 0 \) such that

\[ E \left( \sup_{(\theta, \theta') \in \Theta \cap (k_{\varepsilon, \theta})} \left| \log f(X \mid \theta) - \log f(X \mid \theta') \right| \right) < \varepsilon \]

Note also that if \( B(\theta, k) \) is the ball centered at \( \theta \) and with radius \( k \), then \( \{B(\theta, k_{\varepsilon, \theta}) : \theta \in \Theta \} \) is an open covering of \( \Theta \) and by compactness, there exists a finite open subcovering \( \{B(\theta_j, k_{\varepsilon, \theta_j}) : j = 1,2,\ldots,m \} \). Since

\[ \left| l'_n(\theta) - l_n(\theta) \right| \leq \left| l'_n(\theta_j) - l'_n(\theta_j) \right| + \left| l'_n(\theta_j) - l_n(\theta_j) \right| + \left| l_n(\theta_j) - l_n(\theta) \right| \]

and if we denote \( B_j = B(\theta_j, k_{\varepsilon, \theta_j}) \cap \Theta \), we have that

\[ \sup_{\theta \in \Theta} \left| l'_n(\theta_j) - l_n(\theta) \right| = \max_{j=1,2,\ldots,m} \sup_{\theta \in B_j} \left| l'_n(\theta) - l_n(\theta) \right| \]
\[
\leq \max_{j=1,2,...,m} \sup_{\theta \in B_j} \left| l_n^*(\theta) - l_n^*(\theta_j) \right| + \max_{j=1,2,...,m} \left| l_n^*(\theta) - l_n(\theta_j) \right| + \max_{j=1,2,...,m} \left| l_n(\theta) - l_n(\theta_j) \right|
\]

\[
\leq \max_{j=1,2,...,m} \sup_{\theta \in B_j} \left| l_n^*(\theta) - l_n^*(\theta_j) \right| + \max_{j=1,2,...,m} \left| l_n^*(\theta) - l_n(\theta_j) \right| + \max_{j=1,2,...,m} \sup_{\theta \in B_j} \left| l_n(\theta) - l_n(\theta_j) \right|
\]

Now, \( \max_{j=1,2,...,m} \left| l_n^*(\theta_j) - l_n(\theta) \right| \xrightarrow{p} 0 \) a.s. because of the pointwise convergence proved just above. In addition, since

\[
\max_{j=1,2,...,m} \sup_{\theta \in B_j} \left| l_n(\theta) - l_n(\theta_j) \right| \leq \max_{j=1,2,...,m} \frac{1}{n} \sum_{t=1}^{n} \sup_{\theta \in B_j} \left| f(X|\theta_j) - f(X|\theta) \right|
\]

and

\[
\max_{j=1,2,...,m} \frac{1}{n} \sum_{t=1}^{n} \sup_{\theta \in B_j} \left| f(X|\theta_j) - f(X|\theta) \right| \xrightarrow{a.s.} \max_{\theta \in \Theta} \left( \sup_{\theta \in B_j} \left| f(X|\theta_j) - f(X|\theta) \right| \right)
\]

for large \( n \) we have that

\[
\max_{j=1,2,...,m} \sup_{\theta \in B_j} \left| l_n(\theta_j) - l_n(\theta) \right| < \varepsilon / 2 \quad \text{a.s.}
\]

Finally, note that

\[
E \left( \sup_{\theta \in \Theta} \left| l_n(\theta) - l_n^*(\theta) \right| \right) \cap < Z_1,\ldots,Z_n \right)
\]

\[
\leq \frac{1}{n} \sum_{t=1}^{n} E \left( \sup_{\theta \in \Theta} \left| f(X^*_i|\theta) - f(X^*_i|\theta_j) \right| \right) Z_1,\ldots,Z_n \right)
\]

\[
= \frac{1}{n} \sum_{t=1}^{n} E \left( \sup_{\theta \in \Theta} \left| f(X_i|\theta) - f(X_i|\theta_j) \right| M_{t,n} Z_1,\ldots,Z_n \right)
\]

\[
= \frac{1}{n} \sum_{t=1}^{n} \sup_{\theta \in \Theta} \left| f(X_i|\theta) - f(X_i|\theta_j) \right| E(M_{t,n}) \n
\leq \varepsilon / 2 \quad \text{a.s.}
\]
for $n$ large. Therefore the result holds.

