STATISTICA, anno LXIV, n. 1, 2004

MAXIMUM ENTROPY LEUVEN ESTIMATORS AND MULTICOLLINEARITY

Quirino Paris

1. INTRODUCTION

On a dark and sleepless night in Leuven, Belgium, where my wife Carlene was undergoing a difficult cancer treatment, I was struck by a ray of light coming through the window of our apartment looking onto Ladeuze Plein. The light striking my eyes scattered in a myriad of directions forming all sort of images as in a kaleidoscope. At that moment, my mind wandered to my amateurish readings about the theory of light and to the unexplainable finding of Quantum Electro-Dynamics (QED) according to which *the probability that a photomultiplier is hit by a photon reflected from a sheet of glass is equal to the square of its amplitude*. The amplitude of a photon is an arrow (a vector) that summarizes all the possible ways in which a photon could have reached a given photomultiplier.

This totally implausible discovery about light and matter was presented by Richard Feynman (1985, page 24) in clear and entertaining ways more than fifteen years ago: "The situation today is, we haven't got a good model to explain partial reflection by two surfaces; we just calculate the probability that a particular photomultiplier will be hit by a photon reflected from a sheet of glass. I have chosen this calculation as our first example of the method provided by the theory of quantum electrodynamics. I am going to show you 'how we count the beans' – what the physicists do to get the right answer. I am not going to explain how the photons actually 'decide' whether to bounce back or go through; that is not known. (Probably the question has no meaning.) I will only show how to calculate the correct probability that light will be reflected from a glass of given thickness, because that's the only thing physicists know how to do! … You will have to brace yourself for this – not because it is difficult to understand, but because it is absolutely ridiculous: All we do is draw little arrows on a piece of paper – that's all! Now, what does an arrow have to do with the chance that a particular event will happen? According to the rules of 'how we count the beans,' the probability of an event is equal to the square of the length of the arrow." But why would econometric analysis have anything to do with QED? In fact, it has little to do with it. Except that the analogy between the theory of light and the theory of information became so irresistible in that sleepless night in Leuven.

The analogy can be elaborated along the following lines. Light carries information about the physical environment. When light reaches the eyes (photomultipliers) of a person, the perceived image may be out-of-focus. That person will squint and adjust his eyes in order to improve the reproduction of the image in his brain. It is an astonishing fact of life that every individual, whether wearing glasses or not, knows when a picture is in focus and can adjust a projector to put a picture in focus for an entire audience. Economic data carry information about economic environments and the decision processes that generated those data. As with any picture, the economic information reaching a researcher may correspond to an image that is out-of-focus. Unfortunately, our brain is not wired to recognize when an economic picture is in focus. The goal of econometric analysis, then, is to reconstruct the best possible image of an economic decision process as the way to better understand the economic agent's environment.

This description of econometric analysis is of old vintage. The means to achieving a "better" statistical image of the economic process relies heavily upon the estimator selected by the researcher for this purpose (along with a correctly specified economic model). The novelty of this paper, then, is the proposal of a new class of statistical estimators inspired by the theory of light.

In the next sections, two maximum entropy Leuven (MEL) estimators will be presented. For convenience, they will be numbered Leuven-1 and Leuven-2. The MEL estimators are consistent and asymptotically normal. Properties such as asymptotic unbiasedness, consistency and normality of the parameter estimates will be illustrated by means of Monte Carlo experiments. We will present also a preliminary comparison with rival estimators such as the generalized maximum entropy (GME) estimator (Golan, *et al.*, 1996) and the ordinary least-squares (OLS) estimator. A particularly interesting aspect of this comparison is represented by the behavior of these estimators under a condition of increasing multicollinearity as measured according to Belsley *et al.* (1980) recommendation. The Leuven estimators outperform the OLS estimator for all values of the condition number examined in several Monte Carlo experiments. It is also important to anticipate that, in contrast to the GME estimator, no subjective and a-priori information is necessary in order to implement any of the Leuven estimators.

2. LEUVEN-1 AND LEUVEN-2 ESTIMATORS

The idea of maximum entropy within the context of information was introduced (Jaynes, 1957) as a way to deal with recovering of inverse images under limited information. The Leuven-1 and Leuven-2 estimators are based upon the same formalism and share the same entropy structure. The difference between them consists in the fact that the Leuven-2 estimator extends the entropy specification to the error term.

Let us consider a general, linear statistical model representing some economic relation (production, demand, cost function) that characterizes the following set of data generating processes (DGP):

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u}, \qquad \mathbf{u} \sim \text{IID}(\mathbf{0}, \sigma^2\mathbf{I}), \tag{1}$$

where the dimensions of the various components are $\mathbf{y} \sim (T \times 1)$, $\mathbf{u} \sim (T \times 1)$, $\boldsymbol{\beta} \sim (K \times 1)$ and $\mathbf{X} \sim (T \times K)$. The vector $\mathbf{y}$ and the matrix $\mathbf{X}$ constitute sample information. The vector β represents parameters to estimate and the vector $\mathbf{u}$ contains random disturbances independently and identically distributed (IID) with zero mean and variance σ^2.

In an econometric model with noise, it is impossible to measure exactly the parameters involved in the generation of the sample data. Each parameter depends on every other parameter specified in the model and its measured dimensionality is affected by the available sample information as well as by the measuring procedure. Following the theory of light, it is possible to estimate the probability of such parameters using their revealed image. The revealed image of a parameter can be thought of as the estimable dimensionality that depends on the sample information available for the analysis. Hence, in the Leuven-1 estimator we postulate that the probability of a parameter β_k (which carries economic information) is equal to the square of its "amplitude" where by amplitude we intend its estimated normalized dimensionality. Thus, the Leuven-1 estimator is specified as follows:

$$\min H(\mathbf{p}_\beta, L_\beta, u) = \sum_k p_{\beta_k} \log(p_{\beta_k}) + L_\beta \log(L_\beta) + \sum_t u_t^2 \tag{2}$$

subject to

$$y_t = \sum_k x_{tk}\beta_k + u_t$$

$$L_\beta = \sum_k \beta_k^2$$

$$p_{\beta_k} = \beta_k^2 / L_\beta$$

with $p_{\beta_k} \geq 0$, $k = 1,...,K$, $t = 1,...,T$. The amplitude (or normalized dimensionality) of parameter β_k is given by $\beta_k / \sqrt{L_\beta}$, hence the probability of parameter β_k is given by the square of its amplitude, as in the theory of light. The term $L_\beta \log(L_\beta)$ in the objective function prevents the overflow of the L_β parameter. The Leuven-1 estimator does not require any subjective a-priori information. It utilizes the components of the statistical linear model to define the relevant amplitude of the corresponding parameters.

