Department of Agricultural and Resource Economics
University of California Davis

MELE: Maximum Entropy Leuven Estimators

by

Quirino Paris

April 2001

Working Paper 01-003

California Agricultural Experiment Station
Giannini Foundation for Agricultural Economics

MELE: Maximum Entropy Leuven Estimators

Quirino Paris

Leuven, Belgium, April 6, 2001

Abstract

Multicollinearity hampers empirical econometrics. The remedies proposed to date suffer from pitfalls of their own. The ridge estimator is not generally accepted as a vital alternative to the ordinary least-squares (OLS) estimator because it depends upon unknown parameters. The generalized maximum entropy (GME) estimator of Golan, Judge and Miller depends upon subjective exogenous information that affects the estimated parameters in an unpredictable way. This paper presents novel maximum entropy estimators inspired by the theory of light that do not depend upon any additional information. Monte Carlo experiments show that they are not affected by any level of multicollinearity and dominate OLS uniformly. The Leuven estimators are consistent and asymptotically normal.

1

MELE: Maximum Entropy Leuven Estimators

Quirino Paris

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 my apartment looking onto Ladeuze Plein. A characteristic of my age is that, often, it comes with cataracts, and 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 ElectroDynamics (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 (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 thick-

ness, 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."

I have always been envious of physicists because they have a nice story to tell and because the story is so rich although the premises are so simple. Although it is too late for me to become a physicist, it is not too late for attempting to emulate their reasoning. 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 with cataracts, 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. Also economic data carry information about economic environments and the decision processes that generated those data. As with any picture, the information reaching a researcher may correspond to an image that is out-of-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. These estimators are maximum entropy estimators and are named after the town of Leuven that has played such an important part in my life and in the life of my wife Carlene.

In the next sections, two MELE specifications will be presented. For convenience, they will be numbered 1 and 2. They are all maximum entropy estimators. 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 of Golan, Judge and Miller 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.'s (1980) recommendation. The Leuven estimators outperform the OLS estimator for all values of the condition number examined in several Monte Carlo experiments. They also outperform the GME estimator when a large intercept is part of the econometric model. 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.

Leuven-1 and Leuven-2 Estimators

The Leuven-1 and Leuven-2 estimators share the same entropy structure. Their difference 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):

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

where the dimensions of the various components are $\mathbf{y} \sim (T \times 1)$, $\mathbf{u} \sim (T \times 1)$, $\beta \sim (K \times 1)$ 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.

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:

$$(2) \qquad \min H(\mathbf{p}, L, \mathbf{u}) = \sum_k p_k \log(p_k) + L \log(L) + \sum_t u_t^2$$

subject to

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

$$L = \sum_k \beta_k^2$$

$$p_k = \beta_k^2 / L$$

with $p_k \ge 0$, $k=1,...,K$, $t=1,...,T$. The amplitude (or normalized dimensionality) of parameter β_k is given by $\beta_k / \sqrt{L}$, hence the probability of parameter β_k is given by the square of its amplitude, as in the theory of light. The term $L \log(L)$ in the objective function prevents the overflow of the L 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. Although this structure is relatively simple, it admits a series of sophisticated variations, as discussed in subsequent sections.

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

$$(3) \qquad \min H(\mathbf{p}, L, \mathbf{u}) = \mathbf{p}' \log(\mathbf{p}) + L \log(L) + \mathbf{u}'\mathbf{u}$$

subject to

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

$$L = \beta'\beta$$

$$\mathbf{p} = \beta \odot \beta / L$$

where $\mathbf{p} \ge \mathbf{0}$ and the symbol $\odot$ indicates the element-by-element Hadamard product.

In order to better clarify the nature of the probability relations, we must notice that $\mathbf{p} = \beta \odot \beta / L$ is not a restriction as in the traditional sense since a minimization of the sum of squared residuals $\mathbf{u}'\mathbf{u}$ subject to the same relations of problem (3) will pro-

duce the unrestricted least-squares estimator. Traditionally, in other words, a set of parameter restrictions are appended to the model without changing the objective function. On the contrary, in the Leuven-1 estimator the "restrictions" make sense only if accompanied by the entropy part of the objective function and vice versa. In other words, these probability relations are simply definitions that characterize the Leuven-1 estimator. Surely, these probability relations could have been hypothesized to have a different functional expression in terms of the β parameters but here is where the inspiration to the theory of light provides an intuitive story involving economic information. In the end, whether or not the specification of the probability relations according to the theory of light will have any econometric relevance will depend upon the empirical performance of the Leuven estimators when confronted with many "real life" samples of economic data.