\[\square\]

**Proof of Corollary 4.** Consider $(C(\Theta),d_n)$ the space of continuous functions on $\Theta$ endowed with the sup metric. In the previous lemmas it was shown that for almost all samples, $l_n^*, l_n \in C(\Theta)$ and that $d_n(l_n^*, l_n) \xrightarrow{p} 0$ a.s. Since arg max is a continuous function in this space, the result follows applying the continuous mapping theorem.

\[\square\]

**Proof of Theorem 8.** We assume without any loss of generality that $\Theta$ is a subset of the real line. Applying the mean value theorem to the first order condition solved for the artificial sample we obtain

\[
0 = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial}{\partial \theta} f(X_t^* | \tilde{\theta}_n^*)
\]

\[
= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( \frac{\partial^2}{\partial \theta^2} f(X_t^* | \tilde{\theta}_n^*) + \left( \frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2}{\partial \theta^2} f(X_t^* | \tilde{\theta}_n^*) \right) \right) \sqrt{n} (\tilde{\theta}_n^* - \tilde{\theta}_n^*),
\]

where $\tilde{\theta}_n^*$ is a point between $\tilde{\theta}_n^*$ and $\tilde{\theta}_n^*$. Therefore

\[
\sqrt{n} (\tilde{\theta}_n^* - \tilde{\theta}_n^*) = \left( \frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2}{\partial \theta^2} f(X_t^* | \tilde{\theta}_n^*) \right) \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial}{\partial \theta} f(X_t^* | \tilde{\theta}_n^*).
\]

Following steps similar to those in the proof of the previous lemmas it can be easily shown that

\[
\frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2}{\partial \theta^2} f(X_t^* | \tilde{\theta}_n^*) \xrightarrow{p} \frac{1}{n} \sum_{t=1}^{n} (\partial / \partial \theta) f(X_t^* | \tilde{\theta}_n^*) \quad \text{a.s.} \quad (A.24)
\]

Consider $(\partial / \partial \theta) f(X_t^* | \tilde{\theta}_n^*)$ and notice that

\[
E \left[ (\partial / \partial \theta) f(X_t^* | \tilde{\theta}_n^*) \mid Z_1, \ldots, Z_n \right] = \frac{1}{n} \sum_{t=1}^{n} (\partial / \partial \theta) f(X_t^* | \tilde{\theta}_n^*) = 0,
\]
given the first order conditions of the original problem. It is also important to note that

\[ s_n^2 = E\left\{ \left( \frac{\partial}{\partial \theta} f(X_t^* | \tilde{\theta}_n) \right)^2 \right\} \approx \frac{1}{n} \sum_{t=1}^{n} \left( \frac{\partial}{\partial \theta} f(X_t | \tilde{\theta}_n) \right)^2 \]

which is a positive finite quantity. Therefore, for any \( \delta > 0 \) there is a finite \( k \) such that, for large \( n \)

\[ E\left\{ \left( \frac{\partial}{\partial \theta} f(X_t^* | \tilde{\theta}_n) \right)^2 \right\} \approx \frac{1}{n} \sum_{t=1}^{n} \left( \frac{\partial}{\partial \theta} f(X_t | \tilde{\theta}_n) \right)^2 > k \] \( Z_t = Z_{n,t} \) a.s.

Thus, for all \( \varepsilon > 0 \) and \( n \) large enough, \( n^{1/2} \varepsilon \geq k \) and then

\[ \sum_{i=1}^{n} \frac{1}{s_n^2} E\left\{ \left( n^{-1/2} \frac{\partial}{\partial \theta} f(X_t^* | \tilde{\theta}_n) \right)^2 \right\} \left\{ \left( n^{-1/2} \frac{\partial}{\partial \theta} f(X_t^* | \tilde{\theta}_n) \right) > \varepsilon \right\} < \delta \text{ a.s.} \]

and this means that the triangular array \( \left\{ n^{-1/2} \frac{\partial}{\partial \theta} f(X_t^* | \tilde{\theta}_n) : t = 1,2,\ldots,n \right\} \) is independent conditionally on the sample by construction and also that the Linderberg condition holds for almost all samples. Apply the central limit theorem to obtain the result.