In matrix notation, the Leuven-1 estimator assumes the following specification:

$$\min H(\mathbf{p}_\beta, L_\beta, \mathbf{u}) = \mathbf{p}'_\beta \log(\mathbf{p}_\beta) + L_\beta \log(L_\beta) + \mathbf{u}'\mathbf{u} \tag{3}$$

subject to

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

$$L_\beta = \boldsymbol{\beta}'\boldsymbol{\beta}$$

$$\mathbf{p}_\beta = \boldsymbol{\beta}\Theta\boldsymbol{\beta} / L_\beta$$

where $\mathbf{p}_\beta \geq \mathbf{0}$ and the symbol Θ indicates the element-by-element Hadamard product.

The Leuven-1 estimator does not possess a closed form representation. Its solution requires the use of a computer code for nonlinear programming problems such as GAMS (Brooke *et al.*, 1988). In order to examine the intricate structure of the Leuven-1 estimator it is useful to derive the corresponding Karush-Kuhn-Tucker (KKT) conditions. The corresponding Lagrangean function is given as

$$\begin{aligned} \boldsymbol{L} = \mathbf{p}'_\beta \log(\mathbf{p}_\beta) + L_\beta \log(L_\beta) + \mathbf{u}'\mathbf{u} + \lambda'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{u}) + \\ + \mu(L_\beta - \boldsymbol{\beta}'\boldsymbol{\beta}) + \boldsymbol{\eta}'(\mathbf{p}_\beta - \boldsymbol{\beta}\Theta\boldsymbol{\beta} / L_\beta) \end{aligned} \tag{4}$$

where the symbols $\boldsymbol{\lambda}, \mu, \boldsymbol{\eta}$ are the Lagrange multipliers of the corresponding constraints.

The relevant KKT conditions of problem (3) are stated as follows:

$$\begin{aligned} \frac{\partial \boldsymbol{L}}{\partial \mathbf{p}_\beta} &= \log(\mathbf{p}_\beta) + \boldsymbol{\iota}_K + \boldsymbol{\eta} = \mathbf{0} \\ \frac{\partial \boldsymbol{L}}{\partial L_\beta} &= \log(L_\beta) + 1 + \mu + \boldsymbol{\eta}'\boldsymbol{\beta}\Theta\boldsymbol{\beta} / L_\beta^2 = \log(L_\beta) + 1 + \mu + \boldsymbol{\eta}'\mathbf{p}_\beta / L_\beta = 0 \\ \frac{\partial \boldsymbol{L}}{\partial \beta} &= -\mathbf{X}'\boldsymbol{\lambda} - 2\mu\boldsymbol{\beta} - 2\boldsymbol{\eta}\Theta\boldsymbol{\beta} / L_\beta = \mathbf{0} \\ \frac{\partial \boldsymbol{L}}{\partial \mathbf{u}} &= 2\mathbf{u} - \boldsymbol{\lambda} = \mathbf{0} \end{aligned} \tag{5}$$

where the symbol $\boldsymbol{\iota}_K$ represents a vector of unit elements of dimension K. The solution of these KKT conditions, if it exists, will produce always a vector of probabilities with all positive components. It is apparent that the Leuven-1 estimator is nonlinear in the parameters but, in spite of its complexity, the empirical solution of numerous test problems was swift and efficient on the same level of rapidity of the least-squares estimator.

The Leuven-2 estimator extends the probability specification to the error term $\mathbf{u}$ resulting in the following symmetric structure:

$$\min H(\mathbf{p}_\beta, L_\beta, \mathbf{p}_u, L_u) = \sum_k p_{\beta_k} \log(p_{\beta_k}) + \sum_t p_{u_t} \log(p_{u_t}) + \\ + L_\beta \log(L_\beta) + L_u \log(L_u) \quad (6)$$

subject to

$$y_t = \sum_k x_{tk}\beta_k + u_t$$

$$L_\beta = \sum_k \beta_k^2$$

$$p_{\beta_k} = \beta_k^2 / L_\beta$$

$$L_u = \sum_t u_t^2$$

$$p_{u_t} = u_t^2 / L_u$$

with $p_{\beta_k} \geq 0$ and $p_{u_t} \geq 0$, $k = 1,...,K$, $t = 1,...,T$. Except for the probability specification of the error term, the Leuven-2 estimator shares the same structure and characteristics of the Leuven-1 estimator. Again, the Leuven-2 estimator does not require any subjective exogenous information as does the GME estimator.

3. THE CLASS OF MEL ESTIMATORS AS RIVAL TO THE OLS AND GME ESTIMATORS

In 1996, Golan, Judge and Miller proposed a way to extend Jaynes' (1957) maximum entropy formalism in econometrics to any sort of linear statistical models. Their assumption is that a parameter β_k is regarded as the mathematical expectation of some discrete support values Z_{km} such that

$$\beta_k = \sum_{m=1}^{M} Z_{km} p_{km} \quad (7)$$

where $p_{km} \geq 0$, $k = 1,...,K$, and $m = 1,...,M$ are probabilities and, of course, $\sum_{m=1}^{M} p_{km} = 1$ for $k = 1,...,K$. The element Z_{km} constitutes a-priori information provided by the researcher, while p_{km} is an unknown probability whose value must be determined by solving a maximum entropy problem.

Golan, Judge and Miller (1996) present a thorough discussion of the generalized maximum entropy (GME) estimator. In this estimator, the error terms in model (1) are also reparametrized with given discrete supports V_{tg}, $g = 1,...,G$, $t = 1,...,T$. Let $u_t = \sum_{g=1}^{G} V_{tg} w_{tg}$, with $w_{tg} \geq 0$, where the w_{tg} elements are regarded as probabilities associated with the error support values. Then, the GME estimator can be stated as

$$\max H(\mathbf{p},\mathbf{w}) = -\sum_{k=1}^{K}\sum_{m=1}^{M} p_{km}\log(p_{km}) - \sum_{t=1}^{T}\sum_{g=1}^{G} w_{tg}\log(w_{tg}) \tag{8}$$

subject to

$$y_t = \sum_{k=1}^{K} X_{tk}\sum_{m=1}^{M} Z_{km}p_{km} + \sum_{g=1}^{G} V_{tg}w_{tg}\,, \qquad t = 1,\dots,T$$

$$\sum_{m=1}^{M} p_{km} = 1\,, \qquad k = 1,\dots,K$$

$$\sum_{g=1}^{G} w_{tg} = 1\,, \qquad t = 1,\dots,T\,.$$

The GME estimator is not sensitive to multicollinearity because the matrix $\mathbf{X}'\mathbf{X}$ does not appear on the main diagonal of the appropriate KKT conditions.

