Fac. CC. Económicas y Empresariales
Universidad de La Laguna

Fac. CC. Económicas y Empresariales
Univ. de Las Palmas de Gran Canaria

A note on the asymptotic efficiency of the restricted estimation

José Antonio Hernández *

DOCUMENTO DE TRABAJO 2005-01

* Universidad de Las Palmas de G.C. Departamento de Análisis Económico Aplicado.

A note on the asymptotic efficiency of the restricted estimation*

José A. Hernández†

University of Las Palmas de Gran Canaria

Spain

Abstract

This paper provides a unified framework for the analysis of the stochastic and deterministic constrained estimation. In a general framework it is show that stochastic restrictions method estimates can be asymptotically more efficient than estimates ignoring prior information, and can achieve efficiency of the restricted estimate if prior information grows faster than the sample information in the asymptotics. As an example of the applicability of the previous result, the maximum likelihood stochastically restricted criterion is provided.

Keywords: Prior information, stochastic restrictions, efficiency, maximum likelihood.

JEL classification code: C11, C13, C15

*Author aknowledges the financial support provided by the grant PI2003/042 (Dirección General de Universidades, Gobierno de Canarias).

†University of Las Palmas de Gran Canaria. Facultad de CC. EE y EE. Despacho D312. Campus de Tafira. 35017 Las Palmas de Gran Canaria. jhernandez@daea.ulpgc.es. Tel. 0034928458206; Fax 0034928458183.

1 Introduction

Economic theory provides parameters restrictions for some models, which generates more efficient estimates if restrictions are taken into account within the estimation process. A different type of information can also be available, coming from experience, that is, previous estimations of similar models or different samples. This is called *prior information,* and can be seen as an approximation to the value of the parameter. Including prior information as deterministic restrictions will bring non-consistent estimators, but ignoring it is wasting the possibility of disposing efficiency gains if the priors informs about the true values of parameters. An intermediate solution between ignoring prior information and including it as deterministic, is taking it into account with uncertainty, given by the perception of the quality of priors, i.e., the perception that the researches has about its likelihood of being true. This is the idea behind stochastic restrictions approach, which formally consists on a set of equations that contain restrictions on parameters and an error term whose variance captures uncertainty about priors. In other words, if the quality of priors is poor (high), the researcher could include a stochastic restrictions with a high (low) variance error term.

For finite samples, taking into account stochastic restrictions bring efficiency gains, as show Theil and Goldberger (1961) and Shiller (1973) for a linear model and normal errors, depending inversely on the variance of the error term of the stochastic restriction. Nevertheless, stochastic restrictions seem not to have much impact in the classic econometric literature, possibly due to its asymptotic irrelevance. As sample size increases, stochastic restriction weights decreases and efficiency gains disappears in the limit. Some question of interest derives from this result as can be seen. The main goal of asymptotic theory is to provide finite sample approximated distributions of estimators. If available prior information is correct, finite sample results show that stochastic restrictions are useful, then it follows that asymptotic result about it are not! Therefore, although theoretically correct, the resulting asymptotic irrelevance of stochastic restrictions is not satisfactory, since it says very little about the finite sample distribution of the restricted estimator. Also it would be interesting to provide a more satisfactory asymptotic result, according with expected relevance of the asymptotic restrictions. The aim of this paper is to answer *yes* to that question by providing a unified framework to analyze the asymptotic properties of the general restricted estimator. In a general set up it is show that stochastic restrictions method estimates can be asymptotically more efficient than estimates ignoring prior information, and can achieve efficiency of the restricted estimate if prior information grows faster than the sample information in the asymptotics. As an example of the applicability of the previous result, the maximum likelihood stochastically restricted criterion is provided. The relevance of the application lies on the fact that the maximum likelihood estimator properties are asymptotics by construction, and standard asymptotic theory reject the usefulness of the stochastic restrictions.

The structure of the paper is the following: In section 2, the main assumption of the paper is introduced and motivated. Section 3 proposes the general restricted estimator and shows the theoretical results. Finally, section 4 provides an example of applicability of the previous section results through the maximum likelihood estimator.