■

**Proof of Lemma 6.** We prove the first convergence property. The second can be obtained similarly. Use the dominated convergence theorem to show that \( m(\rho, \theta) \) is continuous in \( \theta \) (uniformly since \( \Theta \) is compact and in each of its coordinates). Thus, for any \( \varepsilon > 0 \) there is some \( \delta(\varepsilon) \) such that if \( \delta < \delta(\varepsilon) \) we have that

\[ \sup_{\theta \in \Theta} \left\| \sup_{\rho \in \rho_n} \left| m(\rho, \theta) - m(\rho, \theta') \right| \right\| < \varepsilon. \]

Hence
The first addend is zero for the choice of $\delta$. The second converges to zero almost surely as a consequence of COROLLARY 4.

**Proof of Corollary 5.** Notice that

$$\sup_{\rho \in \mathbb{R}} \left\| m_n^T(\rho) \bar{I}_n^{-1} m_n(\rho) - m_0^T(\rho) \bar{I}_n^{-1} m_0(\rho) \right\|$$

$$= \sup_{\rho \in \mathbb{R}} \left\| m_n^T(\rho) - m_0^T(\rho) \right\| \bar{I}_n^{-1} \left\| m_n(\rho) + m_0(\rho) \right\|$$

$$\leq \sup_{\rho \in \mathbb{R}} \left\| m_n^T(\rho) - m_0^T(\rho) \right\| \bar{I}_n^{-1} \sup_{\rho \in \mathbb{R}} \left\| m_n(\rho) + m_0(\rho) \right\|$$

Since $\sup_{\rho \in \mathbb{R}} \left\| m_n^T(\rho) - m_0^T(\rho) \right\| \xrightarrow{a.s.} 0$, $\sup_{\rho \in \mathbb{R}} \left\| m_n(\rho) + m_0(\rho) \right\| \xrightarrow{a.s.} 2 \sup_{\rho \in \mathbb{R}} \left\| m_0(\rho) \right\| < \infty,$

and $\bar{I}_n \xrightarrow{a.s.} I_0$, then the result b) follows by usual Slutsky’s theorem.

On the other hand

$$\sup_{\rho \in \mathbb{R}} \left| s_n^*(\rho) - s_n(\rho) \right| \leq \sup_{\rho \in \mathbb{R}} \left\| m_n^*(\rho) \right\| \left\| (\bar{I}_n)^{-1} - \bar{I}_n^{-1} \right\|$$

$$+ \sup_{\rho \in \mathbb{R}} \left\| m_n^*(\rho) - m_n(\rho) \right\| \left\| \bar{I}_n^{-1} \right\| \sup_{\rho \in \mathbb{R}} \left\| m_n^*(\rho) + m_n(\rho) \right\|.$$
Noting that \( \sup_{\rho \in \mathbb{R}} \| m_n^*(\rho) - m_n(\rho) \|_{p^*} \to 0 \) a.s. as a result of the previous lemma, and that \( \| (\tilde{t}_n^*)^{-1} - \tilde{t}_n^{-1} \|_{p^*} \to 0 \) a.s. as a result of (A.24), then we have that
\[
\sup_{\rho \in \mathbb{R}} \| m_n^*(\rho) + m_n(\rho) \|_{p^*} \to 2 \sup_{\rho \in \mathbb{R}} \| m_\theta(\rho) \| < \infty \text{ a.s.}
\]

Then, the result a) also follows using the Cramér-Wold device.