The GME estimator, however, has important weaknesses for which the class of MEL estimators provides a remedy: The estimates of parameter β_k and residual u_t are sensitive, in an unpredictable way, to changes in the support intervals. Caputo and Paris (2000) have done a general and complete analysis of this aspect. A concomitant but distinct weakness of the GME estimator is that the parameter estimates and their variances are affected by the number of discrete support values. Many traditional econometricians reject the GME estimator because of these unsatisfactory properties. In effect, it is somewhat disappointing to inject subjective information into the estimation and data analysis process without knowing in what way this exogenous information will affect the estimated parameters. Also, while knowledge of the bounds for some parameters may be available and, therefore, ought to be used, it is unlikely that this knowledge can cover all the parameters of a model. In other words, the GME estimator depends crucially upon the subjective and exogenous information supplied by the researcher: The same sample data in the hands of different researchers willing to apply the GME estimator will produce different estimates of the parameters and, likely, different policy recommendations.

The class of MEL estimators rivals also the OLS estimator because of its better performance under conditions of increasing multicollinearity, an empirical event that plagues the majority of econometric analyses.

4. DISTRIBUTIONAL PROPERTIES OF MEL ESTIMATORS

The Leuven estimators are consistent and asymptotically normal. A proof of this proposition is presented in the appendix. To illustrate these properties, several Monte Carlo experiments were performed. In particular, consistency, asymptotic unbiasedness and normality of the estimated parameters were considered. In these experiments, the value of the mean squared error criterion tends to zero for a large sample size, supporting the notion that the estimators are consistent and asymptotically unbiased. Furthermore, the behavior of the estimators under increasing levels of multicollinearity was analyzed.

Consistency and asymptotic unbiasedness were measured by the magnitude of the mean squared error (MSE) criterion and of the squared bias in a risk function $\rho(\boldsymbol{\beta},\hat{\boldsymbol{\beta}})$, also called mean squared error loss (MSEL), as suggested by Judge *et al.* (1982, p. 558), where

$$\rho(\boldsymbol{\beta},\hat{\boldsymbol{\beta}}) = tr\mathrm{MSE}(\boldsymbol{\beta},\hat{\boldsymbol{\beta}}) = tr\mathrm{E}[(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'] = \mathrm{E}[(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})] \tag{9}$$

$$= tr\mathrm{COV}(\hat{\boldsymbol{\beta}}) + tr[\mathrm{BIAS}(\hat{\boldsymbol{\beta}}) * \mathrm{BIAS}(\hat{\boldsymbol{\beta}}')] .$$

Tables 1 and 2 present the results of a non-trivial Monte Carlo experiment that deals with a true model exhibiting the following data generating process (DGP). There are ten parameters β_{0k}, $k = 1,...10$, to estimate. Each parameter β_{0k} was drawn from a uniform distribution $U[-1.7, 2.0]$. Each element of the matrix of regressors $\mathbf{X}$ was drawn from a uniform distribution $U[1,5]$. The model has no intercept. Finally, each component of the disturbance vector $\mathbf{u}$ was drawn from a normal distribution $N(0,\sigma_0^2) = N(0,4)$. With this specification, the dependent variable $\mathbf{y}$ was measured in units of tens, ranging from 10 to 100 (in absolute value). Runs of one hundred samples of increasing size, from 50 to 5000 observations, were executed. The GME estimator was implemented with discrete support intervals for the parameters and the error terms selected as $[-5,0,5]$ and $[-10,5,10]$, respectively. The condition number (CN) (Belsley *et al.*, 1980) of the $\mathbf{X}$ matrix is given for each sample size.

TABLE 1

Monte Carlo experiment N. 1: model without intercept. Asymptotic unbiasedness of rival estimators. 100 samples

Estimators	*T*=50 CN=11.5	*T*=200 CN=10.3	*T*=400 CN=9.5	*T*=1000 CN=9.1	*T*= 2000 CN=8.8	*T*=5000 CN=8.1
Leuven-1	0.03630	0.00371	0.00043	0.00078	0.00013	0.00003
Leuven-2	0.00992	0.00179	0.00012	0.00058	0.00013	0.00003
GME	0.04986	0.00451	0.00057	–	–	–
OLS	0.00893	0.00170	0.00012	0.00056	0.00013	0.00003

The GME estimator implemented with the optimization program GAMS failed to reach an optimal solution with a sample size of $T > 400$. This event might be due to the large number of probabilities that must be estimated for an increasing number of error terms. The GME estimator produces results that approximate very closely uniform probabilities and this characteristic of the GME estimator may make it difficult with large samples to locate a maximum value of the objective function. Invariably, the GAMS program terminated with a feasible but non-optimal solution when $T > 400$.

The levels reported in table 1 represent the sum of the squared bias over ten parameters. It would appear that the Leuven-2 estimator performs as well as the OLS estimator in small samples. When the sample size increases, both Leuven estimators rival the OLS estimator. This result is confirmed in table 2 that presents the levels of MSEL for the same experiment and sample sizes.

TABLE 2

Monte Carlo experiment N. 1: model without intercept. MSEL for rival estimators. 100 samples

Estimators	*T*=50 CN=11.5	*T*=200 CN=10.3	*T*=400 CN=9.5	*T*=1000 CN=9.1	*T*=2000 CN=8.8	*T*=5000 CN=8.1
Leuven-1	0.5448	0.1334	0.0709	0.0295	0.0132	0.0052
Leuven-2	0.5661	0.1347	0.0714	0.0294	0.0132	0.0052
GME	0.5469	0.1341	0.0715	–	–	–
OLS	0.5882	0.1351	0.0715	0.0294	0.0132	0.0052

The MSEL values of the Leuven estimators in table 2 tend to zero as T increases at the same rate as the MSEL value of the OLS estimator. This evidence supports the proposition that the Leuven estimators are consistent.

The hypothesis that the parameter estimates are distributed according to a normal distribution was tested by the Bera-Jarque (1981) statistic involving the coefficients of skewness and kurtosis that the authors show to be distributed as a χ^2 variable with two degrees of freedom. In all the runs associated with tables 1 and 2, the normality hypothesis was not rejected with ample margins of safety.

The above results provide evidence that the Leuven-1 and Leuven-2 estimators perform as well as the OLS estimator, under a well-conditioned $\mathbf{X}'\mathbf{X}$ matrix. The Leuven estimators out-perform the OLS estimator under a condition of increasing multicollinearity. Following Belsley *et al.* (1980), multicollinearity can be detected in a meaningful way by means of a condition number computed as the square root of the ratio between the maximum and the minimum eigenvalues of a matrix $\mathbf{X}'\mathbf{X}$ (not a moment matrix) whose columns have been normalized to a unit length. These authors found that the negative effects of multicollinearity begin to surface when the condition number is around 30. A Monte Carlo experiment was conducted to examine the behavior of the MSEL criterion under increasing values of the condition number with a given sample size of $T = 50$. The experiment's structure is identical to that one associated with tables 1 and 2. The results are presented in table 3.

