North-Eastern Hill University, Shillong (India)
Department of Economics : Working Papers Series

Multicollinearity and Modular Maximum Entropy Leuven Estimator

Dr. SK Mishra

An earlier version of this paper was published in
Economics Bulletin, Vol 3 (25), 2004

Department of Economics
North-Eastern Hill University
Shillong, Meghalaya – 793022 (India)

Multicollinearity and Modular Maximum Entropy Leuven Estimator

Dr. SK Mishra
Dept. of Economics
NEHU, Shillong, India

1. Introduction: A high degree of multicollinearity often has detrimental effects on the estimation of regression coefficients, β, of a linear model $y = X\beta + u.$ The *Ordinary Least Squares* (OLS) estimate of β, $\hat{\beta}_{OLS} = (X'X)^{-1}X'y$, is often inaccurate, usually far from the true coefficients, if $X'X$ exhibits a high degree of multicollinearity. This is due to an intricate internecine sharing among the estimated coefficients.

2. Measures of the Degree of Multicollinearity: Let the explanatory variables, $X(n,m)$, in the model $y = X\beta + u$ be measured such that each of its columns has a zero mean and unit standard deviation. In that case, $X'X = nR$, where R is the intercorrelation matrix and r_{ij} is the cosine of the angle between x_i and x_j vectors. Ideally, the $X'X$ matrix should be diagonal. That signifies a total absence of multicollinearity. However, this is far from the real world situation. Since the cosine of an angle must lie between –1 and 1, multicollinearity reaches its zenith when any one or more off-diagonal element(s) of R is (are) $\pm$ 1.

In a multivariate situation there can be several possible measures of multicollinearity. Neumann & Goldstine (1947) suggested a measure of near-singularity of a square matrix (say, A of order m x m). The measure is $M = abs(\lambda_1)/abs(\lambda_m),$ where λ_1 and λ_m are (magnitude wise) the largest and the smallest eigenvalues (respectively) of the matrix A. Although M may be used directly (Kaçiranlar et al., 1999) to measure the degree of multicollinearity, one may use $\mu = +\sqrt{M}$ as the measure. Since μ is a monotonic function of M, both the measures are equally good.

Turing (1948) suggested two alternative measures of near-singularity of a matrix. The first, M(A) = $m(\max_{i,j}|a_{ij}|)(\max_{i,j}|b_{ij}|)$ and the second, N(A)=$m^{-1}\|A\|\|B\|$, where $B = A^{-1}$ and the symbols $|\ .\ |$ and $\|\ .\ \|$ stand for modulus and norm respectively. We may use any norm (absolute, Euclidean, Max, or any other).

Belsley et al. (1980, Chap. 3) measured the severity of multicollinearity by the condition number defined as $c_n = \sqrt{\lambda_1^*/\lambda_m^*}$, where λ_1^* and λ_m^* are the largest and the smallest eigenvalues (respectively) of $(X'X)^*$, that is normed $X'X$ matrix. Normalization of $X'X$ is done in such a manner that the Euclidean norm of each column of the resulting $(X'X)^*$ matrix is unity. In case $X(n,m)$ in the model $y = X\beta + u$ is measured such that each of its columns has a zero mean and unit standard deviation, Belsley's c_n is trivially larger than Neumann's μ defined earlier. One may ask: by how much quantity 'larger' is just '*trivially larger ?*'. The answer is relative and contextual; by that much quantity 'larger' is just '*trivially larger*' which has hardly any bearing on our conclusion regarding the severity of multicollinearity and its undesirable effects on estimation. Note that 'multicollinearity' is a fuzzy concept, dialectical in nature (see Georgescu-Roegen, 1971), which may only imperfectly be measured by any crisp (arithmomorphic) number such as the so called 'condition number.' In the practice of econometrics, the concept of 'condition number' is vague. This is clearly reflected in the literature. Belsley et al. (1980) measure the degree of multicollinearity in one way, Golan et al. (1996) in another way (Paris, 2001, p.1 footnote), Kaçiranlar et al. (1999) in yet another way and so on. However, it is held that multicollinearity begins to signal its deleterious effects when Belsley's condition number is *around* 30. Beyond this number ($c_n >> 30$) multicollinearity destabilizes the estimation and the estimated regression coefficients, $\hat{\beta}_{OLS}$, are grossly unreliable.

3. Remedial Attempts: Hoerl & Kennard (1970) introduced *Ridge Regression* (RR) as a remedial measure to multicollinearity. Ridge Regression numerically perturbs $X'X$ through adding to it a matrix $\delta I \; : \delta > 0$. Thus, $\hat{\beta}_{Ridge} = (X'X + \delta I)^{-1} X'y$. The value of δ is iteratively obtained. As Theobald (1974) pointed out, the choice of δ depends on unknown parameters (β and σ^2 of population) and replacing the unknown population parameters by their sample estimates does not ensure an advantage of RR over OLS.