2 Priors's quality improves asymptotically

In this section the stochastic restriction approach is described in the framework of a standard general linear model (GLM) for a T periods sample size. This familiar framework makes the presentation easier to understand, though not required for the purposes. The considered model is given by:

$$Y = X\beta + u \tag{1}$$

where β, is a $k-$dimension parameter vector, X is a Txk matrix of regressors, Y, u are respectively $T-$ dimension endogenous variable vector and random disturbance, where $u \sim i.i.d.(0, \sigma_u^2 I_T)$ -a more complex structure for u is also allowed, but omitted to simply the analysis. Available prior information about β, modelled through the stochastic restriction approach, leads to the equation:

$$r = R\beta + v \tag{2}$$

where r is a $qx1$ vector $(q < k)$ of values that prior information allocates to a linear combination of β's and R is a qxk matrix of coefficients. The term v is a random vector assumed to be $v \sim i.i.d.(0, \sigma_v^2 I_q)$, where σ_v^2 is chosen according with the uncertainty about the prior information. Since v is random, also r is. It is also assumed that v is independent of u. Although it is not necessarily, neither realistic assuming that stochastic restrictions are independent and homoscedastic, it makes more direct the understanding the effects of stochastic restrictions on efficiency gains.

In order to shed some light in the understanding of what a stochastic restriction is, consider a Cobb-Douglas production function $Y = AK^{\alpha}L^{\beta}$ using standard macroeconomics notation. Assume that prior information is available about parameters, consisting in the believe that "Returns to scale are probably constant". It means that it is expected that $\alpha + \beta$ were close to one. In terms of a stochastic restriction, prior information can be put forward as $1 = (\alpha + \beta) + v$, where v is the error term capturing that information need not to be exactly true.

The main contribution of this paper relies on assumptions about the asymptotics of σ_v^2. More precisely, it is considered an asymptotically decreasing variance of the stochastic restrictions, or, in terms of Bekker (1994), a parameter sequence. As a result, the relative weights of prior and sample information are preserved in asymptotic terms, and as opposed to the standard approach, this explains efficiency gains.

This kind of assumption might be considered too strong and, as mentioned in Kadane (1971), difficult to justify. However, in the context of IV estimation with weak instruments, in Bekker (1994) and Staiger and Stock (1997) we find a similar assumption, justified by the goal of finding better approximations to the finite sample distribution of the estimator of interest. The approximation is derived mainly from standard asymptotic theory, but also, taking into account the extra assumption of a parameter sequence, designed to improve the properties of the considered estimator. Despite of the objection of Kadane (1971), Bekker (1994) claims that there is no need to make such "realistic" assumption and that "...the quality of the approximation is the only criterion for justifiability". In our context it could also be argued that the goal of the assumption is providing a distribution that fits the finite sample distribution better. But added to that, there is a realistic motivation for it. Considering that the priors are obtained from a sample whose size also increases asymptotically can be interpreted as the idea that experience can be improved, and this could be viewed as natural fact. Hence, informative priors in static terms, leads to improved quality priors with additional sampling.

3 A general approach to restricted estimation

A *general restricted* estimator is described in this section as the outcome of the estimation of a GLM model where priors are taken into account through the stochastic restrictions approach. The resulting estimator variance covariance matrix converges to the non-restricted or the exactly restricted estimators depending on the relative asymptotics of T and T^*.

Consider the standard GLM given in (1) and the stochastic restrictions in (2). The complete model can be seen as the system

$$\begin{bmatrix} Y \\ r \end{bmatrix} = \begin{bmatrix} X \\ R \end{bmatrix} \beta + \begin{bmatrix} u \\ v \end{bmatrix} \tag{3}$$

Calling $\bar{Y}' = [Y', r']$, $\bar{X}' = [X', R']$ and $\bar{u}' = [u', v']$. Since model (3) is heteroscedastic, to be efficiently estimated, it should be premultiplied by the transformation matrix P, with elements I_T/σ_u and I_q/σ_v in the main diagonal. The resulting transformed model is

$$y = x\beta + w \tag{4}$$

where $y = P\bar{Y}$, $x = P\bar{X}$, and $w = P\bar{u}$, with $E(w) = 0$, and $V(w) = I_{T+q}$.