The Leuven-1 estimator reveals a remarkable stability as the condition number increases. On the contrary, and as expected, the OLS estimator shows a dramatic increase in the MSEL levels for values of the condition number that can be easily encountered in empirical econometric analyses. The Leuven-2 estimator reveals a slightly less stable behavior although it seems to converge to the same level of MSEL achieved by the Leuven-1 estimator for higher values of the condition number. Also the Leuven-2 estimator outperforms the OLS estimator uniformly. The GME estimator was implemented in two versions with two different support

TABLE 3

Monte Carlo experiment N. 1: model without intercept. MSEL of rival estimators for an increasing condition number. T=50, 100 samples

Condition Number	Estimators Leuven-1	Leuven-2	GME(-5,5)	GME(-20,20)	OLS
11	0.545	0.576	0.547	0.584	0.588
30	0.800	1.016	0.773	1.059	1.092
60	0.922	1.832	0.858	2.195	2.561
101	0.876	2.457	0.818	3.890	6.120
203	0.792	2.169	0.758	5.316	23.009
304	0.768	1.667	0.742	4.424	50.908
508	0.754	1.183	0.733	2.694	117.601
1,018	0.749	0.898	0.730	1.346	219.350
4,478	1.120	1.123	1.103	1.155	560.338
42,187	1.126	1.133	1.108	1.109	601.108

intervals of the parameters.[1] The first version of GME, with narrow support intervals, reveals a stability comparable to that of the Leuven-1 estimator. The second version of GME, with wider support intervals, exhibits a significant increase in MSEL values. When the number of repeated samples was increased to 300, the results were very similar to those given in tables 1, 2 and 3.

5. SCALING PROPERTIES OF MEL ESTIMATORS

With regard to scaling, the Leuven estimators are "invariant" to an arbitrary change of measurement units of the sample information in the same sense that the OLS estimator is "invariant" to a change of scale of either the dependent variable or the regressors. In reality, a more proper characterization of the OLS and Leuven estimators under different scaling is that their estimates change in a known way due to a known (but arbitrary) choice of measurement units of either the dependent variable or regressors or both. Because of this knowledge, it is always possible to recover the original estimates obtained prior to the scale change and, in this sense, both the OLS and the Leuven estimators are said to be scale invariant.

The proof of scale invariance for the Leuven estimators requires a discussion of the relevant KKT conditions because these estimators lack a closed form solution. The main line of reasoning runs as follows: if the KKT conditions corresponding to two different and arbitrary scaling schemes of the sample informa-

[1] Golan *et al.* in their 1996 book (Chapter 8) analyze the behavior of the GME estimator against the OLS estimator using the wrong notion of condition number. Although they quote Belsley *et al.* (1980), their condition number is simply the ratio of the maximum to the minimum eigenvalues of the $\mathbf{X'X}$ matrix (not the square root of this ratio, as indicated by Belsley et al.). In their empirical analysis, they selected values of the condition number that varied from 1 to 100 which correspond to values of Belsley's condition number from 1 to 10. Because multicollinearity begins to signal its deleterious effects when Belsley's condition number is around 30, the discussion of Golan *et al.* (1996) does not involve empirical problems that are ill-conditioned. The rapidly rising values of the MSEL detected for the OLS estimator are due to the rather small sample size (T=10) selected for their Monte Carlo experiment.

tion produce solutions that can be interchanged in the respective KKT conditions by means of an arbitrary and known linear operator, the Leuven estimators are said to be scale-invariant. The KKT conditions for the Leuven-1 estimator corresponding to an unscaled model are given in system (5).

Notice that the fourth equation of (5) involving only the variables $\mathbf{u}$ and $\boldsymbol{\lambda}$ establishes the symmetric duality between error terms and the Lagrange multipliers of the linear statistical model: in the unscaled Leuven-1 model, therefore, the Lagrange multipliers $\boldsymbol{\lambda}$ are always twice as large as the estimated residuals. With such a general result, the fourth equation can be eliminated and the third equation of (5) can be rewritten as

$$-\mathbf{X}'\mathbf{u} - \mu\boldsymbol{\beta} - \boldsymbol{\eta}\Theta\boldsymbol{\beta} / L_\beta = \mathbf{0}\,. \tag{10}$$

We will regard the first two equations of system (5) plus equation (1) as representing the relevant KKT conditions for deriving the scale-invariance property of the Leuven-1 estimator.

We now scale the dependent variable $\mathbf{y}$ of the linear statistical model in equation (1) by an arbitrary but known scalar parameter R and the matrix of regressors $\mathbf{X}$ by an arbitrary but known linear operator $\mathbf{S}$ regarded as a non-singular matrix of dimensions $(K \times K)$. Under this scaling scheme, the linear statistical model given by equation (1) assumes the following representation:

$$\begin{aligned} \frac{\mathbf{y}}{R} &= \left(\frac{\mathbf{X}}{R}\mathbf{S}^{-1}\right)\mathbf{S}\boldsymbol{\beta} + \frac{\mathbf{u}}{R} \\ \frac{\mathbf{y}}{R} &= \left(\frac{\mathbf{X}}{R}\mathbf{S}^{-1}\right)\boldsymbol{\beta}^* + \mathbf{u}^* \end{aligned} \tag{11}$$

where $\boldsymbol{\beta}^* \equiv \mathbf{S}\boldsymbol{\beta}$ and $\mathbf{u}^* \equiv \mathbf{u}/R$. The specification of the optimization model that will produce scale-invariant estimates of the Leuven-1 estimator can then be stated as

$$\min H(\mathbf{p}_\beta, L_\beta, \mathbf{u}) = \mathbf{p}'_\beta \log(\mathbf{p}_\beta) + L_\beta \log(L_\beta) + R^2 \mathbf{u}^{*\prime}\mathbf{u}^* \tag{12}$$

subject to

$$\frac{\mathbf{y}}{R} = \left(\frac{\mathbf{X}}{R}\mathbf{S}^{-1}\right)\boldsymbol{\beta}^* + \mathbf{u}^*$$

$$L_\beta = \boldsymbol{\beta}^{*\prime}\mathbf{S}^{-1\prime}\mathbf{S}^{-1}\boldsymbol{\beta}^*$$

$$\mathbf{p}_\beta = \mathbf{S}^{-1}\boldsymbol{\beta}^*\Theta\mathbf{S}^{-1}\boldsymbol{\beta}^* / L_\beta\,.$$

If the scalar R is equal to one and the matrix $\mathbf{S}$ is taken as the identity matrix, the model specified in equation (12) is identical to the model exhibited in equation (3). We need to show that the KKT conditions of model (12) produce a solution of the scaled model that can be used to recover a solution of the unscaled model which satisfies the two KKT equations of system (5) plus equation (1).