Sarkar (1992) introduced the Restricted Ridge Regression estimator by grafting the Ridge Regression into the restricted Least Squares method of estimation (Theil, 1971, pp. 43-45). Liu (1993) obtained a new estimator by combining the Stein (1956) estimator with the RR estimator. Kaçiranlar et al. (1999) obtained another new estimator by grafting the Liu estimator into the Restricted Least Squares method of estimation, called the Restricted Liu estimator. All these estimators are the improved Ridge Regression estimators in essence.

Golan et al. (1996) introduced the *Generalized Maximum Entropy* (GME) estimator to resolve the multicollinearity problem. This estimator requires a number of support values supplied subjectively and exogenously by the researcher. The estimates as well as their standard errors depend on those support values. In a real life situation it is too demanding on the researcher to supply appropriate support values, which limits the application of GME.

Paris (2001) introduced the *Maximum Entropy Leuven* (MEL) estimator. It exploits the information available in the sample data more efficiently than the OLS does, and unlike GME estimator, it does not require any additional information to be supplied by the researcher. The MEL estimator maximizing entropy in $\hat{\beta}$ is formulated as (Paris, 2001, p. 3) $\min H(p_\beta, L_\beta, u) = p'_\beta \log(p_\beta) + L_\beta \log(L_\beta) + u'u$, subject to three equality conditions as: $(i)\;\; y = X\beta + u;\;\; (ii) L_\beta = \beta'\beta;\;\; and\;\; (iii) p_\beta = \beta \Theta \beta / L_\beta :\;\; 0 \le p_\beta \le 1$. Here the symbol Θ indicates the element-by-element Hadamard product. Further, $p_{\beta_i} \log(p_{\beta_i}) = 0$ if $p_{\beta_i} = 0$. Of course, p_{β_i} being the probability, $\sum_{i=1}^{m} p_{\beta_i} = 1$. Paris concludes: "*under any level of multicollinearity, MEL estimator uniformly dominates the OLS estimator according to the mean squared error criterion. It rivals also the GME estimator without requiring any subjective additional information*."

4. Objectives of the Present Investigation: Our objectives in this investigation are: (i) to look into the problem of multicollinerarity more closely and (ii) to shed more light on the

MEL estimator through simulation as well as to propose a new estimator. Additionally, we will discuss some computational issues and alternatives also.

5. Multicollinearity, $X'X$ Matrix and $\sigma(u)$: In the literature on multicollinearity we find that the blame of causing trouble with estimation is solely attributed to the structure of X in the linear econometric model $y = X\beta + u$, which structure is reflected into $X'X$ in $\hat{\beta}_{OLS} = (X'X)^{-1}X'y$. However, some role is played by the standard deviation of the residual term as well.

Table 1. Effects of $\sigma(u)$ and Multicollinearity Measure μ on OLS Estimation

$\sigma(u)$	Measure (μ) of multicollinearity	Regression Coefficients Estimated by OLS (Mean of 50 trials)				
		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
0.00	1222.89	10.000	20.000	30.000	40.000	50.000
0.00	32.16	10.000	20.000	30.000	40.000	50.000
0.00	12.02	10.000	20.000	30.000	40.000	50.000
1.00	1229.89	8.884	29.864	48.073	33.719	67.479
1.00	32.16	9.898	20.365	30.486	39.825	50.426
1.00	12.02	9.943	20.206	30.199	39.927	50.116
2.00	1222.89	2.535	54.552	93.735	17.922	108.128
2.00	32.16	9.795	20.731	30.972	39.651	50.852
2.00	12.02	9.320	20.501	31.101	39.879	50.404
5.00	1222.89	-8.662	106.380	189.338	-15.195	195.321
5.00	32.16	9.488	21.826	32.431	39.127	52.130
5.00	12.02	8.301	21.252	32.752	39.697	51.009
10.00	1222.89	-27.324	192.760	348.676	-70.390	340.642
10.00	32.16	8.977	23.653	34.862	38.254	54.259
10.00	12.02	6.602	22.503	35.504	39.393	52.019
20.00	1222.89	-12.306	217.284	391.452	-85.613	399.588
20.00	32.16	7.953	27.305	39.723	36.506	58.578
20.00	12.02	8.857	24.114	33.980	38.538	53.113
30.00	1222.89	-23.460	315.926	572.178	-148.419	574.382
30.00	32.16	6.930	30.958	44.585	34.761	68.777
30.00	12.02	8.285	26.172	35.964	37.808	54.669