Definition. The general restricted estimator of β, $\hat{\beta}_{GR}$ is the OLS estimator of model (4). That is,

$$\hat{\beta}_{GR} = (x'x)^{-1}x'y$$

It is immediate to check that $\hat{\beta}_{GR}$ is unbiased and that its variance is

$$V(\hat{\beta}_{GR}) = \left[\frac{X'X}{\sigma_u^2} + \frac{R'R}{\sigma_v^2}\right]^{-1} \tag{5}$$

which is smaller than $V(\hat{\beta}_{OLS})$.

Although, $\hat{\beta}_{GR}$ is not asymptotically more efficient than $\hat{\beta}_{OLS}$. Under standard assumptions (say SAA) of the GLM[1], asymptotic distribution of $\hat{\beta}_{GR}$ is

$$\sqrt{T}(\hat{\beta}_{GR}-\beta) \xrightarrow{d} N\left(0, p\lim\left(\frac{1}{T}\frac{X'X}{\sigma_u^2}+\frac{1}{T}\frac{R'R}{\sigma_v^2}\right)^{-1}\right)$$

and then, asymptotic variance is

$$AV(\hat{\beta}_{GR}) = p\lim\left(\frac{1}{T}\frac{X'X}{\sigma_u^2}+\frac{1}{T}\frac{R'R}{\sigma_v^2}\right)^{-1} \tag{6}$$

Since $p\lim\left(\frac{1}{T}\frac{R'R}{\sigma_v^2}\right) = 0$, equation (6) implies that $AV(\hat{\beta}_{GR}) = AV(\hat{\beta}_{OLS})$, which is not a very satisfactory result in empirical terms, as Lütkepohl (1991) points out. Far from it, it is reasonable to consider that if good quality priors are available, should be used, although they are established to be asymptotically irrelevant by the standard asymptotic setting. How to overcome this contradiction?. What follows is intended to find compatibility between theory and practice by providing a more general setting for the constrained estimation analysis.

Assumption A1. $V(v) \equiv \sigma_v^{*2} = \frac{\sigma_v^2}{T^*}$, i.e., the variance of the stochastic restriction decreases with T^*, the sample size of the sample that provides prior information.

The asymptotics is analyzed as T and T^* goes to infinite, and different growing rates are allowed for T and T^*. It should be noted that A1 implies that the term r depends on T^* and hence denoted as r^*. Since A1 states that prior information improves with the sample size T^*, in the limit r^* should

[1]SAA are $p\lim(\frac{x'x}{T})$ is a finite and inversible matrix, and $\frac{x'w}{\sqrt{T}} \rightarrow N(0, \Omega)$

equal $R\beta$ and be true. The following assumptions are in order for the theoretical discussion.

Assumption A2. $p\lim(T/T^*) = \infty$.

Assumption A3. $p\lim(T/T^*) = 1$.

Assumption A4. $p\lim(T/T^*) = 0$.

Next it is discussed the efficiency of $\hat{\beta}_{GR}$ in terms on the preceding alternatives assumptions, that is, on the dominating sample size, as shown in the following propositions.

Proposition 1. Under SAA, A1 and A2, $AV(\hat{\beta}_{GR}) = AV(\hat{\beta}_{OLS})$.

In this case, there are no efficiency gain, –as for the case were no increasing informative priors are considered– since T dominates T^*.

Proposition 2. Under SAA, A1 and A3, $AV(\hat{\beta}_{GR}) < AV(\hat{\beta}_{OLS})$.

Hence, if T and T^* increases at the same rate, general restricted estimation brings asymptotic efficiency gains, as opposed to standard asymptotic theory.

Proposition 3. Under SAA, A1, and A4, $AV(\hat{\beta}_{GR}) = AV(\hat{\beta}_R)$, where $\hat{\beta}_R$ is the constrained estimator of the model

$$\begin{aligned} Y &= X\beta + u \\ r &= R\beta \end{aligned}$$

Hence, if prior information increases more rapidly than variables information, priors are as relevant as if they were deterministic in the limit.