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

(4) $\quad L = \mathbf{p}\, \log(\mathbf{p}\,) + L\, \log(L\,) + \mathbf{u}\,\mathbf{u} + \lambda\,(\mathbf{y} - \mathbf{X}\beta - \mathbf{u}) + \mu(L\, - \beta\,\beta) + \eta\,(\mathbf{p}\, - \beta\,\,\beta/L\,)$

where the symbols λ, μ, η are the Lagrange multipliers of the corresponding constraints.

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

(5)
$$\frac{L}{\mathbf{p}} = \log(\mathbf{p}\,) + \iota_K + \eta = \mathbf{0}$$
$$\frac{L}{L} = \log(L\,) + 1 + \mu + \eta\,\beta\,\,\beta/L^2 = \log(L\,) + 1 + \mu + \eta\,\mathbf{p}\,/L\, = 0$$
$$\frac{L}{\beta} = -\mathbf{X}\,\lambda - 2\mu\beta - 2\eta\,\,\beta/L\, = \mathbf{0}$$
$$\frac{L}{\mathbf{u}} = 2\mathbf{u} - \lambda = \mathbf{0}$$

where the symbol ι_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 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 esti-mator.

The Leuven-2 estimator extends the probability specification to the error term **u** resulting in the following symmetric structure:

(6) $\min H(\mathbf{p}, L, \mathbf{p}_u, L_u) = \sum_k p_k \log(p_k) + \sum_t p_{u_t} \log(p_{u_t}) + L \log(L) + L_u \log(L_u)$

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

$L = \sum_k \beta_k^2$

$p_k = \beta_k^2 / L$

$L_u = \sum_t u_t^2$

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

with $p_k \geq 0$ and $p_{u_t} \geq 0$, $k=1,...,K$, $t=1,...,T$. Except for the probability elaboration of the error term, the Leuven-2 estimator shares the same structure and characteristics of the Leuven-1 estimator. In matrix notation the Leuven-2 estimator assumes the following form:

(7) $\min H(\mathbf{p}, L, \mathbf{p}_u, L_u) = \mathbf{p}\log(\mathbf{p}) + L\log(L) + \mathbf{p}_u\log(\mathbf{p}_u) + L_u\log(L_u)$

subject to $\mathbf{y} = \mathbf{X}\beta + \mathbf{u}$

$L = \beta'\beta$

8

$$\mathbf{p} = \beta \quad \beta / L$$

$$L_u = \mathbf{u} \, \mathbf{u}$$

$$\mathbf{p}_u = \mathbf{u} \quad \mathbf{u} / L_u$$

with $\mathbf{p} \quad \mathbf{0}$ and $\mathbf{p}_u \quad \mathbf{0}$. Again, the Leuven-2 estimator does not require any subjective exogenous information as does the GME estimator.

The Class of MELE as Rival to the GME and OLS Estimators

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

(8)
$$_k = \sum_m Z_{km} p_{km}$$

where $p_{km} \quad 0$ are probabilities and, of course, $\sum_m p_{km} = 1$ 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 present a thorough discussion of the Generalized Maximum Entropy (GME) estimator. Let $\beta = \mathbf{Zp}$, with $p_{km} \quad 0$ and $\sum_m p_{km} = 1$ for $k = 1,...,K$. Also, let $\mathbf{u} = \mathbf{Vw}$, with $w_{tg} \quad 0$ and $\sum_g w_{tg} = 1$ for $t = 1,...,T$. Then, the GME estimator can be stated as

(9)
$$\max H(\mathbf{p}, \mathbf{w}) = - \sum_{k,m} p_{km} \log(p_{km}) - \sum_{t,g} w_{tg} \log(w_{tg})$$

subject to
$$\mathbf{y} = \mathbf{XZp} + \mathbf{Vw}$$

$$\sum_m p_{km} = 1 \qquad k = 1,...,K$$

$$\sum_g w_{tg} = 1 \qquad t = 1,...,T.$$

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

The GME estimator, however, has important weaknesses for which the class of MELE 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 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 MELE 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.