In table 1 we present the OLS estimates of β = (10, 20, 30, 40, 50) with three different samples of X giving three different $X'X$ matrices, with μ = 12.02, 32.16 and 1222.89. We have generated $u \sim N(0,\sigma)$ with $\sigma = 0$, 1, 2, 5, 10, 20 and 30. Note that $\sigma = 0$ means that we have not added u to $X\beta$. There is no constant term in the model; each variable has mean = 0 and standard deviation = 1. The sample size is 30. Fifty trials have been made to obtain the results in each case. The results clearly indicate that with an increase in σ the estimates go wild when μ is large. However, for a small value of σ the OLS gives fairly acceptable estimates even if μ is large. On the other hand, for a small μ a larger σ cannot destabilize the estimator. It appears that errors (that introduce inconsistency into the over-determined linear system of equations $y = X\beta$, the strength of which is dependent on $\sigma(u)$) and X (that contains information on the source of variation in the true y or the y net of error) interact to determine $\hat{\beta}$. A large μ implies a weaker power of X in explaining the variations in y, which may yet be effective if $\sigma(u)$ is small enough and vice versa. From this we learn that large μ coupled with a large σ destabilizes the estimator; either of the two in isolation cannot cause much harm. Yet, of the two, μ is more potent in determining the stability of the OLS estimator of regression coefficients. An investigation is also needed as to the origins of quirks and antiquirks (Bertrand & Holder, 1988; Bertrand, 1998) in the intricate interactions between multicollinearity (measured by μ, say) and $\sigma(u)$.

6. Estimation of Regression Coefficients when $X'X$ may have a Large μ or c_n: In the conventional scheme the regression parameters, β in $y = X\beta + u$, are estimated by $\hat{\beta}_{OLS} = (X'X)^{-1}X'y.$ The matrix $X'X$ is inverted by conventional methods such as Gauss-Jordan, Gauss-Seidel, Cholesky's triangular factorization, etc. (Krishnamurthy & Sen, pp. 134-213). Near-singular matrices (that have a large condition number, c_n) are often ill-conditioned to such inversion methods. Once an ill-conditioned matrix is inverted by one of the conventional methods, the $(X'X)^{-1}$ often does not strictly satisfy the relationship $(X'X)^{-1}X'X = I.$

As we know, it is possible to express $X'X$ as VDV', where D and V are the (diagonal) matrix of eigenvalues and (orthogonal) matrix of eigenvectors of $X'X$, respectively. Symmetric matrices that are ill-conditioned to inversion are quite well-conditioned to computing their eigenvalues and eigenvectors. Now, using the celebrated Cayley-Hamilton theorem (Fröberg, 1965, pp. 57-62), $(X'X)^{-1} = VD^{-1}V'$ if all the elements in the principal diagonal of D are non-zero (absence of perfect multicollinearity). In case some elements (at least one, but not all, of course) in the principal diagonal of D are zero (perfect multicollinearity), one may obtain D^{+} (the Moore-Penrose generalized inverse of D, see Theil, 1971. pp. 268-270), which is very simple to compute. Since D is a diagonal matrix, $d_{ii}^{+} = d_{ii}^{-1}$ if $d_{ii} \neq 0$ else $d_{ii}^{+} = 0.$ In this case, $(X'X)^{+} = VD^{+}V'$, with which one may obtain $\hat{\beta}^{*} = (X'X)^{+}X'y.$ If all the principal diagonal elements of D are non-zero, $D^{-1} = D^{+} \Rightarrow \hat{\beta}_{OLS} = \hat{\beta}^{*}.$ We will call this method of estimation as (the generalized) Cayley-Hamilton (CH) method to obtain $\hat{\beta}_{OLS}$. Computationally, this method has the advantages of withstanding against a very large c_n, even if the smallest eigenvalue of $X'X$ is zero, and of being well suited to programming on a computer. Additionally, it yields the measure of multicollinearity, μ, as an inexpensive byproduct.

Alternatively, one may minimize $u'u = f(\beta \mid y, X) = (y - X\beta)'(y - X\beta)$ by some quadratic programming or search algorithm for non-linear programming, such as Nelder & Mead, Hooke & Jeeves, Rosenbrock, Powell, Fletcher & Reeves or Fletcher & Powell methods (Kuester & Mize, 1973; pp. 297-366). Among the search methods, the Random Walk (RW) method is quite flexible, although slow. For a comparative view of performance of the RW and the CH methods of computation, table 2 may be referred to.