Proposition 4. Under A1, if $\lim(T/T^{*}) = M$, $0 < M < \infty$, then,

$$i) \quad AV(\hat{\beta}_{GR}) < (AV(\hat{\beta}_{OLS})$$

$$ii) \quad \frac{d(AV(\hat{\beta}_{GR}))}{dM} > 0$$

$$iii) \quad \lim_{M\to 0} AV(\hat{\beta}_{GR}) = AV(\hat{\beta}_{R})$$

$$iv) \quad \lim_{M\to \infty} AV(\hat{\beta}_{GR}) = AV(\hat{\beta}_{OLS})$$

This proposition states that $AV(\hat{\beta}_{GR})$ increases continuously from the variance of the unrestricted estimator to the variance of the deterministic restricted estimator as prior's quality decreases.

4 Example: maximum likelihood and prior information

In this section it is described an application of the main result of the previous section through the maximum likelihood (ML) estimation method. Since ML estimator properties are defined asymptotically, under standard approach there is no place for stochastic restrictions. Nevertheless, under assumptions A1 and A3 or A4 introduced in Section 3 it is possible to support efficiency gains into the ML estimation method.

First, the extension of the ML method to the stochastic restrictions approach is defined. Then, its distribution and the resulting efficiency gain is shown for the cases in which prior information is not dominated asymptoticaly by sample information.

It is considered a parametric model in which y_t, for $t = 1, ...T$, is the endogenous variable and x_t, $t = 1, ...T$ is the set of exogenous variables of the model. Let $f_0(y_1, ..., y_T/x_1, ..., x_T)$ be the conditioned distribution function of $y_1, ..., y_T$ given $x_1, ..., x_T$. If observations are independent, then the conditioned distribution function can be decomposed as $\prod_{t=1}^{T} f_0(y_t/x_1, ..., x_T)$. It is also assumed strong exogeneity on variable x, that is, $f_0(y_t/x_1,...,x_T) = f_0(y_t/x_1,...,x_t)$ and that there is a $p-$dimension parameter vector β_0 such that $f_0(y_t/x_t) = f(y_t/x_t;\beta_0)$. The likelihood function is defined as the conditional distribution of sample, which for independent variables equals the marginal densities of the sample. That is,

$$L_1(y,x;\beta_0) = \prod_{t=1}^{T} f(y_t/x_t;\beta_0) \tag{7}$$

The maximum likelihood estimator of β_0 is defined as

$$\hat{\beta}_{ML} = \arg\max_{\beta} \{\log(L_1(y,x;\beta_0))\} \tag{8}$$

The asymptotic distribution of $\hat{\beta}_{ML}$ can be easily derived as

$$\sqrt{T}(\hat{\beta}_{ML} - \beta_0) \xrightarrow{d} N\left(0, \left\{-E\left[\frac{1}{T}H_1(\beta_0)\right]\right\}^{-1}\right) \tag{9}$$

where $H_1(\beta_0)$ is the Hessian of the log-likelihood, i.e.,

$$H_1(\beta_0) = \frac{\partial^2 \ln(L_1(y,x;\beta_0)}{\partial\beta\partial\beta'}$$

For simplifying porposes, call *standard assumptions* (SA) (see for instance Amemiya (1985) for details) the set of assumptions that are required on the previous model to derive the asymptotic properties of the ML estimator.

Now, available prior information on parameter β is considered in the form of q stochastic restrictions, given by the equation:

$$r^* = R\beta + v \tag{10}$$

where $v \sim iid(0, \sigma_v^* I_q)$, independent of u, σ_v^* choosen according with the quality of priors, r^* is a q dimension vector containing the expected values of the restrictions describing priors and matrix R of dimension qxk describes the set of linear restrictions about parameters.