Uniqueness of the Estimates (Estimator Identification by the Sample Data)

The Leuven-1 estimator is composed by a strictly convex objective function, a convex linear model, the length of the parameter vector which is a convex function, and a relation between probabilities and parameters which is neither concave nor convex. It is this last relation that makes it difficult to decide whether the Leuven estimators have a unique global solution. A limited number of empirical examples have been analyzed by restarting the solution routine with many different starting values. In particular, a grid for ten parameters of a Monte Carlo experiment described below was explored with initial values that ranged from –100 to +100 (when the true value of the parameters ranged between -2 and +2) using a step size equal to 5. In all the runs, the Leuven estimators converged to the same solution. These results justify the conjecture that the Leuven estimators have a unique global solution given an arbitrary sample of data.

Distributional Properties of MELE

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 $\rho(\beta,\hat{\beta})$, also called mean squared error loss (MSEL), as suggested by Judge et al. (p. 558), where

$$(10) \quad \rho(\beta,\hat{\beta}) = trMSE(\beta,\hat{\beta}) = trE[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] = E[(\hat{\beta} - \beta)'(\hat{\beta} - \beta)]$$

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

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}, $k = 1,...K$, 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 [-10,0,10], respectively. The condition number (CN) (see Belsley et al.) 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 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 (1980) 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. Equivalently, the same condition number can be obtained by computing the singular value decomposition of a normalized matrix, $norm\mathbf{X}$, such that $norm\mathbf{X} = \mathbf{UDV}'$, where $\mathbf{D}$ is a diagonal matrix of singular values while $\mathbf{U}$ and $\mathbf{V}$ are matrices such that $\mathbf{U}'\mathbf{U} = \mathbf{I}$ and $\mathbf{V}'\mathbf{V} = \mathbf{I}$. The condition number of the $norm\mathbf{X}$ matrix, then, is the ratio between the maximum and the minimum singular values measured as absolute values. Because $norm\mathbf{X}'\mathbf{X} = \mathbf{VDU}'\mathbf{UDV}' = \mathbf{VD}^2\mathbf{V}' = \mathbf{VLV}'$, with eigenvalues $\mathbf{L} = \mathbf{D}^2$, the two definitions of condition number are consistent. Belsley et al. 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.

14

Table 3. Monte Carlo Experiment N. 1: Model without intercept. MSEL of rival estimators for an increasing condition number. [1]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

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 intervals of the parameters. 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.

Scaling Properties of MELE

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 scale-invariance property of the OLS estimator is well-known. That the same property might be shared by nonlinear estimators may be a novel result.

The lack of a closed form solution for the Leuven estimators requires a discussion of the relevant KKT conditions. The main line of reasoning runs as follows: if the KKT conditions corresponding to two different and arbitrary scaling schemes of the sample information 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) and will be reproduced below for convenience:

$$\frac{\partial L}{\partial \mathbf{p}} = \log(\mathbf{p}) + \iota_K + \eta = \mathbf{0}$$

$$\frac{\partial L}{\partial L} = \log(L) + 1 + \mu + \eta \beta \quad \beta/L^2 = \log(L) + 1 + \mu + \eta \mathbf{p}/L = 0$$

(11)

$$\frac{\partial L}{\partial \beta} = -\mathbf{X}\lambda - 2\mu\beta - 2\eta \quad \beta/L = \mathbf{0}$$

$$\frac{\partial L}{\partial \mathbf{u}} = 2\mathbf{u} - \lambda = \mathbf{0}$$

Notice that the fourth equation involving only the variables $\mathbf{u}$ and λ establishes the symmetric duality between error terms and the Lagrange multipliers of the linear statistical model: in unscaled models, therefore, the Lagrange multipliers 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 (11) can be rewritten as

(12) $-\mathbf{X}\mathbf{u} - \mu\beta - \eta \quad \beta/L = \mathbf{0}$.