$\sigma(u)$	Algorithm	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
10	Random Walk Search	9.43192	23.23360	32.99955	40.31155	52.82256
10	C-H; $(X'X)^{+} \equiv VD^{+}V'$	9.43192	23.23359	32.99954	40.31155	52.82256
20	Random Walk Search	8.86384	26.46719	35.99909	40.62309	55.64511
20	C-H; $(X'X)^{+} \equiv VD^{+}V'$	8.86384	26.46719	35.99909	40.62310	55.64511

Table 2. Relative Performance of Generalized Cayley-Hamilton (C-H) and Random Walk Algorithms

[Sample size (n) = 20; $\sigma(u) = 10, 20$; Condition Number (c_n)=5.12; $\mu = 5.02$]

7. Generation of Multicollinear Explanatory Variables (X): For the investigation at hand we will require to generate multicollinear X, such that $X'X$ has a large μ or c_n. It is easier to obtain $X(n,m)$, in n rows (observations or the sample size) and m columns, such that each variable (column of X) is uniformly distributed within a specified range. However, generation of X with some control over the degree of multicollinearity is quite involved. The following procedures generates X with a high degree of multicollinearity.

A. *Orthogonalization consisting of six steps* - (i) Generate $Z(n,m) \sim U(0,1)$, uniformly distributed random numbers lying between 0 and 1 with a suitable random number generator - n stands for sample size and m for the number of variables. (ii) Standardize Z such that for each variable (column) its mean ($\bar{z}_j; j=1,2,...,m$) is zero and standard deviation ($\sigma(z_j)=\sigma_j$, say ; $j=1,2,...,m$) is unity. (iii) Compute the correlation matrix $R(m,m)$ from Z. (iv) Compute all m eigenvalues (L) and eigenvectors (E) of R. (v) normalize E to yield ω such that $\omega'\omega = L$, a diagonal matrix. Note that eigenvectors are impervious to a multiplication by any non-zero constant. (vi) compute $Q = Z\omega$. The columns of Q are pair-wise orthogonal. Moreover, the variance (σ_j^2) of the jth column (variable) of Q, i.e. q_j, is $\lambda_j = L_{jj}$.

B. *Multicollinearization consisting of three steps* – (i) Choose a suitable positive diagonal matrix L^* such that its elements are in a descending order, the first element (L^*_{11}) is quite large and the last element (L^*_{mm}) is quite small; note that $L^*_{ij} > 0$ *iff* $i=j$, $L^*_{ij}=0$ otherwise; also note that $trace(L^*)=m.$ (ii) Normalize E (obtained in A.(iv) above) to yield ε such that $\varepsilon'\varepsilon = L^*$, (iii) compute $X = Z\varepsilon'$.

The resulting X in B above yields $X'X$ with a large μ or c_n. However, it is important to know that in case the step A(v) is not undergone and $\omega \equiv E$ is used, a mild *multicollinearization* is achieved than if all the said steps in A and B were undergone. On the other hand, if step B(ii) is not undergone, *protogonalization* (no further m*ulticollinearization)* is achieved.

8. A Comparative Study of Performance of OLS and MEL Estimators: Armed with the algorithms and procedures detailed out above, we proceed to conduct some Monte Carlo experiments to study the relative performance of OLS and MEL (Paris, 2001) estimators.

Table 3-A. Relative Performance of OLS and MEL Estimators

[Sample size (n) = 20; Condition Number (c_n)=5.12; μ=5.02]

$\sigma(u)$	Estimate	Estimator	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
$\sigma(u)=0$	$\hat{\beta}$	MEL	5.994	0.810	7.258	30.748	23.971
		MMEL	10.023	18.431	28.566	40.251	48.950
		OLS	10.000	20.000	30.000	40.000	50.000
	RMS	MEL	4.006	19.190	22.742	9.252	26.029
		MMEL	0.023	1.569	1.434	0.251	1.050
		OLS	0.000	0.000	0.000	0.000	0.000
$\sigma(u)=5$	$\hat{\beta}$	MEL	6.005	0.719	7.147	30.739	24.003
		MMEL	10.118	17.937	28.111	40.344	48.809
		OLS	10.095	19.504	29.543	40.092	49.859
	RMS	MEL	4.069	19.294	22.862	9.293	26.007
		MMEL	1.359	3.445	2.872	1.490	2.134
		OLS	1.357	2.811	2.218	1.454	1.782
$\sigma(u)=10$	$\hat{\beta}$	MEL	6.014	0.629	7.035	30.725	24.032
		MMEL	10.213	17.444	27.656	40.436	48.669
		OLS	10.190	19.008	29.087	40.185	49.717
	RMS	MEL	4.275	19.422	23.000	9.399	26.010
		MMEL	2.716	6.081	4.920	2.932	3.783
		OLS	2.715	5.622	4.436	2.907	3.563
$\sigma(u)=15$	$\hat{\beta}$	MEL	6.021	0.539	6.921	30.708	24.057
		MMEL	10.308	16.951	27.202	40.528	48.530
		OLS	10.285	18.513	28.630	40.277	49.576
	RMS	MEL	4.604	19.574	23.155	9.569	26.037
		MMEL	4.074	8.819	7.066	4.381	5.511
		OLS	4.072	8.432	6.654	4.361	5.345
$\sigma(u)=20$	$\hat{\beta}$	MEL	6.028	0.449	6.807	30.688	24.078
		MMEL	10.410	16.540	26.813	40.596	48.436
		OLS	10.381	18.017	28.173	40.370	49.434
	RMS	MEL	5.031	19.750	23.328	9.800	26.088
		MMEL	5.358	11.369	9.081	5.797	7.206
		OLS	5.430	11.243	8.872	5.814	7.126

