Software-Illustrated Explanations of Econometrics Contributions by CR Rao for his 100-th Birthday

H. D. Vinod[1]

Abstract

Since the vast array of scientific contributions by Dr. C. R. Rao are difficult to summarize in a short paper, we focus on (1) concepts relevant for applied statisticians including econometricians, (2) using the R software to explain the concepts to be better understood by Economics students who may lack sufficient background in mathematical statistics, (3) topics that are covered in publicly and freely available R packages. The range of R packages influenced by Rao's research is impressive. Dr. Rao will be 100 years old in 2020, and it is high time the Economics profession recognizes and honors his foundational contributions to econometrics.

Introduction

A review of Professor C. R. Rao's contributions in Statistics is available in Kumar et al. (2010). It shows the important role played by Professor Rao in the founding of the Indian Econometric Society, which recently held a 54-th Annual meeting in Jammu, India, attended by over 400 econometricians from all parts of India and abroad. Rao helped late Prof. P. C. Mahalanobis, his colleague and mentor at the Indian Statistical Institute in Calcutta, in formulating economic policies of free India upon achieving independence in 1947. He encouraged sound statistical training for economists and encouraged spending resources for good economic data collection. The journal *Econometric Theory* paid tribute to Rao by printing a 70-page long interview with him in Bera (2003), which includes detailed lists of Rao's 458 publications and various honors received. The Government of India acknowledged Rao's

✉ H. D. Vinod
vinod@fordham.edu

1   Fordham University, Bronx, New York 10458, USA

leadership in helping to improve Indian economic data collection and policy—a worthy activity for a potential Nobelist in Economics.

Instead of revisiting Kumar et al. (2010) in detail, this paper provides a short technical review, providing hands-on explanations of Rao's ideas with the help of R software absent in the earlier reviews and retrospectives. We hope to encourage young Econometricians to find inspiration and new insights from Rao's publications in their own research. Rao's ideas are explained anew with references to over 20 R packages, which should allow students to get a quick start with practical applications, even though some packaged examples might be from other sciences.

The input code snippets in R software and outputs produced by R are distinctly highlighted in the sequel, such that the reader can copy and paste the input and compare her locally produced output with our output reported here. Since the regression model is the bread and butter of applied econometrics, we use regressions to illustrate some ideas, even though some ideas can be more simply explained using the sample mean.

The usual linear regression model in matrix notation is

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \varepsilon = X\beta + \varepsilon, \tag{1}$$

where we have T observations, X is the $T \times (p + 1)$ matrix of data on all regressor variables including the first column of ones to represent the intercept β_0, y is a $T \times 1$ vector of data on the dependent variable, β is a $(p + 1) \times 1$ vector of regression coefficients, and ε is a $T \times 1$ vector of unknown true errors.

For example, we let y be the stopping distance ('dist') of a car (in feet), and x_1 be the speed of the car in miles per hour ('speed'), using Ezekiel's data called 'cars' always available in R. Our first R input code is:

```
#??datasets #this command lists all data sets available in R
cbind(head(cars,3),tail(cars,3)) # y=dist, x1=speed
nrow(cars) #how many observations=T?
```

The following output has the first set of three row-values in the first two columns and the last set of three rows-values in the last two columns. R does not print to the screen the tailing (last) three row numbers 48–50.

```
> cbind(head(cars,3),tail(cars,3)) # y=dist, x1=speed
    speed dist speed dist
1     4    2    24   93
2     4   10    24  120
3     7    4    25   85
> nrow(cars) #how many observations=T?
[1] 50
```

Note that 'cars' data in R has $T = 50$ with y for a (50×1) vector of stopping distances (2, 10, 4, ...,93, 120, 85). We need to insert the first column of T ones in our (50×2) X matrix for the intercept. The second column of X has x_1 representing car speeds. An R function for linear models (lm) creates an R object containing all regression results. We name the R object 'reg' in the following code.

```
library(stargazer)# for nice regression output
attach(cars) #to access columns by name
reg=lm(dist~speed) #standard R method for regression
stargazer(reg)
```

The output of the above code is in a form suitable to produce the following (Latex) Table 1, where standard errors of regression coefficients are in parentheses under the coefficient values.