We will regard the first two equations of system (11) plus equation (12) 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:

(13)

$$\frac{\mathbf{y}}{R} = \frac{\mathbf{X}}{R}\mathbf{S}^{-1} \quad \mathbf{S}\beta + \frac{\mathbf{u}}{R}$$

$$\frac{\mathbf{y}}{R} = \frac{\mathbf{X}}{R}\mathbf{S}^{-1} \quad \beta^* + \mathbf{u}^*$$

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

17

(14) $\min H(\mathbf{p}\ , L\ , \mathbf{u}) = \mathbf{p}\ \log(\mathbf{p}\ ) + L\ \log(L\ ) + R^2 \mathbf{u}^*\ \mathbf{u}^*$

subject to $\dfrac{\mathbf{y}}{R}\ = \dfrac{\mathbf{X}}{R}\mathbf{S}^{-1}\ \beta^* + \mathbf{u}^*$

$L\ = \beta^*\ \mathbf{S}^{-1}\ \mathbf{S}^{-1}\beta^*$

$\mathbf{p}\ = \mathbf{S}^{-1}\beta^*\ \ \mathbf{S}^{-1}\beta^* \ / L\ .$

If the scalar R is equal to one and the matrix $\mathbf{S}$ is taken as the identity matrix, the model specified in equation (14) is identical to the model exhibited in equation (3). We need to show that the KKT conditions of model (14) 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 (11) plus equation (12).

After setting up the Lagrangean function corresponding to model (14), the KKT condition are derived as follows:

$\dfrac{L}{\mathbf{p}} = \log(\mathbf{p}\ ) + \iota_K + \eta = \mathbf{0}$

$\dfrac{L}{L} = \log(L\ ) + 1 + \mu + \eta\ \mathbf{S}^{-1}\beta^*\ \ \mathbf{S}^{-1}\beta^* \ / L^2 = \log(L\ ) + 1 + \mu + \eta\ \mathbf{p}\ / L\ = 0$

(15)

$\dfrac{L}{\beta^*} = -\mathbf{S}^{-1}\dfrac{\mathbf{X}}{R}\lambda - 2\mu\mathbf{S}^{-1}\ \mathbf{S}^{-1}\beta^* - 2\mathbf{S}^{-1}\ \eta\ \ \mathbf{S}^{-1}\beta^* \ / L\ = \mathbf{0}$

$\dfrac{L}{\mathbf{u}^*} = R^2 2\mathbf{u}^* - \lambda = \mathbf{0}$

Now, let us assume that the vector $(\hat{\mathbf{u}},\hat{\beta}, \hat{L}\ ,\hat{\mathbf{p}}\ ,\hat{\mu},\hat{\lambda},\hat{\eta})$ represents a solution of the system of KKT conditions (15). 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 (11) and equation (12). First of all, the first two equations of system (15) have a structure that is identical to the structure of the first two equations of system (11). Hence, the values of $\hat{\mathbf{p}}\ ,\hat{\eta},\hat{L}$ and $\hat{\mu}$ that satisfy the first two equations of system (15) by assumption, satisfy

18

also the first two equations of system (11). We can thus state that $\hat{\hat{\mathbf{p}}} = \hat{\mathbf{p}}$, $\hat{\hat{L}} = \hat{L}$, $\hat{\hat{\eta}} = \hat{\eta}$ 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 (15) can then be re-stated as

$$(16) \qquad R^2 2\hat{\mathbf{u}} - \hat{\lambda} = R^2 2\frac{\hat{\hat{\mathbf{u}}}}{R} - \hat{\lambda} = R2\hat{\hat{\mathbf{u}}} - \hat{\lambda} = \mathbf{0}$$

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, $2\mathbf{u} = \lambda$. We thus have $\hat{\hat{\lambda}} = \hat{\lambda}/R$. The solution value of the Lagrange multiplier $\hat{\lambda}$ can be replaced by its equivalent expression [equation (16)] in the third equation of system (15) after pre-multiplying it by the matrix $\mathbf{S}$ to obtain

$$(17) \qquad -\frac{\mathbf{X}}{R} R2\hat{\hat{\mathbf{u}}} - 2\hat{\mu}\mathbf{S}^{-1}\hat{\beta} - 2\hat{\eta} \, \mathbf{S}^{-1}\hat{\beta}/\hat{L} = \mathbf{0}.$$