After setting up the Lagrangean function corresponding to model (12), the KKT conditions are derived as follows:

$$\begin{aligned}
&\frac{\partial \boldsymbol{L}}{\partial \mathbf{p}_\beta} = \log(\mathbf{p}_\beta) + \boldsymbol{\iota}_K + \boldsymbol{\eta} = \mathbf{0} \\
&\frac{\partial \boldsymbol{L}}{\partial L_\beta} = \log(L_\beta) + 1 + \mu + \boldsymbol{\eta}'\mathbf{S}^{-1}\boldsymbol{\beta}^* \Theta \mathbf{S}^{-1}\boldsymbol{\beta}^* / L_\beta^2 = \log(L_\beta) + 1 + \mu + \boldsymbol{\eta}'\mathbf{p}_\beta / L_\beta = 0 \\
&\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{\beta}^*} = -\mathbf{S}^{-1\prime}\frac{\mathbf{X}'}{R}\boldsymbol{\lambda} - 2\mu\mathbf{S}^{-1\prime}\mathbf{S}^{-1}\boldsymbol{\beta}^* - 2\mathbf{S}^{-1\prime}\boldsymbol{\eta}\Theta\mathbf{S}^{-1}\boldsymbol{\beta}^* / L_\beta = \mathbf{0} \\
&\frac{\partial \boldsymbol{L}}{\partial \mathbf{u}^*} = R^2 2\mathbf{u}^* - \boldsymbol{\lambda} = \mathbf{0}
\end{aligned} \tag{13}$$

Now, let us assume that the vector $(\hat{\mathbf{u}}, \hat{\boldsymbol{\beta}}, \hat{L}_\beta, \hat{\mathbf{p}}_\beta, \hat{\mu}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\eta}})$ represents a solution of the system of KKT conditions (13). We will show that this solution can be used to recover a vector of the same parameters that solves the first two equations of the KKT system (5) and equation (1). First of all, the first two equations of system (13) have a structure that is identical to the structure of the first two equations of system (5). Hence, the values of $\hat{\mathbf{p}}_\beta, \hat{\boldsymbol{\eta}}, \hat{L}_\beta$ and $\hat{\mu}$ that satisfy the first two equations of system (13) by assumption, satisfy also the first two equations of system (5). We can thus state that $\hat{\hat{\mathbf{p}}}_\beta = \hat{\mathbf{p}}_\beta$, $\hat{\hat{L}}_\beta = \hat{L}_\beta$, $\hat{\hat{\boldsymbol{\eta}}} = \hat{\boldsymbol{\eta}}$ and $\hat{\hat{\mu}} = \hat{\mu}$, where a double hat indicates a solution of the unscaled model. Furthermore, using the identity $\mathbf{u}^* \equiv \mathbf{u}/R$ we can obtain an estimate of the unscaled residuals as $\hat{\hat{\mathbf{u}}} = \hat{\mathbf{u}}R$ or $\hat{\mathbf{u}} = \hat{\hat{\mathbf{u}}}/R$. The fourth equation of system (13) can then be re-stated as

$$R^2 2\hat{\mathbf{u}} - \hat{\boldsymbol{\lambda}} = R^2 2\frac{\hat{\hat{\mathbf{u}}}}{R} - \hat{\boldsymbol{\lambda}} = R2\hat{\hat{\mathbf{u}}} - \hat{\boldsymbol{\lambda}} = \mathbf{0} \tag{14}$$

to signify that the Lagrange multiplier in the scaled linear model is R times as large as the corresponding Lagrange multiplier in the unscaled model since we know that, in unscaled models, $2\mathbf{u} = \boldsymbol{\lambda}$. We thus have $\hat{\hat{\boldsymbol{\lambda}}} = \hat{\boldsymbol{\lambda}}/R$. The solution value of the Lagrange multiplier $\hat{\boldsymbol{\lambda}}$ can be replaced by its equivalent expression [equation (14)] in the third equation of system (13) after pre-multiplying it by the matrix $\mathbf{S}'$ to obtain

$$-\frac{\mathbf{X}'}{R}R2\hat{\mathbf{u}} - 2\hat{\mu}\mathbf{S}^{-1}\hat{\boldsymbol{\beta}} - 2\hat{\boldsymbol{\eta}}\Theta\mathbf{S}^{-1}\hat{\boldsymbol{\beta}} / \hat{L}_{\beta} = \mathbf{0}\,. \tag{15}$$

Finally, equation (15) reduces to equation (10) after using the identity $\boldsymbol{\beta}^{*} \equiv \mathbf{S}\boldsymbol{\beta}$ from which we can obtain an estimate of the unscaled parameter $\boldsymbol{\beta}$ as $\hat{\hat{\boldsymbol{\beta}}} = \mathbf{S}^{-1}\hat{\boldsymbol{\beta}}$. To be explicit,

$$-\mathbf{X}'\hat{\hat{\mathbf{u}}} - \hat{\hat{\mu}}\hat{\hat{\boldsymbol{\beta}}} - \hat{\boldsymbol{\eta}}\Theta\hat{\hat{\boldsymbol{\beta}}} / \hat{\hat{L}}_{\beta} = \mathbf{0} \tag{16}$$

by making use also of the equalities dealing with parameters $L_{\beta}, \boldsymbol{\eta}$ and μ as stated above. Equation (16) has the same structure of equation (10) and, furthermore, we have found an unscaled solution (based upon the solution of the scaled model) that satisfies it. This completes the proof of the scale-invariance property of the Leuven-1 estimator.

In the OLS estimator, the parameter estimates are affected in a known way by arbitrary changes in the measurement units of both the dependent variable and the regressors (except for the special case in which both sets of variables change in the same way). On the contrary, the parameter estimates of the Leuven estimators do not change for an arbitrary variation of the measurement units of the dependent variables. They change only for a scale variation of the regressors.