Now we discuss the theory of linear regression to set up our notation needed for explaining Rao's contributions, including Cramer-Rao lower bound on variance in the regression context. We have the following semi-parametric (when normality is not assumed) probabilistic structure for y and errors in matrix notation:

Table 1 Regression of stopping distance on speed by standard R method

	Dependent variable Dist
Speed	3.932***
	(0.416)
Constant	−17.579**
	(6.758)
Observations	50
R^2	0.651
Adjusted R^2	0.644
Residual Std. error	15.380 (df = 48)
F statistic	89.567*** (df = 1; 48)

*p <0.1; **p < 0.05; ***p < 0.01

$$y = X\beta + \varepsilon, \quad E(\varepsilon) = 0, \quad E(\varepsilon\varepsilon') = \sigma^2 I_T, \tag{2}$$

where I_T is a $T \times T$ identity matrix, suggesting homoscedasticity or constant variances and zero covariances among errors. Note that $E(y) = X\beta$ is also the conditional mean $E(y|X)$. If we further assume that (a) the X matrix of regressors has full column rank, (b) all columns of X are uncorrelated with errors, and (c) that y (and ε) are multivariate Normal, $y \sim N(\mu, \sigma^2 I)$, then the usual t-tests and F tests on coefficients and the overall model are available. In particular, let the t-th error ε_t have the following Normal density:

$$f(\varepsilon_t) = (2\pi)^{-1/2}\sigma^{-1} \exp\{-(\varepsilon_t'\varepsilon_t)/2\sigma^2\}, \tag{3}$$

where ε_t is the t-th element of the vector $(y - X\beta)$ and is known only when β is known. Hence, the density $f(\varepsilon_t)$ is a function of data on y, X at time t, given specified numerical values of β and σ^2. The density is usually written as: $f(y_t, X_t|\beta, \sigma^2)$, where we assume that the numerical values of all items after the vertical bar are known. The joint density for all T observations is the product of such densities for all t from $t = 1$ to $t = T$:

$$f_{\text{joint}}(y, X|\beta, \sigma^2) = \prod_{t=1}^{T} f(\varepsilon_t), \tag{4}$$

where $f(\varepsilon_t)$ is defined in (3) and does depend on β through $\varepsilon = (y - X\beta)$.

Sir Fisher re-interpreted the same joint density function $f_{\text{joint}}(\cdot)$ as $f_{lkhd}(\beta, \sigma^2|y, X)$. Its log is called the log-likelihood (LL) function. It is useful when the object is to find the unknown parameters from the given data. For the regression model (2), we are not assuming the form of the density of ε, only that we know its mean and variance-covariance matrix. A quasi-log-likelihood function can be defined from the assumptions of (1), by pretending (hence quasi) that errors have the same mean and variance as the corresponding Normal without having to be actually Normal.

Now the quasi-LL and the usual LL by expanding the product in (4) after using (3) are given by

$$LL = -(T/2)\log(2\pi\sigma^2) + \varepsilon'\varepsilon/(2\sigma^2), \tag{5}$$

where

$$\varepsilon'\varepsilon = (y - X\beta)'(y - X\beta) = y'y - 2\beta'X'y + \beta'X'X\beta. \tag{6}$$

Now we assume that σ^2 is known for simplifying the following derivations and minimize $\varepsilon'\varepsilon$ with respect to β. An extension to the case where σ^2 is also unknown is obtained by maximizing the LL with respect to σ^2 is straightforward and available in textbooks. The matrix derivative of $2\beta'X'y$ with respect to β is $2X'y$, and the derivative of the quadratic form $\beta'X'X\beta$ is $2X'X\beta$. Setting the derivative equal to zero, we get the first-order condition (FOC) from calculus given by $-2X'y + 2X'X\beta = 0$. Upon canceling the 2 and solving for β, we have the ordinary least squares (OLS) estimator b as the solution:

$$b = (X'X)^{-1}X'y. \tag{7}$$

Let us use the 'cars' data to illustrate a numerical implementation of the above formula.

```
attach(cars) #access names speed, dist
beta0=rep(1,50)#vector of ones
X=cbind(beta0,speed)#create X matrix
XTX=t(X)%*%X #get X'X
XTy=t(X)%*%dist #get X'y
XTXinv=solve(XTX) #inverse of X'X
b=solve(XTX)%*%XTy #get OLS/ML estimator b
```

The output below of the above code shows that the OLS estimator b from first principles using the formula of Eq. (7) agrees with the R output in Table 1.

```
> b# OLS/ML estimator from first principles
[,1]
beta0 -17.579095
speed   3.932409
> reg=lm(dist~speed) #b by standard R methods
> coef(reg)
(Intercept)        speed
-17.579095     3.932409
```