In order to support asymptotic efficiency gains, A1 is assumed, - that is, $V(v) = \sigma_v^* = \sigma_v / T^* -$, and also A3 or A4 – say A3 without loss of generality. Since v_i are independent, by applying the Lindberg Levy Central Limit Theorem on (10), it is easy to check that

$$r^* \sim N(R\beta, \sigma_v^* I_q) \tag{11}$$

and the density of the above distribution (11) can be added to the conditional distributions of the sample information. The marginal density of each one of q terms on the stochastic restrictions is

$$h(r_i^*) = \frac{1}{\sigma_v^* \sqrt{2\pi}} e^{-\frac{v_i^2}{2\sigma_v^{*2}}} \tag{12}$$

where $v_i = r_i - R_i^{'} \beta$, $i = 1, ..., q$, and $R_i^{'}$ is the ith row of R. The likelihood function related to stochastic restrictions is derived as considering priors as independent sample information, which equals the marginal densities of the sample. That is,

$$L_2(y, x; \beta_0) = \prod_{i=1}^{q} h(r_i^*; \beta_0) \tag{13}$$

This amounts to writing the whole likelihood for all *observations*, including that on the prior values of β, as

$$\bar{L}(y, x, r; \beta_0) = L_1.L_2 = \left(\prod_{t=1}^{T} f(y_t/x_t; \beta_0)\right)\left(\prod_{i=1}^{q} h(r_i, \beta_0)\right)$$

and the criterion for the restricted ML estimator is fully defined.

Definition. The stocastically restricted maximum likelihood (RML) estimator of β is

$$\hat{\beta}_{RML} = \arg\max_{\beta} \left\{\log \bar{L}\right\} \tag{14}$$

Next, it is a matter of interest deriving the properties of the defined $\hat{\beta}_{RML}$ estimator in order to discuss the efficiency gains with respect to the unrestricted ML estimator. Assuming A3, v independent of y and SA, the asymptotic distribution of $\hat{\beta}_{RML}$ is given by

$$\sqrt{T}(\hat{\beta}_{RML} - \beta_0) \xrightarrow{d} N(0, W) \tag{15}$$

where $W = \left\{-E\left[\frac{1}{T} H(\beta_0)\right]\right\}^{-1}$ and $H(\beta_0) = \frac{\partial^2 \log(\bar{L}(\beta_0))}{\partial\beta\partial\beta'}$

Proposition 5 Under SA, A1 and A3, $\hat{\beta}_{RML}$ is asymptotically more efficient than $\hat{\beta}_{ML}$

As shown, Proposition 5 provides an example of applicability of the result suggested in Proposition 2 where the asymptotic relevance of the stochastic restrictions is sustained on the bases of the asymptotic informative requirement for priors. By taking into account priors on the maximum likelihood estimation in the form of stochastic restrictions, it is shown the existence of asymptotic efficiency gains, which is intended to bring more accurate estimates of the considered model. This result illustrate the applicability of the theoretical results described in section 3 and provides an empirical motivation for them in the setting of a familiar econometrics method.

5 Conclusions

This paper shows that prior information, modelled through stochastic restrictions could bring efficiency gains asymptotically. This result is found by assuming that sample information does not grows faster than prior information asymptotically. Moreover, a general setting is provided in which the variance covariance matrix of the stochastically restricted estimator converges asymptotically to the efficient restricted estimator if prior information grows faster than sample information.

As a corollary of the previous result about the asymptotic relevance of prior information, the main result is implemented over the Maximum Likelihood (ML) estimator. The proposed Restricted Maximum Likelihood estimator is shown to be more efficient than the basic ML estimator. Hence it is suggested as a solution to incorporate prior information into the general and useful ML procedure to improve efficiency, as opposed to the standard asymptotic analysis concludes. This example of the theoretical result that supports the relevance of the stochastic restrictions approach on the ML method, suggest an analytical framework to be considered for empirical applications.

Appendix

The proofs of the propositions are provided next.

Proposition 1. Under SAA, A1 and A2, $AV(\hat{\beta}_{GR}) = AV(\hat{\beta}_{OLS})$.

Proof.

$$\begin{aligned} AV(\hat{\beta}_{GR}) &= p\lim\left(\frac{X'X}{T}\frac{1}{\sigma_u^2} + \frac{T^*}{T}\frac{R'R}{\sigma_v^2}\right)^{-1} \\ &= E\left(\frac{X'X}{\sigma_u^2}\right)^{-1} = AV(\hat{\beta}_{OLS}) \end{aligned}$$

■

Proposition 2. Under SAA, A1 and A3, $AV(\hat{\beta}_{GR}) < AV(\hat{\beta}_{OLS})$.