The scale invariant specification of the Leuven-2 estimator assumes the following structure:

$$\min H(\mathbf{p}_{\beta}, L_{\beta}, \mathbf{p}_{u}, L_{u}) = \mathbf{p}'_{\beta}\log(\mathbf{p}_{\beta}) + L_{\beta}\log(L_{\beta}) + \mathbf{p}'_{u}\log(\mathbf{p}_{u}) + L_{u}\log(L_{u}) \tag{17}$$

subject to

$$\frac{\mathbf{y}}{R} = \left(\frac{\mathbf{X}}{R}\mathbf{S}^{-1}\right)\boldsymbol{\beta}^{*} + \mathbf{u}^{*}$$

$$L_{\beta} = \boldsymbol{\beta}^{*\prime}\mathbf{S}^{-1\prime}\mathbf{S}^{-1}\boldsymbol{\beta}^{*}$$

$$\mathbf{p}_{\beta} = \mathbf{S}^{-1}\boldsymbol{\beta}^{*}\Theta\mathbf{S}^{-1}\boldsymbol{\beta}^{*} / L_{\beta}$$

$$L_{u} = R^{2}\mathbf{u}^{*\prime}\mathbf{u}^{*}$$

$$\mathbf{p}_{u} = R^{2}\mathbf{u}^{*}\Theta\mathbf{u}^{*} / L_{u}\,.$$

The proof of scale invariance of the Leuven-2 estimator follows a line of reasoning that is similar to that developed for the Leuven-1 estimator.

6. CHANGE OF ORIGIN

The change of origin of the sample information (deviations from the mean, for example) produces two opposite results depending on whether or not the linear model has an intercept. For models without intercept, the parameter estimates of the Leuven estimators are invariant to a change of origin of the measurement units. In order to prove this result it is sufficient to show that a solution derived from a model whose sample information is defined in deviations from the mean satisfies also the KKT conditions of a model whose sample information is measured in natural units. The relevant KKT conditions of this latter model are given, again, by system (5).

In order to set up a model defined in deviations from the mean, it is convenient to define a deviaton operator $\mathbf{D} \equiv \left(\mathbf{I}_T - \frac{\boldsymbol{\iota}_T \boldsymbol{\iota}'_T}{T} \right)$ that will generate a dependent variable and regressors in deviations from their respective means. The vector $\boldsymbol{\iota}_T$ has T unit elements. The $\mathbf{D}$ operator is an idempotent symmetric matrix. Operating on vectors $\mathbf{y}$, $\mathbf{u}$ and matrix $\mathbf{X}$, the model in deviations from the mean is stated as

$$\min H(\mathbf{p}_\beta, L_\beta, \mathbf{u}) = \mathbf{p}'_\beta \log(\mathbf{p}_\beta) + L_\beta \log(L_\beta) + \mathbf{u}'\mathbf{u} \tag{18}$$

subject to

$$\mathbf{Dy} = \mathbf{DX}\boldsymbol{\beta} + \mathbf{Du}$$

$$L_\beta = \boldsymbol{\beta}'\boldsymbol{\beta}$$

$$\mathbf{p}_\beta = \boldsymbol{\beta}\boldsymbol{\Theta}\boldsymbol{\beta} / L_\beta .$$

The relevant KKT conditions of problem (18) are given by

$$\begin{aligned}
\frac{\partial \boldsymbol{L}}{\partial \mathbf{p}_\beta} &= \log(\mathbf{p}_\beta) + \boldsymbol{\iota}_K + \boldsymbol{\eta} = \mathbf{0} \\
\frac{\partial \boldsymbol{L}}{\partial L_\beta} &= \log(L_\beta) + 1 + \mu + \boldsymbol{\eta}'\boldsymbol{\beta}\boldsymbol{\Theta}\boldsymbol{\beta} / L_\beta^2 = \log(L_\beta) + 1 + \mu + \boldsymbol{\eta}'\mathbf{p}_\beta / L_\beta = 0 \\
\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{\beta}} &= -\mathbf{X}'\mathbf{D}\boldsymbol{\lambda} - 2\mu\boldsymbol{\beta} - 2\boldsymbol{\eta}\boldsymbol{\Theta}\boldsymbol{\beta} / L_\beta = \mathbf{0} \\
\frac{\partial \boldsymbol{L}}{\partial \mathbf{u}} &= 2\mathbf{u} - \mathbf{D}\boldsymbol{\lambda} = \mathbf{0}
\end{aligned} \tag{19}$$

Now, let us assume that the vector $(\hat{\mathbf{u}}, \hat{\boldsymbol{\beta}}, \hat{L}_\beta, \hat{\mathbf{p}}_\beta, \hat{\mu}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\eta}})$ represents a solution of the system of KKT conditions (19). By replacing $\mathbf{D}\boldsymbol{\lambda}$ in the third equation of system (19) by its equivalent expression given in the fourth equation of (19), the KKT conditions of the model in deviations from the mean have a structure that is identical to the KKT conditions (5). Hence, the solution $(\hat{\mathbf{u}}, \hat{\boldsymbol{\beta}}, \hat{L}_\beta, \hat{\mathbf{p}}_\beta, \hat{\mu}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\eta}})$ of system (19) will satisfy also system (5).

For models with intercept, the parameter estimates of the Leuven estimators are not invariant to a change of origin of the measurement units. This implies that the familiar practice of defining regressors and dependent variables in deviations from their mean is not admissible. The reason for this result depends upon the different dimension of the parameter space in the two specifications. The KKT conditions of the model estimated with an explicit intercept are articulated in six sets of relations (associated with $\mathbf{p}_G, p_1, L_\beta, \mathrm{b}_G, \beta_1, \mathbf{u}$, where β_1 is the intercept and $\boldsymbol{\beta}_G$ is the vector of the remaining parameters) whereas the KKT conditions of the model defined in deviations from the mean exhibits only five sets of relations (associated with $\mathbf{p}_G, p_1, L_\beta, \mathrm{b}_G, \mathbf{u}$). In other words, in models with intercept, the parameter space collapses by one dimension when the sample information is defined in deviations from the mean and no information is available to recover the parameter of the lost dimension.[2] The same reduction in the dimension of the parameter space occurs also in the OLS estimator but with it there exists a specific relation (based upon average sample information) that recovers the "missing" parameter β_1.

7. MODELS WITH INTERCEPT

The Monte Carlo experiment presented above dealt with a model without intercept. The nature of an intercept in a linear statistical model is different from the nature of all the other slope parameters. While slope parameters may be interpreted as elasticities (in a double logarithmic model), the intercept term is a catch-all parameter related, for example, to regressors that, for lack of sample information, are assumed to be kept at some unknown constant level. In principle, completely specified econometric models have no intercept since the great majority of economic relations (cost, profit, demand, and supply functions), are homogeneous (of either degree one or zero). In reality, many empirical econometric studies present large intercept values that are order of magnitude larger than the value of the remaining slope parameters. Aside from ignorance about relevant regressors, a large value of the intercept suggests that the dependent variable was not scaled properly. Whatever the reasons for the presence of an intercept, we now assume a model with an intercept that is order of magnitude larger (in absolute value) than the other slope parameters. In this case, it is convenient to separate the intercept from the other parameters and to define the probability relation only for these slope parameters. The intercept is regarded as the first parameter β_1. Then, the Leuven-1 estimator of this model with intercept is stated as