Finally, equation (17) reduces to equation (12) after using the identity $\beta^* \equiv \mathbf{S}\beta$ from which we can obtain an estimate of the unscaled parameter β as $\hat{\hat{\beta}} = \mathbf{S}^{-1}\hat{\beta}$. To be explicit,

$$(18) \qquad -\mathbf{X}\hat{\hat{\mathbf{u}}} - \hat{\hat{\mu}}\hat{\hat{\beta}} - \hat{\hat{\eta}} \, \hat{\hat{\beta}}/\hat{\hat{L}} = \mathbf{0}$$

by making use also of the equalities dealing with parameters L, η and μ as stated above. Equation (18) has the same structure of equation (12) 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 regres-

sors (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:

(19) $\qquad \min H(\mathbf{p}, L, \mathbf{p}_u, L_u) = \mathbf{p}\ \log(\mathbf{p}\ ) + L\ \log(L\ ) + \mathbf{p}_u\ \log(\mathbf{p}_u) + L_u \log(L_u)$

subject to $\qquad\qquad\qquad \dfrac{\mathbf{y}}{R} = \dfrac{\mathbf{X}}{R}\mathbf{S}^{-1}\ \beta^{*} + \mathbf{u}^{*}$

$$L = \beta^{*}\,\mathbf{S}^{-1}\,\mathbf{S}^{-1}\beta^{*}$$

$$\mathbf{p} = \mathbf{S}^{-1}\beta^{*}\ \mathbf{S}^{-1}\beta^{*} / L$$

$$L_u = R^2 \mathbf{u}^{*}\ \mathbf{u}^{*}$$

$$\mathbf{p}_u = R^2 \mathbf{u}^{*}\ \mathbf{u}^{*} / L_u.$$

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.

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 inter-cept. 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

20

sample information is measured in natural units. The relevant KKT conditions of this latter model are given, again, by system (11).

In order to set up a model defined in deviations from the mean, it is convenient to define a "deviator" operator $\mathbf{D} \equiv \mathbf{I}_T - \frac{\iota_T \iota_T'}{T}$ that will generate a dependent variable and regressors in deviations from their respective means. The vector ι_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

(20)
$$\min H(\mathbf{p}, L, \mathbf{u}) = \mathbf{p}' \log(\mathbf{p}) + L' \log(L) + \mathbf{u}'\mathbf{u}$$

subject to
$$\mathbf{Dy} = \mathbf{DX}\beta + \mathbf{Du}$$

$$L = \beta'\beta$$

$$\mathbf{p} = \beta \odot \beta / L .$$

The relevant KKT conditions of problem (20) are given by

(21)
$$\frac{\partial L}{\partial \mathbf{p}} = \log(\mathbf{p}) + \iota_K + \eta = \mathbf{0}$$

$$\frac{\partial L}{\partial L} = \log(L) + 1 + \mu + \eta'\beta \odot \beta / L^2 = \log(L) + 1 + \mu + \eta'\mathbf{p}/L = 0$$

$$\frac{\partial L}{\partial \beta} = -\mathbf{X}'\mathbf{D}\lambda - 2\mu\beta - 2\eta \odot \beta / L = \mathbf{0}$$

$$\frac{\partial L}{\partial \mathbf{u}} = 2\mathbf{u} - \mathbf{D}\lambda = \mathbf{0}$$

Now, let us assume that the vector $(\hat{\mathbf{u}}, \hat{\beta}, \hat{L}, \hat{\mathbf{p}}, \hat{\mu}, \hat{\lambda}, \hat{\eta})$ represents a solution of the system of KKT conditions (21). By replacing $\mathbf{D}\lambda$ in the third equation of system (21) by its equivalent expression given in the fourth equation of (21), the KKT conditions of the model in deviations from the mean have a structure that is identical to the KKT conditions (11). Hence, the solution $(\hat{\mathbf{u}}, \hat{\beta}, \hat{L}, \hat{\mathbf{p}}, \hat{\mu}, \hat{\lambda}, \hat{\eta})$ of system (21) will satisfy also system (11).

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 variable 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 $\mathbf{p}_G, p_1, L, \beta_G, \lambda_1, \mathbf{u}$, where λ_1 is the intercept and β_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 $\mathbf{p}_G, p_1, L, \beta_G, \mathbf{u}$). 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 λ_1.