Proof. Under A1, $AV(\hat{\beta}_{GR})$ is now

$$\begin{aligned} AV(\hat{\beta}_{GR}) &= p\lim\left(\frac{1}{T}\frac{X'X}{\sigma_u^2} + \frac{T^*}{T}\frac{R'R}{\sigma_v^2}\right)^{-1} \\ &< E\left(\frac{X'X}{\sigma_u^2}\right)^{-1} = AV(\hat{\beta}_{OLS}) \end{aligned}$$

■

Proposition 3. Under SAA, A1, and A4, $AV(\hat{\beta}_{GR}) = AV(\hat{\beta}_R) < AV(\hat{\beta}_{OLS})$, where $\hat{\beta}_R$ is the constrained estimator of the model

$$\begin{aligned} Y &= X\beta + u \\ r &= R\beta \end{aligned}$$

Proof. First I use a matrix algebra result. Let A, kxk, B kxq and C qxq, be matrixes such that $\exists$ A^{-1}, C^{-1}. Then,[2]

$$(A + BCB')^{-1} = A^{-1} - A^{-1}B(C^{-1} + B'A^{-1}B)^{-1}B'A^{-1} \tag{E1}$$

Now, taking

$$\begin{aligned} A &= \frac{1}{\sigma_u^2}\frac{X'X}{T}, \\ B &= R', \\ C &= I_q \frac{1}{T}\frac{1}{\sigma_v^{*2}} \overset{(A1)}{=} I_q \frac{T^*}{T}\frac{1}{\sigma_v^2} \end{aligned}$$

from (6) and (E1), $AV(\hat{\beta}_{GR})$ can be rewritten as

$$\begin{aligned} &p\lim \left[\left(\frac{X'X}{T}\frac{1}{\sigma_u^2} \right)^{-1} - \left(\frac{X'X}{T}\frac{1}{\sigma_u^2} \right)^{-1} R' \left[I_q \sigma_v^2 \frac{T}{T^*} + R \left(\frac{X'X}{T}\frac{1}{\sigma_u^2} \right)^{-1} R' \right]^{-1} R \left(\frac{X'X}{T}\frac{1}{\sigma_u^2} \right)^{-1} \right] \\ &\overset{(A4)}{=} \left[E\left(\frac{X'X}{\sigma_u^2} \right)^{-1} - E\left(\frac{X'X}{\sigma_u^2} \right)^{-1} R' \left[RE\left(X'X\right)^{-1} R' \right]^{-1} RE\left(X'X\right)^{-1} \right] \end{aligned} \tag{E2}$$

since by A4, $p\lim T/T^* = 0$. It can be easily checked that the last term of the above equation is $AV(\hat{\beta}_R) < AV(\hat{\beta}_{OLS})$ shown in Proposition 2.

■

Proposition 4. Under (A01), if $\lim(T/T^*) = M$, $0 < M < \infty$, then,

[2] See Greene (1997), for instance.

a) $\frac{d(AV(\hat{\beta}_{GR}))}{dM} > 0$

b) $\lim_{M\to 0} AV(\hat{\beta}_{GR}) = AV(\hat{\beta}_R)$

c) $\lim_{M\to\infty} AV(\hat{\beta}_{GR}) = AV(\hat{\beta}_{OLS})$

Proof. Calling $H(M)$

$$H(M) = \left[I_q \sigma_v^2 M + Rp\lim \left(\frac{X'X}{T\sigma_u^2} \right)^{-1} R' \right]^{-1} \tag{E3}$$

and back to equation (16),

$$AV(\hat{\beta}_{GR}) = E\left(\frac{X'X}{\sigma_u^2}\right)^{-1} - E\left(\frac{X'X}{\sigma_u^2}\right)^{-1} R'H(M)R'^{-1}RE\left(\frac{X'X}{\sigma_u^2}\right)^{-1}$$

Since $\frac{dAV(\hat{\beta}_{GR})}{dH} < 0$ and $\frac{dH(M)}{dM} < 0$, then $\frac{dAV(\hat{\beta}_{GR})}{dM} > 0$, what proves a).