[2] Presumably, the same result applies to the GME estimator. In this case, the variant of the GME estimator proposed by van Akkeren and Judge (2000) is in jeopardy when dealing with models that exhibit an intercept because its implementation depends on defining the regressors and the dependent variable in deviations from the mean.

$$\min H(\mathbf{p}_\beta, L_\beta, \mathbf{u}) = \sum_{k=2}^{K} p_{\beta_k} \log(p_{\beta_k}) + L_\beta \log(L_\beta) + \sum_t u_t^2 \tag{20}$$

subject to

$$y_t = \beta_1 + \sum_{k=2}^{K} x_{tk}\beta_k + u_t$$

$$L_\beta = \sum_{k=2}^{K} \beta_k^2$$

$$p_{\beta_k} = \beta_k^2 / L_\beta$$

with $p_{\beta_k} \geq 0$, $k = 2,...,K$, $t = 1,...,T$. As we will illustrate by means of a second Monte Carlo experiment, the Leuven-1 estimator specified in (20) performs very well when a large intercept is present. A similar specification can easily be extended to the Leuven-2 estimator.

The second Monte Carlo experiment was generated by the following DGP: There are ten parameters β_{0k}, $k = 1,...10$, and β_{01} is considered the intercept with a true value of 15. Each remaining parameter β_{0k}, $k = 2,...,10$, was drawn from a uniform distribution $U[-2,3]$. Each element of the matrix of regressors $\mathbf{X}$ (other than the first column which has all unit values) was drawn from a uniform distribution $U[1,10]$. Finally, each component of the disturbance vector $\mathbf{u}$ was drawn from a normal distribution $N(0,\sigma_0^2) = N(0,4)$. With this specification, the dependent variable $\mathbf{y}$ was measured in units of tens, ranging from 10 to 100. One hundred samples of size T=50 were replicated. The GME estimator was implemented with discrete support intervals for the parameters and the error terms selected as [-20,0,20] and [-10,0,10], respectively. The condition number (CN) of the $\mathbf{X}$ matrix is given for each sample size.

TABLE 4

Monte Carlo experiment N. 2: model with intercept. MSEL and squared bias of rival estimators for an increasing condition number. T=50, 100 samples

Estimators	CN=23	CN=55	CN=135	CN=539	CN=898	CN=2,692
MSEL						
Leuven-1	4.862	3.780	4.460	4.606	4.593	4.573
Leuven-2	5.023	3.952	5.312	5.217	5.016	4.843
GME	15.508	13.707	15.916	15.891	15.014	14.966
OLS	5.061	4.253	8.972	93.053	212.582	488.583
Squared Bias						
Leuven-1	0.0053	0.5368	1.2814	1.5419	1.5368	1.5206
Leuven-2	0.0130	0.1260	0.6595	1.6257	1.6628	1.6333
GME	12.8909	10.9858	10.8893	11.6888	12.0451	12.3789
OLS	0.0187	0.1401	0.2251	1.8467	6.5685	32.5240

The main information presented by table 4 is that, given the DGP of this Monte Carlo experiment, the GME estimator exhibits MSEL values that are three times as large as the Leuven estimators. Furthermore, the levels of squared bias of the GME estimator are very large in comparison to those of the Leuven estimators. This evidence suggests that, in the presence of a model with a large value of the intercept (relative to the value of the other slope parameters), the use of the GME estimator may be unnecessarily too risky. Considerable level of risk can be avoided in this case by using one of the Leuven estimators. The OLS estimator outperforms the GME estimator for levels of multicollinearity associated with a condition number smaller than 150.

8. CONCLUSION

The class of MEL estimators is inspired by the theory of light and rivals the GME estimator of Golan *et al.*, (1996) by performing very well under the MSEL risk function while avoiding the requirement of subjective exogenous information that is a necessary component of the GME estimator. In a specific Monte Carlo experiment they outperform the GME estimator when a model has an intercept measured by orders of magnitude larger than the other slope parameters. The Leuven estimators are invariant to a change of scale in the sense of the OLS estimator. Furthermore, they are consistent and asymptotically normal.

In comparison to the GME estimator, the class of Leuven estimators is parsimonious with respect to the number of parameters to be estimated. For example, the solution of the Leuven-1 estimator has $(2K+T)$ components (K parameters β_k, K probabilities p_{β_k}, and T error terms u_t). The solution of the GME estimator for a similar model has $(MK+GT)$ components, where M is the number of discrete supports for the parameter β_k and G is the number of discrete supports for the error term L_β. The empirical GME literature indicates that, in general, $M=5$ and $G=3$.

The Leuven estimators appear to succeed where the ridge estimator failed: Under any levels of multicollinearity, the Leuven estimators uniformly dominate the OLS estimator according to the mean squared error criterion. For small samples $(T=50)$ or $(T=100)$, the Leuven estimators produce estimates that are different from those of the OLS estimator. These estimates are radically different under multicollinearity as the MSEL of the Leuven estimators is stable and very small relative to the MSEL of the OLS estimator.

Department of Agricultural and Resource Economics
University of California, Davis
QUIRINO PARIS

ACKNOWLEDGMENTS

This paper was written in honor and loving memory of my wife, Carlene Paris, who died of leiomyosarcoma – a very rare cancer – on May 5, 2001. I am indebted to Michael R. Caputo and Art Havenner for stimulating discussions and incisive comments that have contributed substantial improvements of an earlier version. All errors are mine.

REFERENCES

D. A. BELSLEY, E. KUH, R.E. WELSCH (1980), *Regression Diagnostics*, John Wiley, New York.

A. BERA, C. JARQUE (1981), *Efficient test for normality, heteroskedasticity and serial independence of regression residuals: Monte Carlo evidence*, "Economics Letters" 7, pp. 313-318.

A. BROOKE, D. KENDRICK, A. MEERAUS (1988), *GAMS: A User's guide*. The Scientific Press, New York.

M. R. CAPUTO, Q. PARIS (2000), *Comparative statics of the generalized maximum entropy estimator of the general linear model*, Department of Agricultural and Resource Economics, University of California, Davis, Working Paper, 25 pages.

R. FEYNMAN (1985), *QED: the strange theory of light and matter*, Princeton University Press, Princeton, New Jersey.

A. GOLAN, G. G. JUDGE, D. MILLER (1996), *Maximum entropy econometrics*, Wiley and Sons, Chichester, UK.

E. T. JAYNES (1957), *Information theory and statistical mechanics*, "Physics Review" 106 (1957): 620-630, Part II, 108 (1957): 171-190.

G.G. JUDGE, T.C. HILL, W.E. GRIFFITHS, H. LÜTKEPOHL, T.C. LEE (1982), *Introduction to the theory and practice of econometrics*. John Wiley, New York.