Proof of Lemma 7. To simplify notation, consider \( \Theta \) to be a subset of the real line. First, apply the Linderberg-Feller central limit Theorem as in the proof of Theorem 8 to show that
\[
\sqrt{n} \sum_{i=1}^{n} \left( \frac{\partial}{\partial \theta} f(X_i^* | \tilde{\theta}_n) \right) - \left[ \frac{\partial}{\partial \theta} f(X_i | \tilde{\theta}_n) \right] \xrightarrow{p^*} \mathcal{N}(0, \Gamma_0) \text{ a.s.}
\]
for some variance covariance matrix \( W \). Now notice that, from the proof of Theorem 8,
\[
\sqrt{n}(\tilde{\theta}_n^* - \tilde{\theta}_n) = - \left[ E \left( \frac{\partial^2}{\partial \theta^2} f(X | \theta_0) \right) \right]^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left( \frac{\partial}{\partial \theta} f(X_i | \tilde{\theta}_n) \right) + O_n^*
\]
where \( O_n^* \xrightarrow{p^*} 0 \) a.s. Analogously, following steps similar to those in the proof of the previous lemmas it can be easily shown that, for some \( \tilde{\theta}_n^* \) such that \( |\tilde{\theta}_n^* - \tilde{\theta}_n| \leq |\tilde{\theta}_n^* - \tilde{\theta}_n| \)
and
\[
\sqrt{n}(\tilde{t}_n^* - \tilde{t}_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ \left( \frac{\partial}{\partial \theta} f(X_i^* | \tilde{\theta}_n) \right) \right]^2 - \left[ \left( \frac{\partial}{\partial \theta} f(X_i | \tilde{\theta}_n) \right) \right]^2
\]
It is not hard to show that
\[
\left[ \frac{1}{n} \sum_{i=1}^{n} 2(\partial / \partial \theta) f(X_i^* | \tilde{\theta}_n) (\partial^2 / \partial \theta^2) f(X_i^* | \tilde{\theta}_n) \right] \sqrt{n} (\tilde{\theta}_n^* - \tilde{\theta}_n).
\]

where, in this context, \( R_n = E \left[ 2(\partial / \partial \theta) f(X | \theta_o) (\partial^2 / \partial \theta^2) f(X | \theta_o) \right] \). Therefore, using the Cramér-Wold device, we have that
\[
\begin{bmatrix} \frac{\sqrt{n}(\tilde{\theta}_n^* - \tilde{\theta}_n)}{\sqrt{n}(\tilde{t}_n^* - \tilde{t}_n)} \end{bmatrix} \rightarrow^{d} \begin{bmatrix} -J_{\theta_o'}^{-1} \ \ 0 \\ -R_oJ_{\theta_o'}^{-1} \ \ 1 \end{bmatrix} N(0, \Gamma_o) \quad a.s.
\]
\[N \begin{bmatrix} (0) \\ (0) \end{bmatrix} \begin{bmatrix} V_o & V_{o1} \\ V_{o1}^* & V_I \end{bmatrix} \] a.s.

Apply the \( \delta \)-method to obtain the result.

---

**Proof of Theorem 9.** Without any loss, we suppose that \( \Theta, \Re \) are subsets of the real line. From Notation 2, let \( \hat{M}_n^* = M(\hat{\rho}_n^*, \hat{\theta}_n^*) \) and \( \bar{M}_n^* = M(\bar{\rho}_n^*, \bar{\theta}_n^*) \) respectively, where \( \bar{\rho}_n^* \) is such that \( |\bar{\rho}_n^* - \hat{\rho}_n^*| \leq |\hat{\rho}_n^* - \hat{\rho}_n| \). Use the first order conditions as in the proof of Theorem 4 to obtain that
\[
0 = \hat{M}_n^* (\tilde{r}_n^*)^\dagger m_n^*(\hat{\rho}_n^*)
\]
\[
= \hat{M}_n^* (\tilde{r}_n^*)^\dagger m_n^*(\hat{\rho}_n^*) + \bar{M}_n^* (\tilde{r}_n^*)^\dagger \bar{M}_n^*(\hat{\rho}_n^* - \hat{\rho}_n)
\]
and therefore, if \( \tilde{N}_n^* = (\partial / \partial \theta) m(\hat{\rho}_n^*, \tilde{r}_n^*) \) and \( \tilde{\theta}_n^* \) being a point between \( \hat{\theta}_n \) and \( \hat{\theta}_n \) we write,
\[
\hat{M}_n^* (\tilde{r}_n^*)^\dagger \bar{M}_n^* \sqrt{n}(\hat{\rho}_n^* - \hat{\rho}_n)
\]
\[
= -\hat{M}_n^* (\tilde{r}_n^*)^\dagger \sqrt{n} [m_n^*(\hat{\rho}_n^*)-m_n^*(\hat{\rho}_n)] - \sqrt{n}\bar{M}_n^* (\tilde{r}_n^*)^\dagger m_n^*(\hat{\rho}_n)
\]
\[- \dot{M}_n^* (\hat{I}_n^*)^t \hat{N}_n \sqrt{n(\hat{\theta}_n^* - \hat{\theta}_n^*)} - \left[ (\partial / \partial \rho) M(\hat{\rho}_n^*, \hat{\theta}_n^*) \sqrt{n(\hat{\rho}_n^* - \hat{\rho}_n^*)} \right] (\hat{I}_n^*)^t m_n(\hat{\rho}_n^*) \]

\[- \left[ (\partial / \partial \theta) M(\hat{\rho}_n^*, \hat{\theta}_n^*) \sqrt{n(\hat{\theta}_n^* - \hat{\theta}_n^*)} \right] (\hat{I}_n^*)^t m_n(\hat{\rho}_n^*) - \sqrt{n} M(\hat{\rho}_n^*, \hat{\theta}_n^*) (\hat{I}_n^*)^t (\hat{I}_n^*)^t m_n(\hat{\rho}_n^*) \]