Note: MEL = Estimator à la Paris (2001); MMEL obtains $p(\beta)$ differently using absolute norm.

We have fixed the sample size to 20, since small sample properties of these estimators have a relatively more practical significance. We have also fixed the model size (m = no. of explanatory variables in the regression model $y = X\beta + u$) to five variables. The disturbance vector, u, is normally distributed with 0 mean and $\sigma(u) = 0$, 5, 10 15 or 20. Choice of

$\sigma(u)=0$ amounts to adding no error term to $X\beta$ that implies an over-determined (since $n>m$) but *consistent* system of linear equations, $y=X\beta$. Multicollinear $X(20,5)$ matrices yielding $X'X$ with different c_n have been generated by the procedure laid down earlier. The estimation of $\hat{\beta}_{OLS}$ has been made by the Cayley-Hamilton method while the estimation of $\hat{\beta}_{MEL}$ has been made by the Random Walk method (Rao, 1978; pp. 252-257). The author wrote his own program (in FORTRAN) for computation needed in this work.

Table 3-B. Relative Performance of OLS and MEL Estimators

[Sample size (n) = 20; Condition Number (c_n)=28.55; μ =27.87]

$\sigma(u)$	Estimate	Estimator	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
$\sigma(u)=0$	$\hat{\beta}$	MEL	-2.564	-2.325	1.909	22.489	16.460
		MMEL	11.083	0.000	16.497	42.853	42.728
		OLS	10.000	20.000	30.000	40.000	50.000
	RMS	MEL	12.564	22.325	28.091	17.511	33.540
		MMEL	1.083	20.000	13.503	2.853	7.272
		OLS	0.000	0.000	0.000	0.000	0.000
$\sigma(u)=5$	$\hat{\beta}$	MEL	-2.567	-2.348	1.835	22.455	16.526
		MMEL	11.001	0.754	16.876	42.678	43.072
		OLS	10.352	17.659	28.445	40.321	49.443
	RMS	MEL	12.586	22.352	28.173	17.561	33.481
		MMEL	2.798	19.458	13.337	3.111	7.277
		OLS	3.251	13.308	8.140	2.857	4.377
$\sigma(u)=10$	$\hat{\beta}$	MEL	-2.570	-2.369	1.760	22.415	16.588
		MMEL	11.059	0.588	16.683	42.645	43.152
		OLS	10.705	15.317	26.891	40.642	48.886
	RMS	MEL	12.646	22.386	28.271	17.650	33.442
		MMEL	5.560	22.033	14.880	4.458	8.466
		OLS	6.501	26.616	16.280	5.714	8.755
$\sigma(u)=15$	$\hat{\beta}$	MEL	-2.572	-2.390	1.685	22.367	16.644
		MMEL	11.446	-0.913	15.730	42.826	42.934
		OLS	11.057	12.976	25.336	40.963	48.330
	RMS	MEL	12.743	22.426	28.384	17.780	33.424
		MMEL	8.179	27.098	17.770	6.347	10.450
		OLS	9.752	39.923	24.420	8.572	13.132
$\sigma(u)=20$	$\hat{\beta}$	MEL	-2.574	-2.409	1.609	22.311	16.696
		MMEL	11.653	-1.409	15.402	42.851	42.982
		OLS	11.410	10.635	23.781	41.284	47.773
	RMS	MEL	12.875	22.474	28.513	17.948	33.424
		MMEL	11.005	32.344	20.626	8.222	12.477
		OLS	13.003	53.231	32.560	11.429	17.510

Note: MEL = Estimator à la Paris (2001); MMEL obtains $p(\beta)$ differently using absolute norm.