M. VAN AKKEREN, G.G. JUDGE (1999), *Extended empirical likelihood, estimation and inference*, University of California, Berkeley, Working Paper, 49 pages.

APPENDIX

The goal of this appendix is to prove that the Leuven-1 estimator is consistent and asymptotically normal. The strategy is based upon the realization that, in the limit, the objective function of the Leuven-1 estimator converges to the limit value of the objective function of the OLS estimator. This implies that the Leuven-1 estimator converges to the OLS estimator.

The Monte Carlo experiment N. 1 reported above provides empirical evidence that the Leuven-1 estimator might be consistent and asymptotically unbiased. Since it is well known that the OLS estimator is consistent and asymptotically normal, it will be sufficient to show convergence of the Leuven-1 estimator's objective function to the limit value of the OLS objective function in order to attain our stated goal. In other words, we will demonstrate that the sequence of random variables representing the objective function of the Leuven-1 estimator converges to the limit value of the objective function of the OLS estimator as the sample size tends to infinity. The simplest way to obtain this result is to make sure that the model's parameters of the Leuven-1 estimator are bounded by finite values so that, when the sample size will tend to infinity, the probability limit of certain expressions in the objective function will tend to zero. We must recall that the Leuven-1 estimator does not have a closed form solution and, therefore, the structure of the Leuven-1 estimator is given by its nonlinear optimization program or, equivalently, its set of KKT conditions. For convenience we restate the Leuven-1 estimator and its associated KKT conditions:

$$\min H(\mathbf{p}_\beta, L_\beta, \mathbf{u}) = \sum_k p_{\beta_k} \log(p_{\beta_k}) + L_\beta \log(L_\beta) + \sum_t u_t^2 \tag{21}$$

subject to

$$y_t = \sum_k x_{tk}\beta_k + u_t \qquad \lambda_t \tag{22}$$

$$p_{\beta_k} = \beta_k^2 / L_\beta \qquad \eta_k \tag{23}$$

$$L_\beta = \sum_k \beta_k^2 \qquad \mu \tag{24}$$

with $p_{\beta_k} \geq 0$, k=1,…,K, t=1,…,T, and where λ_t, η_k and μ are Lagrange multipliers of the corresponding constraints. We will assume that the above specification follows from a specific DGP where $u_t \sim N(0, \sigma_0^2)$.

In order to establish finite bounds on the parameters and the Lagrange multipliers we need to state the KKT conditions of this problem:

$$\frac{\partial \boldsymbol{L}}{\partial p_{\beta_k}} = \log(p_{\beta_k}) + 1 + \eta_k = 0 \tag{25}$$

$$\frac{\partial \boldsymbol{L}}{\partial L_\beta} = \log(L_\beta) + 1 + \mu + \sum_k \eta_k \beta_k^2 / L_\beta^2 = \log(L_\beta) + 1 + \mu + \sum_k \eta_k p_{\beta_k} / L_\beta = 0 \tag{26}$$

$$\frac{\partial \boldsymbol{L}}{\partial \beta_k} = -\sum_t x_{tk} \lambda_t - 2\mu\beta_k - 2\eta_k \beta_k^2 / L_\beta = 0 \tag{27}$$

$$\frac{\partial \boldsymbol{L}}{\partial u_t} = 2u_t - \lambda_t = 0\,. \tag{28}$$

Now we assume that, for any randomly selected sample of data, a feasible solution exists for both the primal and dual problems. This means that the Leuven-1 estimator has an optimal solution and all the unknown variables are bounded away from infinity. Then, from (23) and (24) we have $\sum_{k=1}^K p_{\beta_k} = \sum_{k=1}^M \beta_k^2 / \sum_{j=1}^K \beta_j^2 = 1$ while, from (25), each probability p_{β_k} is strictly positive since $p_{\beta_k} = e^{-(1+\eta_k)} > 0$ and since the Lagrange multiplier η_k is bounded by the assumption of a feasible primal problem. Hence, we conclude that

$$1 > p_{\beta_k} > 0 \tag{29}$$

for each $k = 1,...,K$. Using (23) again, we also conclude that each term β_k^2 cannot be equal to zero and cannot assume the value of infinity because either event violates relation (29). The second part of this result is equivalent to an upper bound on the parameter L_β.

Having established finite bounds on every component of the Leuven-1 estimator, we are ready to take the probability limit for $T \to \infty$ of the entropy criterion (21) and prove the proposition that

$$\underset{T\to\infty}{p\lim}\, T^{-1} H^T(\mathbf{y}^T, \boldsymbol{\beta}^T, L_\beta^T, \mathbf{p}_\beta^T) = \underset{T\to\infty}{p\lim}\, T^{-1} SSR^T(\mathbf{y}^T, \boldsymbol{\beta}^T) \tag{30}$$

where $SSR^T(\mathbf{y}^T, \boldsymbol{\beta}^T) = \sum_t^T (y_t - \sum_k^K x_{tk}\beta_k)^2$ represents the sum of squared residuals of the linear model (22). The superscript "T" on every argument of (30) indicates its dependence on the sample size T. We thus have

$$\underset{T\to\infty}{p\lim}\, T^{-1} \sum_{k=1}^K p_k^T \log(p_k^T) = 0$$

$$\underset{T\to\infty}{p\lim}\, T^{-1} L_\beta^T \log(L_\beta^T) = 0 \tag{31}$$

$$\underset{T\to\infty}{p\lim}\, T^{-1} \sum_{t=1}^T u_t^2 = \sigma_0^2$$

This result demonstrates that the probability limit of the entropy objective function (21) converges to the limiting value of the objective function of the OLS estimator, QED. Thus, the asymptotic properties of the OLS estimator carry over to the Leuven-1 estimator. A similar development can be elaborated for the Leuven-2 estimator.

RIASSUNTO

Stimatori di Leuven di massima entropia e multicollinearità

Si presenta una nuova classe di stimatori, denominati stimatori di Leuven di massima entropia (MEL). Tali stimatori sono consistenti e asintoticamente normali. Essi sono ispirati alla teoria della luce. Gli stimatori MEL rivaleggiano con lo stimatore di massima entropia generalizzata (GME). Usando il criterio dello scarto quadratico medio, tali stimatori sono superiori allo stimatore dei minimi quadrati quando l'informazione campionaria è affetta da multicollinearità.

SUMMARY

Maximum entropy Leuven estimators and multicollinearity

A novel class of estimators, called maximum entropy Leuven (MEL) estimators, is presented and its performance is illustrated by Monte Carlo experiments. These estimators are inspired by the theory of light. The MEL estimators are consistent and asymptotically normal. They rival the generalized maximum entropy estimator (GME). Based on the mean squared error criterion, the MEL estimators outperform the ordinary least-squares estimator in the presence of multicollinerity.