From (E3),

$$\begin{aligned} \lim_{M\to 0} H(M) &= \left[RE\left(\frac{X'X}{\sigma_u^2}\right)^{-1} R' \right]^{-1} \\ \lim_{M\to\infty} H(M) &= 0 \end{aligned}$$

By substituting these last terms into (16), it is obtained: $\lim_{M\to 0} AV(\hat{\beta}_{GR}) = AV(\hat{\beta}_R)$, and $\lim_{M\to\infty} AV(\hat{\beta}_{GR}) = AV(\hat{\beta}_{OLS})$, what proves b) and c).

■

Proposition 5 Under SA, A1 and A3, $\hat{\beta}_{RML}$ is asymptotically more efficient than $\hat{\beta}_{ML}$.

Proof.

The asymptotic distributions of $\hat{\beta}_{RML}$ and $\hat{\beta}_{ML}$ are respectively

$$\sqrt{T}(\hat{\beta}_{RML} - \beta_0) \xrightarrow{d} N\left(0, (-A)^{-1}\right)$$

$$\sqrt{T}(\hat{\beta}_{ML} - \beta_0) \xrightarrow{d} N\left(0, (-B)^{-1}\right)$$

where $A = E\left[\frac{1}{T}\frac{\partial^2 \log(\bar{L})}{\partial\beta\partial\beta^{'}}\right]$ and $B = E\left[\frac{1}{T}\frac{\partial^2 \log(L_1)}{\partial\beta\partial\beta^{'}}\right]$.

Statemente to be proved is $-A^{-.1} < -B^{-1}$, which holds if and only if

$$A^{-1} \quad > \quad B^{-1}$$

$$A \quad < \quad B$$

By construction, $\bar{L} = L_1.L_2$ and by operating adequately,

$$E\left[\frac{1}{T}\frac{\partial^2 \log(\bar{L})}{\partial\beta\partial\beta^{'}}\right] = E\left[\frac{1}{T}\frac{\partial^2 \log(L_1)}{\partial\beta\partial\beta^{'}}\right] + E\left[\frac{1}{T}\frac{\partial^2 \log(L_2)}{\partial\beta\partial\beta^{'}}\right]$$

that is,

$$A = B + C \text{ (say)}$$

Since second order condition holds for the likelihood criterion, it follows from equation that $A < B$, if $C \neq 0$, which have to be proved.

By equation (13)

$$\frac{\partial^2 \log(L_2)}{\partial\beta\partial\beta^{'}} = \frac{\partial^2 \left(\sum_{i=1}^{q} \log h(r_i^*; \beta_0)\right)}{\partial\beta\partial\beta^{'}}$$

and by (12), A1 and A3,

$$
\begin{aligned}
p\lim\left(\frac{1}{T}\frac{\partial^2 \log h(r_i^*)}{\partial\beta\partial\beta^{'}}\right) &= -p\lim\left(\frac{1}{T}\frac{R_i'R_i}{\sigma_v^*}\right) \\
&= -E\left(\frac{R_i'R_i}{\sigma_v}\right) \neq 0
\end{aligned}
$$

Hence

$$
E\left[\frac{1}{T}\frac{\partial^2 \log(L_2)}{\partial\beta\partial\beta^{'}}\right] = -E\left[\frac{R'R}{\sigma_v}\right] = C \neq 0
$$

what ends the proof.

■

References

Amemiya, T., 1985, Advanced Econometrics (Hardvard University Press, Cambridge)

Bekker, P., 1994, Alternative approximations to the distributions of instrumental variable estimators, Econometrica 62, 657-681.

Greene, W 1997, Econometric Analysis (Prentice Hall, New York).

Kadane, J. B., 1971, Comparison of k-class Estimator when the Disturbances are Small, Econometrica 39, 723-737.

Lütkepohl, H., 1993, Introduction to Multiple Time Series Analysis (Springer Verlag, Berlin).

Shiller, R., 1973, A Distributed Lag Estimator Derived from Smoothness Priors, Econometrica 41, 775-778.

Staiger, D., Stock, J. 1997, Instrumental Variables Regression with Weak Instruments, Econometrica 65, pp. 557-586.

Theil, H and A. Goldberger, 1961, On Pure and Mixed Statistical Estimation in Economics, International Economic Review 2, 65-78.