Table 3-C. Relative Performance of OLS and MEL Estimators [Sample size (n) = 20; Condition Number (c_n)=92.98; $\mu = 91.48$]							
$\sigma(u)$	Estimate	Estimator	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
$\sigma(u)=0$	$\hat{\beta}$	MEL	-6.215	-1.874	0.577	17.346	12.857
		MMEL	4.982	0.000	14.510	42.225	39.482
		OLS	10.000	20.000	30.000	40.000	50.000
	RMS	MEL	16.215	21.874	29.423	22.654	37.143
		MMEL	5.019	20.000	15.490	2.225	10.518
		OLS	0.000	0.000	0.000	0.000	0.000
$\sigma(u)=5$	$\hat{\beta}$	MEL	-6.222	-1.873	0.507	17.305	12.931
		MMEL	5.601	-0.147	14.541	42.169	39.970
		OLS	11.016	12.672	25.567	40.982	48.331
	RMS	MEL	16.231	21.876	29.501	22.707	37.075
		MMEL	6.227	20.173	15.663	2.871	10.572
		OLS	7.566	41.834	24.434	6.646	11.231
$\sigma(u)=10$	$\hat{\beta}$	MEL	-6.226	-1.872	0.438	17.256	13.000
		MMEL	7.159	-3.759	12.739	42.593	39.904
		OLS	12.032	5.345	21.135	41.963	46.662
	RMS	MEL	16.259	21.884	29.595	22.793	37.026
		MMEL	9.074	25.335	18.385	4.854	12.015
		OLS	15.132	83.668	48.867	13.292	22.462
$\sigma(u)=15$	$\hat{\beta}$	MEL	-6.226	-1.870	0.369	17.197	13.063
		MMEL	8.487	-8.084	10.407	43.143	39.444
		OLS	13.048	-1.983	16.702	42.945	44.993
	RMS	MEL	16.301	21.897	29.705	22.913	36.996
		MMEL	13.538	31.059	21.434	7.031	14.323
		OLS	22.698	125.502	73.301	19.938	33.693
$\sigma(u)=20$	$\hat{\beta}$	MEL	-6.223	-1.868	0.300	17.128	13.121
		MMEL	9.336	-10.169	9.326	43.386	39.423
		OLS	14.064	-9.311	12.270	43.927	43.324
	RMS	MEL	16.355	21.914	29.830	23.067	36.984
		MMEL	18.259	34.072	23.238	8.924	16.758
		OLS	30.265	167.337	97.734	26.585	44.924
Note: MEL = Estimator à la Paris (2001); MMEL obtains $p(\beta)$ differently using absolute norm.							

The results are presented in the tables 3-A through 3-D with different c_n values, viz. 5.12, 28.55, 92.98 and 421.34, implying negligible, critical (threshold), substantially high and very high degrees of multicollinearity respectively. We have not computed any overall measure (such as MSEL – Judge et al., 1982, p. 558) of performance of an estimator. Instead, we have computed Root Mean Square (RMS) of deviations of each individual $\hat{\beta}_j$ (given by RMS$_j$ $=[\{\sum_{i=1}^{ntrial}(\hat{\beta}_{ij}-\beta_j)^2\}/ntrial]^{1/2}$; $\hat{\beta}_{ij}$ is β_j estimated in the ith trial; ntrial=50; β_j=known parameter). It is obvious that if the RMS_a of most of the coefficients are smaller than the RMS_b of their counterpart, and no RMS_a is much larger than RMS_b , then $MSEL_a$ will be smaller than $MSEL_b$. Moreover, multicollinearity is not of so much concern to the overall model fit as to the individual coefficients of the model.

Table 3-D. Relative Performance of OLS and MEL Estimators

[Sample size (n) = 20; Condition Number (c_n)=421.34; $\mu = 419.43$]