The last term equals zero for all \( n \) using first order conditions of \( \hat{\rho}_n \). Hence

\[
\left[ \dot{M}_n^* (\hat{I}_n^*)^t \hat{N}_n + (\partial / \partial \rho) M(\hat{\rho}_n^*, \hat{\theta}_n^*) (\hat{I}_n^*)^t m_n(\hat{\rho}_n^*) \right] \sqrt{n(\hat{\rho}_n^* - \hat{\rho}_n^*)}
\]

\[
= - \left[ \dot{M}_n^* (\hat{I}_n^*)^t \hat{N}_n + (\partial / \partial \theta) M(\hat{\rho}_n^*, \hat{\theta}_n^*) (\hat{I}_n^*)^t m_n(\hat{\rho}_n^*) \right] \sqrt{n(\hat{\theta}_n^* - \hat{\theta}_n^*)}
\]

\[- M(\hat{\rho}_n^*, \hat{\theta}_n^*) \sqrt{n} \left( (\hat{I}_n^*)^t - (\hat{I}_n^*)^t \right) m_n(\hat{\rho}_n^*) \]

and the result follows using procedures analogous of the previous lemmas, lemma 6 and

the continuous mapping theorem.

\[ \blacksquare \]

**Proof of Corollary 7.** For the sake of brevity, denote \( W_n^* = w\left( \sqrt{n}(\hat{\rho}_n^* - \hat{\rho}_n^*) \right) \),

\( W_n = w\left( \sqrt{n}(\hat{\rho}_n - \rho_0) \right) \), and \( W_\infty = w\left( N(0, V_\rho) \right) \). Therefore, given that \( W_n \xrightarrow{\text{a.s.}} W_\infty \) and

that \( W_n^* \xrightarrow{\text{a.s.}} W_\infty \), we have to show that

\[
\lim_{n \to \infty} P\left[ W_n \geq F_{W_\infty}^{-1}(\alpha) \right] = 1 - \alpha.
\]

This statement is equivalent to (see Serfling, 1980, p. 3)
\[ \lim_{n \to \infty} P \left[ F_{W_n}(W_n) \geq \alpha \right] = 1 - \alpha. \]

In addition, since \( F_{W_n} \to F_{W_n} \) for almost all samples, we have that for any \( \epsilon, \delta > 0 \) there is some \( n_{\epsilon, \delta} \) such that if \( n > n_{\epsilon, \delta} \), \( P \left( W_n : \sup_x |F_{W_n}(x) - F_{W_n}(x)| > \epsilon \right) < \delta \). Thus for any \( n > n_{\epsilon, \delta} \)

\[ P \left[ F_{W_n}(W_n) \geq \alpha + \epsilon \right] - \delta \leq P \left[ F_{W_n}(W_n) \geq \alpha \right] \leq P \left[ F_{W_n}(W_n) \geq \alpha - \epsilon \right] + \delta. \]

Since \( F_{W_n} \) is continuous, \( F_{W_n}(W_n) \xrightarrow{d} F_{W_n}(W_n) \) that has a uniform distribution if \( F \) is continuous. Therefore

\[ 1 - \alpha - \epsilon - \delta = \lim_{n \to \infty} P \left[ F_{W_n}(W_n) \geq \alpha + \epsilon \right] - \delta \]

\[ \leq \lim_{n \to \infty} P \left[ F_{W_n}(W_n) \geq \alpha \right] \]

\[ \leq \lim_{n \to \infty} P \left[ F_{W_n}(W_n) \geq \alpha - \epsilon \right] + \delta \]

\[ = 1 - \alpha + \epsilon + \delta. \]

Since \( \epsilon, \delta \) were arbitrary quantities, the result follows.

\[ \blacksquare \]

**REFERENCES**