Score Function and Score Equation

We have verified that R software correctly estimates the OLS estimator derived by matrix algebra formulas. Next we verify that the OLS formula also maximizes the log-likelihood (LL). First, we define the *score vector* g^* as the derivative of the LL function:

$$g^* = (\partial LL/\partial \beta) = 2X'(y - X\beta)/(2\sigma^2), \tag{8}$$

evaluated at the true parameter values. Since the first-order condition (FOC) for maximization of LL states that $E(g^*) = 0$, also known as the score equation. We can cancel the 2 from the numerator and denominator of (8) and write the FOC as:

$$X'(y - X\beta)/\sigma^2 = 0 = (1/\sigma^2)X'\varepsilon. \tag{9}$$

Now the LL maximizing solution upon setting the left-hand-side equal to zero is the same b defied above in eq. (7):

Fisher Information Matrix

Fisher information matrix is defined as the variance-covariance matrix of the score vector (from the outer product of the score vector) as:

$$I_{nf} = E(g^* g^{*\prime}) = (1/\sigma^4) X' E(\varepsilon \varepsilon') X = (1/\sigma^2) X' X, \tag{10}$$

where we have used $E(\varepsilon \varepsilon') = \sigma^2 I_T$ from Eq. (2).

The second-order condition for a maximum from calculus is that the matrix of second-order partial derivatives should be negative definite. It can be stated in terms of the Fisher information matrix based on the second-order partials as:

$$I_{inf} = -E\left[\frac{\partial^2}{\partial \beta'} LL\right] \tag{11}$$

Since the Fisher information matrix is a projection matrix, it is non-negative definite. Thus the ML estimator of the β in Eq. (7) satisfies both first and second-order conditions for a maximum. We use the usual estimate s^2 of σ^2 from the residual sum of squares (RSS) divided by the degrees of freedom (df $= T - p - 1$).

```
df=length(speed)-2;df
s2=sum(resid(reg)^2)/df;s2
```

In the above code, s2 equals RSS/df.

```
> df=length(speed)-2;df
[1] 48
> s2=sum(resid(reg)^2)/df;s2
[1] 236.5317
```

Note that this output agrees with the square of 'Residual Std. Error' of 15,380, reported in Table 1. A slight discrepancy is due to the superior numerical accuracy of R.

```
XTX=t(X)%*%X #get X'X
s2*XTX
```

The information matrix for our cars example is $s^2 X' X$ with the R output given next.

```
> XTX=t(X)%*%X #get X'X
> s2*XTX
beta0      speed
beta0 11826.58  182129.4
speed 182129.40 3128841.2
```

Having set up the basic notation, we are ready to discuss one foundational contribution by Rao.

The Unified Theory of Linear Estimation

The familiar notation A^- to denote generalized inverse was first developed by Rao in 1955 based on an application to a study of long-term effects of radiation on Nagasaki-Hiroshima victims, (Rao 1991, p. 33), independently of Penrose's 1955 work leading to Moore-Penrose inverse. The properties of generalized inverse were fully worked out by Rao. For example, if A^- is a left inverse of A, then pre-multiplication should yield the identity matrix: $A^-A = I$

If we choose the left inverse $A^- = (A'A)^{-1}A'$ as our generalized inverse of A, then the generalized inverse of A is very easy to compute by the following one-line code in R:

```
gen.inv=solve(t(A)%*%A) %*% A
```

Rao's unified theory applies the generalized inverse to report some almost unbelievable properties described in the sequel. What is the motivation? Equation (2) assumes that X is of full column rank, implying that $X'X$ can be inverted and also that $V = \varepsilon\varepsilon'$ is non-singular. What if these matrices are singular? Rao has provided an elegant solution to both problems and introduced an "inverse partitioned matrix," which yields summary statistics for linear models.

$$\begin{bmatrix} V & X \\ X' & 0 \end{bmatrix}^- = \begin{bmatrix} C_1 & C_2 \\ C_3 & -C_4 \end{bmatrix}, \tag{12}$$

where superscript $(-)$ denotes any generalized inverse, and the matrices C_1 to C_4 are numerically known from the elements in a generalized inverse.

Now we have several relations providing all estimators and statistics needed for inference.

$$\hat{\beta} = C_2'y = C_3y, \tag{13}$$

$$df = rank(V : X) - rank(X), \tag{14}$$

$$\hat{\sigma}^2 = y'C_1y/df, \tag{15}$$

$$Var(P'\hat{\beta}) = \sigma^2 P'C_4P, \tag{16}$$

where P matrix defines a linear combination, df denotes degrees of freedom, where $rank(V : X)$ denotes the rank of a matrix obtained by writing the two matrices side-by-side, and where Var denotes variance-covariance matrix.