$\sigma(u)$	Estimate	Estimator	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
$\sigma(u)=0$	$\hat{\beta}$	MEL	-2.806	0.641	-3.836	3.657	6.937
		MMEL	-0.001	-0.078	9.596	36.202	32.255
		OLS	10.000	20.000	30.000	40.000	50.000
	RMS	MEL	12.806	19.359	33.836	36.343	43.063
		MMEL	10.001	20.080	20.406	3.808	17.747
		OLS	0.000	0.000	0.000	0.000	0.000
$\sigma(u)=5$	$\hat{\beta}$	MEL	-2.817	0.651	-3.901	3.612	7.004
		MMEL	-2.789	-0.289	8.160	36.225	30.408
		OLS	14.384	-12.610	10.679	43.999	42.114
	RMS	MEL	12.822	19.353	33.909	36.393	43.001
		MMEL	13.892	20.314	22.070	6.295	20.128
		OLS	27.836	186.564	110.175	25.228	48.386
$\sigma(u)=10$	$\hat{\beta}$	MEL	-2.821	0.656	-3.953	3.562	7.055
		MMEL	-6.539	-0.195	6.693	36.513	28.162
		OLS	18.768	-45.221	-8.643	47.997	34.227
	RMS	MEL	12.842	19.357	33.985	36.458	42.965
		MMEL	20.308	20.216	24.152	10.424	23.758
		OLS	55.672	373.128	220.350	50.456	96.772
$\sigma(u)=15$	$\hat{\beta}$	MEL	-2.819	0.659	-3.994	3.504	7.091
		MMEL	-7.701	-0.955	5.925	27.167	27.579
		OLS	23.152	-77.831	-27.964	51.996	26.341
	RMS	MEL	12.867	19.370	34.064	36.539	42.954
		MMEL	24.900	21.216	25.576	14.776	25.967
		OLS	83.509	559.691	330.526	75.684	145.159
$\sigma(u)=20$	$\hat{\beta}$	MEL	-2.809	0.661	-4.026	3.435	7.111
		MMEL	-7.818	-1.573	5.412	37.261	27.443
		OLS	27.536	-110.441	-47.286	55.994	18.455
	RMS	MEL	12.895	19.389	34.149	36.640	42.970
		MMEL	29.723	21.944	26.998	19.859	28.341
		OLS	111.345	746.255	440.701	100.912	193.545

Note: MEL = Estimator à la Paris (2001); MMEL obtains $p(\beta)$ differently using absolute norm.

9. A Revisit to *probability* in Paris' MEL Estimator: Paris has defined $prob(\beta_i) = \beta_i^2 / \beta'\beta = \{\beta_i /(\beta'\beta)^{1/2}\}^2$. Thus, he has normalized β_i using the ***Euclidean norm*** of β. In the estimation procedure, $prob(\hat{\beta})$ is obtained accordingly. He draws a justification for this operation from physics. However, we must note that the physical system relating to light may not be archetypal to all systems (e.g. the economic system, see Georgescu-Roegen) that throw up data with the multicollinearity problem.

Therefore, let us part with the Euclidean norm, normalize β_i using the ***absolute norm*** of β and investigate into its effects on the performance of the Maximum Entropy estimator. We obtain a new estimator of β by solving the problem stated as:

$$\min H(p_\beta, L_\beta, u) = p'_\beta \log(p_\beta) + L_\beta \log(L_\beta) + u'u$$

subject to:

$$(i)\ y = X\beta + u$$

$$(ii)\ L_\beta = \sum_{j=1}^{m} \left|\beta_j\right|$$

$$(iii)\ p_{\beta_j} = \left|\beta_j\right| / L_\beta :\ \ 0 \le p_{\beta_j} \le 1;\ \ p_{\beta_j} \log(p_{\beta_j}) = 0\ \ if\ \ p_{\beta_j} = 0.$$

This new estimator is not fully à la Paris (2001) and hence we would call it the ***Modular Maximum Entropy Leuven*** (MMEL) estimator. The results of this enterprise are presented in the tables 3-A through 3-D, between the results of MEL and OLS estimators. We observe that overall, the performance of MMEL (in terms of mean estimated coefficients as well as the RMS values in 50 trials) is much superior to that of the MEL estimator.

It is pertinent to note here that obtaining $prob(\beta)$ is the most crucial task before the scientist if he chooses to use the maximum entropy estimator (MEL, MMEL or any variant thereof). After all, the mathematics of probability suggests us that given a sample description space S, probability is a function which assigns a non-negative real number to every event A, denoted by P(A) and it is called the probability of the event A. The probability function is defined on a Borel field of events conformal to the axioms of positiveness, certainty and union. Under these axioms, there could be several different rules of assignment, ranging from subjective judgement backed up by a rational belief to counting the number of success in the repeated trials. In our study MEL does this assignment in the one way and the MMEL does that in the other way. There could be many more (possibly better) rules of assignment. Thus, the subjective (or exogenous) element that was explicit in Golan et al. reappears in the MEL, although in another garb.