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

$$(22) \qquad \min H(\mathbf{p}, L, \mathbf{u}) = \sum_{k=2}^{K} p_k \log(p_k) + L \log(L) + \sum_{t} u_t^2$$

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

$$L = \sum_{k=2}^{K} \beta_k^2$$

$$p_k = \beta_k^2 / L$$

with $p_k \geq 0$, $k=2,...,K$, $t=1,...,T$. As we will illustrate by means of a second Monte Carlo experiment, the Leuven-1 estimator specified in (22) 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}, $k = 1,...K$, and β_{01} is considered the intercept with a true value of 15. Each remaining parameter β_{0k}, $k = 2,...,K$, was drawn from a uniform distribution

$U[-2,3]$. Each element of the matrix of regressors **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 **u** was drawn from a normal distribution $N(0, \sigma_0^2) = N(0,4)$. With this specification, the dependent variable **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) (see Belsley et al.) of the **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.[1] 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 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.

Conclusion

The class of MEL estimators is inspired by the theory of light and rivals the GME estimator of Golan et al. 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. 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 $_k$, K probabilities p_k, and T error terms u_t). The solution of the GME estimator for a similar model has $(MK+JT)$ components, where M is the number of discrete supports for the parameter $_k$

and *J* is the number of discrete supports for the error term u_t. The empirical GME literature indicates that, in general, *M*=5 and *J*=3.

Another distinctive feature of the Leuven estimators regards the parameter probabilities that, in general, do not approach the uniform distribution as do the corresponding probabilities of the GME estimator. In order to illustrate this proposition, the parameter and probability estimates for one data sample of the Monte Carlo experiment N. 1 described above are reported in Table 5. The condition number for the **X** matrix of this sample is equal to 1018. There are ten parameters with true values as reported in the first column. The GME estimator was implemented with three support values for the parameters and a support interval of [-20,0,20]. The three support values for each error term were selected as [-10,0,10].

Table 5. Estimates of parameters and probabilities in rival estimators

		Leuven-1		GME[-20,20]				OLS
	True Beta	Beta	Prob(Beta)	Beta	Prob1	Prob2	Prob3	Beta
1	-0.0258	-0.0301	0.0001	-0.0208	0.3339	0.3333	0.3328	-0.0048
2	-1.0752	-0.8594	0.0957	-0.9108	0.3564	0.3328	0.3108	-0.9552
3	0.4149	0.8509	0.0937	0.8469	0.3124	0.3329	0.3547	0.8855
4	1.4772	1.1832	0.1815	1.2400	0.3028	0.3324	0.3648	1.2203
5	-1.5673	-1.1938	0.1847	-0.9819	0.3582	0.3327	0.3091	7.5817
6	-0.3852	-0.5096	0.0337	-0.7394	0.3520	0.3330	0.3150	-9.2844
7	-0.4499	-0.8148	0.0861	-1.0607	0.3602	0.3326	0.3072	-9.5962
8	0.1000	-0.1732	0.0039	-0.1822	0.3379	0.3333	0.3288	-0.2112
9	-0.7403	-0.7781	0.0785	-0.7923	0.3535	0.3329	0.3136	-0.7938
10	1.5974	1.3662	0.2420	1.4385	0.2980	0.3320	0.3699	1.4385

As anticipated, the probabilities of the GME estimator tend toward the uniform distribution with all values very near to 1/3. On the contrary, the probability values of the Leuven-1 estimator are far from the uniform distribution. Because of the presence of a high level of multicollinearity, some of the OLS parameter estimates are very far from the true values.

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.

Footnotes

1. Golan, Judge and Miller 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, 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. 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.

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 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.

References

Belsley, D.A., E. Kuh, and R.E. Welsch. *Regression Diagnostics*. John Wiley, New
York, 1980.

Bera, A. and C. Jarque. "Efficient Test for Normality, Heteroscedasticity and Serial
Independence of Regression residuals: Monte Carlo Evidence." Economics
Letters 7,1981, pp.313-318.

Brooke, A., D. Kendrick, and A. Meeraus. *GAMS: A User's Guide*. The Scientific
Press, 1988.

Caputo, M. and Q. Paris, "Comparative Statics of the Generalized Maximum Entropy
Estimator of the General Linear Model," Department of Agricultural and
Resource Economics, University of California, Davis, June 2000, 25 pages.

Feynman, Richard, *QED: The Strange Theory of Light and Matter*, Princeton
University Press, Princeton, New Jersey, 1985.

Golan, Amos, George Judge and Douglas Miller, *Maximum Entropy Econometrics*,
Wiley and Sons, Chichester, UK, 1996.