Now the R ode to verify these results for cars data is as follows.

```
attach(cars) #access names speed, dist
beta0=rep(1,50)#vector of ones
X=cbind(beta0,speed)#create X matrix
y=dist
T=nrow(cars)
p=2
V=diag(T)
zer=matrix(0,nrow=p,ncol=p)
A0=cbind(V,X)
A1=cbind(t(X),zer)
A=rbind(A0,A1)
dim(A)
Ainv=solve(t(A)%*%A) %*% A
C1=Ainv[1:T,1:T]
C2=Ainv[1:T,(T+1):(T+p)]
C3=Ainv[(T+1):(T+p),1:T]
C4=-Ainv[(T+1):(T+p),(T+1):(T+p)]
```

Now we can verify Eq. (13) by the following code:

```
bhat=C3 %*% y;bhat
t(C2)%*%y
```

The output agrees with Table 1.

```
> bhat=C3 %*% y;bhat
beta0 -17.579095
speed   3.932409
> t(C2)%*%y
beta0 -17.579095
speed   3.932409
```

Next, we can verify Eq. (14) for degrees of freedom.

```
library(fBasics)
rk(cbind(V,X))-rk(X)
```

The output agrees with Table 1.

```
> rk(cbind(V,X))-rk(X)
[1] 48
```

Next, we verify Eq. (15) for error variance.

```
s2=t(y)%*%C1 %*%y /df; s2
```

The output s2 below agrees with its square root reported in Table 1.

```
> s2=t(y)%*%C1 %*%y /df; s2
[,1]
[1,] 236.5317
```

Note that the matrix on the left side of Eq. (12) can be huge. If there are 500 elements in the rows of y, this is a 502×502 matrix. Numerical mathematics using computers suggests that one should avoid computing inverses of large matrices, since numerical rounding and truncation errors can be difficult to control, Vinod and McCullough (2003).

Hence if V^{-1} and $(X'X)^{-1}$ exist, Eq. (12) is mainly of theoretical interest for its ability to yield all summary statistics of interest in one matrix operation. It is particularly useful when the inverses of V and $(X'X)$ do not exist.

If the left inverse is difficult to compute, one will need to use the singular value decomposition by the 'svd' function in R. We define an R object sv1 to store the decomposition of A into three matrices UDV, where D is diagonal. Note that $A = UDV'$ and $A^- = UD^{-1}V'$. Once the generalized inverse is computed, it is quite simple to compute the matrices C_1 to C_4 by partitioning the generalized inverse matrix.

```
sv1=svd(A)
U=sv1$u
d=sv1$d
V=sv1$v
Ainv=U%*%diag(1/d)%*%t(V)
C1=Ainv[1:T,1:T]
C2=Ainv[1:T,(T+1):(T+p)]
C3=Ainv[(T+1):(T+p),1:T]
C4=-Ainv[(T+1):(T+p),(T+1):(T+p)]
bhat=C3 %*% y;bhat
t(C2)%*%y
```

If the output of the above code does not give two nearly identical estimates of $\hat{\beta}$ (from C3 and C2) we have serious numerical (collinearity) problems, rendering the results from inverse partitioned matrix of Eq. (12) unreliable.

Cramer-Rao Lower Bound (CRLB)

Let a log-likelihood function similar to Eq. (5) be twice differentiable, and the support of y does not depend on unknown parameters. If b is an unbiased estimator of β, then using (10) we have:

$$Var(b) \geq I_{inf}^{-1} = \sigma^2 (X'X)^{-1}, \tag{17}$$

which suggests that the variance-covariance matrix of the OLS/ML estimator, $\sigma^2(X'X)^{-1}$, reaches the Cramer-Rao lower bound. Hence b is Cramer-Rao efficient. Rao (1991) states that he established the inequality in 1943, when Rao was only 23, even though its publication was delayed due to the Second World War until 1945.