10. A Multi-Objective Optimization Interpretation of the MEL Estimator: The MEL estimator (MEL proper as well as MMEL) purports to minimize a combination of $\{u'u\}$ and $\{p'_\beta \log(p_\beta) + L_\beta \log(L_\beta)\}$. Consider the 2-objective minimization problem given as:

$$Min\ \ H_1(\beta \mid y, X) = u'u$$

$$Min\ \ H_2(\beta \mid y, X) = p'_\beta \log(p_\beta) + L_\beta \log(L_\beta)$$

subject to:

$$(i)\ \ y = X\beta + u;\ \ (ii) L_\beta = \beta'\beta;\ \ (iii) p_\beta = \beta \Theta \beta / L_\beta;\ \ (iv)\ \ 0 \le p_{\beta_j} \le 1.$$

$$(iv)\ \ p_{\beta_j} \log(p_{\beta_j}) = 0\ \ if\ \ p_{\beta_j} = 0.$$

Now, one of the simplest procedures to solve a multi-objective programming problem is to construct an auxiliary (composite) single objective function from the multiple objective functions and optimize it under the given constraints. One may write the resulting composite minimand objective function as a convex combination of the original H_1 and H_2 functions, such as $\kappa[u'u]+(1-\kappa)[p'_\beta \log(p_\beta)+L_\beta \log(L_\beta)]: 0<\kappa<1$. If $\kappa=1$, it gives us the conventional OLS estimator and if $\kappa=0$ it maximizes '*entropy*' leading to equalization of the regression coefficients, ignoring $u'u$ altogether. The MEL estimator chooses $\kappa=0.5$.

Nevertheless, one may choose to make κ a decision variable obeying some additional constraint(s). This approach to resolving the multicollinearity problem requires investigation. Further, there could be several alternative approaches to solve a multi-objective programming problem, other than the one outlined above. Note that all these observations apply to MMEL estimator also.

References

1. **Belsley, DA, E Kuh & RE Welsch** (1980). *Regression Diagnostics*, *Identifying Influential Data and Sources of Collinearity*, Wiley, New York.
2. **Bertrand, PV** (1998)."Constructing Explained and Explanatory Variables with Strange Statistical Analysis Results", *The Statistician: J of the Royal Stat. Soc.*, Series D 47, pp. 377-383.
3. **Bertrand, PV & RL Holder** (1988)."A Quirk in Multiple Regression: The Whole Regression can be greater than the Sum of its Parts", *The Statistician*, 37, pp. 371-374.
4. **Fröberg, CE** (1965). *Introduction to Numerical Analysis*, Addison-Wesley, London.
5. **Georgescu-Roegen, N** (1971). *The Entropy Law and The Economic Process*, Harvard Univ. Press, Cambridge, Mass.
6. **Golan, A, G George & D Miller** (1996). *Maximum Entropy Econometrics*, Wiley, London.
7. **Hoerl, AE & RW Kennard** (1970). "Ridge Regression: Biased Estimation for Non-Orthogonal Problems", *Technometrics*, pp. 55-68.
8. **Hoerl, AE & RW Kennard** (1970). "Ridge Regression: Applications to Non-Orthogonal Problems", *Technometrics*, pp. 69-82.
9. **Judge, GG, RC Hill, WE Griffiths, H Lütkepohl & TC Lee** (1982). *Introduction to the Theory and Practice of Econometrics*, Wiley, New York.
10. **Kaçiranlar, S, S Sakallioglu, F Akdeniz, GPH Styan & HJ Werner** (1999)"A New Biased Estimator in Linear Regression and a Detailed Analysis of the Widely-Analysed Dataset on Portland Cement", *Sankhyā: The Indian J. of Stat.*, 61.B.3, pp. 443-459.
11. **Krishnamurthy, EV & SK Sen** (1976). *Computer-Based Numerical Algorithms*, Affiliated East-West Press, New Delhi.
12. **Kuester, JL & JH Mize** (1973). *Optimization Techniques with Fortran*, McGraw Hill, New York.

13. **Liu, K** (1993).”A New Class of Biased Estimate in Linear Regression”, *Communications in Statistics – Theory and Methods*, 22, pp. 393-402.
14. **Paris, Q** (2001). “Multicollinearity and Maximum Entropy Estimators”, *Economics Bulletin*, 3(11), pp. 1-9.
15. **Rao, SS** (1978). *Optimization : Theory and Applications*, Wiley Eastern, New Delhi.
16. **Sarkar, N** (1992).”A New Estimator Combining the Ridge Regression and the Restricted Least Squares Methods of Estimation”, *Communications in Statistics – Theory and Methods*, 21, pp. 1987-2000.
17. **Stein, C** (1956).”Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution”, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1, 197-206.
18. **Theil, H** (1971). *Principles of Econometrics*, Wiley, New York.
19. **Theobald, CM** (1974). “Generalizations of Mean Square Error Applied to Ridge Regression”, *J. of the Royal Stat. Soc., Series B*, 36, pp. 103-106.
20. **Turing, AM** (1948) “Rounding-off errors in Matrix Processes”, *Qrtly J. Mech. Appl. Math.,* 1, pp. 287-308.
21. **von Neumann, J & HH Goldstine** (1947). “Numerical Inverting of Matrices of High Order”, *Bull. Amer. Math. Soc*., 53, pp. 1021-1099.