Jaynes, E. T., "Information Theory and Statistical Mechanics," *Physics Review*
106(1957):620-630, Part II, 108(1957):171-190.

Judge, G.G., T.C. Hill, W.E. Griffiths, H. Lütkepohl, T.C. Lee. *Introduction to the
Theory and Practice of Econometrics*. John Wiley, New York, 1982.

Van Akkeren, M. and G.G. Judge. "*Extended Empirical Likelihood Estimation and
Inference*" University of California, Berkeley, 1999, 49 pages.

Appendix: Distributional Properties of the Leuven-1 Estimator

The goal of this appendix is to prove that the Leuven-1 estimator is consistent and as-ymptotically normal. The strategy is based upon the realization that, in the limit (that is, for a sample size that tends to infinity), 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. In order to see this result clearly it is sufficient to realize that the following model (where the entropy components of the Leuven-1 estimator have been removed)

A.0 $\qquad\qquad\qquad$ $\min\{\mathbf{u}\,\mathbf{u}\}$

subject to $\qquad\qquad\qquad$ $\mathbf{y} = \mathbf{X}\beta + \mathbf{u}$

$\qquad\qquad\qquad\qquad$ $L\; = \beta\,\beta$

$\qquad\qquad\qquad\qquad$ $\mathbf{p}\; = \beta\;\;\beta\,/\,L$

is nothing else but a rather baroque way of stating the unrestricted 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 suf-ficient 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 func-tion 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 re-sult 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:

A.1
$$\min H(\mathbf{p}, L, \mathbf{u}) = \sum_k p_k \log(p_k) + L \log(L) + \sum_t u_t^2$$

subject to

A.2
$$y_t = \sum_k x_{tk}\beta_k + u_t \qquad \forall t$$

A.3
$$p_k = \beta_k^2 / L \qquad \forall k$$

A.4
$$L = \sum_k \beta_k^2 \qquad \mu$$

with $p_k \geq 0$, $k=1,...,K$, $t=1,...,T$, and where λ_t, θ_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:

A.5
$$\frac{\partial L}{\partial p_k} = \log(p_k) + 1 + \theta_k = \mathbf{0}$$

A.6
$$\frac{\partial L}{\partial L} = \log(L) + 1 + \mu + \sum_k \theta_k \beta_k^2 / L^2 = \log(L) + 1 + \mu + \sum_k \theta_k p_k / L = 0$$

A.7
$$\frac{\partial L}{\partial \beta_k} = -\sum_t x_{tk}\lambda_t - 2\mu\beta_k - 2\theta_k\beta_k^2 / L = 0$$

A.8
$$\frac{\partial L}{\partial u_t} = 2u_t - \lambda_t = 0$$

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 (A.3) and (A.4) we have $\sum_k p_k = \sum_k \lambda_k^2 / \sum_j \lambda_j^2 = 1$ while, from (A.5), each probability p_k is strictly positive since $p_k = e^{-(1+\mu_k)} > 0$ and since the Lagrange multiplier μ_k is bounded by the assumption of a feasible primal problem. Hence, we conclude that

A.9 $\qquad\qquad 1 > p_k > 0$

for each $k = 1,...,K$. Using (A.3) again, we also conclude that each term λ_k^2 cannot be equal to zero and cannot assume the value of infinity because either event violates relation (A.9). The second part of this result is equivalent to an upper bound on the parameter L.

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 (A.1) and prove the proposition that

A.10 $\qquad\qquad p\lim_{T} T^{-1} H^T(\mathbf{y}^T, \boldsymbol{\beta}^T, L^T, \mathbf{p}^T) = p\lim_{T} T^{-1} SSR^T(\mathbf{y}^T, \boldsymbol{\beta}^T)$

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 (A.2). The superscript "T" on every argument of (A.10) indicates its dependence on the sample size T. We thus have

$$p\lim_{T} T^{-1} \sum_{k=1}^{K} p_k^T \log(p_k^T) = 0$$

A.11 $\qquad\qquad p\lim_{T} T^{-1} L^T \log(L^T) = 0$

$$p\lim_{T} T^{-1} \sum_{t=1}^{T} u_t^2 = \sigma_0^2$$

This result demonstrates that the probability limit of the entropy objective function (A.1) 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.