Now we illustrate the Cramer-Rao lower bound of Eq. (17) using 'cars' data. We use the usual estimate of σ^2 from the residual sum of squares (RSS) divided by the degrees of freedom (df $= T - p - 1$).

```
df=length(speed)-2;df
s2=sum(resid(reg)^2)/df;s2
CRLB=s2*XTXinv;CRLB
```

In the above code, s2 equals (RSS/df), and CRLB denotes the Cramer-Rao lower bound matrix. The usual standard errors of regression coefficients are simply square roots of the diagonal elements of CRLB, which agree with the R output in Table 1.

```
> df=length(speed)-2;df
[1] 48
> s2=sum(resid(reg)^2)/df;s2
[1] 236.5317
> CRLB=s2*XTXinv;CRLB #matrix
          beta0        speed
beta0 45.676514 -2.6588234
speed -2.658823  0.1726509
> sqrt(diag(CRLB))
beta0      speed
6.7584402 0.4155128
```

It can be shown that the usual unbiased estimator of $\sigma^2 = (\hat{\varepsilon}, \varepsilon)/(T - p - 1)$ does not reach the CRLB, since its variance $2\sigma^4/T$.

Applications and Extensions of CRLB

In addition to the median, economists are often interested in other quantiles. For example, the top 5% or bottom 5% of incomes help understand the richest and the poorest persons. Estimation of such quantiles is optimal in the sense of reaching the Cramer-Rao lower bound is established in Godambe (2001) by showing that the quantile is also the score function under appropriate distributional assumptions. Kale (1962) expounds the role of CRLB to "estimating functions," which has spawned a vast literature.

Economists are often interested in certain functions of regression coefficients $f(\beta)$ besides the individual coefficients β. A practical problem arises regarding the efficiency of the plug-in estimator $f(\hat{\beta})$ obtained by replacing the unknown by its estimate, where CRLB can help.

Econometric practice often involves assuming normality or linearity in their parametric models. However, one fears misspecification in making these assumptions. Anytime functional forms of some components of the model are unknown, and we have so-called semi-parametric models. Newey (1990) reviews the statisticians' approach to these problems by using the CRLB with explicit econometric applications in mind. One considers parametric submodels of the semiparametric model and use the CRLB for submodels and their supremum. He proves that "The asymptotic variance of any semiparametric estimator is no smaller than the supremum of the Cramer-Rao bounds for all parametric submodels."

If one admits autocorrelation and heteroscedasticity among regression errors, we must rewrite Eq. (2) to replace the identity matrix by V. That is, we assume a non-spherical variance-covariance matrix of regression errors: $E(\varepsilon \varepsilon') = \sigma^2 V$. A feasible generalized least squares (FGLS) estimator extends the OLS estimator of (7) by inserting an estimate $\hat{V}$ of the matrix V as follows:

$$b_{gls} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}y. \qquad (18)$$

Newey argues that the semiparametric efficiency bound for FGLS is the asymptotic variance of the GLS estimator, which does reach CRLB.

The Generalized method of moments (GMM) is a popular econometric tool. Its main selling point is that it reaches the CRLB. GMM is closely related to 'optimal estimating functions' (OptEF) similar to g^* of Eq. (8). Vinod (2008) (Sec. 10.3) shows that the OptEF are optimal in the sense that they reach the CRLB.

Rao showed in a Sankhya paper in 1961 a notion of the first-order efficiency of a consistent estimator readily verified by the unit correlation between it and the score. This avoids linking efficiency to the variance and also avoids an anomaly created by Hodge's example of the nonexistence of lower bound on variance in some cases, (Rao 1991, p. 30). Hall and Mathiason (1990), almost never cited in Econometrics literature, shows with detailed proofs that CRLB provides a basis for defining efficiency and that ML estimators and certain 'one-step' variations achieve such efficiency. These results extend to models with nuisance parameters commonly present in Econometrics. Three useful results based on CRLB are:

1) Roots of score equations, $g^* = 0$, in the neighborhood of any preliminary root-n-consistent estimates provide efficient estimates.
2) Replacing nuisance parameters with consistent estimates continues to retain asymptotic efficiency.
3) A test statistic based on a quadratic form (e.g. $g^{*\prime}I_{inf}^{-1}g^*$) is asymptotically efficient and is distributed as a Chi-square random variable.

Rao's Tests in R Packages

We list some R packages inspired by Rao's contributions with potential applications in Econometrics.

1) 'aSPU' Package by Il-Youp Kwak provides R code for 'adaptive sum of powered score test.' It has applications in genetics focusing on pathways for genetic traits. See details at https://github.com/ikwak2/aSPU. These may be useful for the economic traits of sub-populations in econometric panel (longitudinal) data studies.
2) 'weightedScores' Package by Aristidis K. Nikoloulopoulos has R code for weighted scores for regression models with dependent data. It reports the results of an intermediate step for variable selection for longitudinal categorical and count data.
3) 'mdscore' Package by Antonio Hermes M. da Silva-Junior has a set of R functions to obtain a modified score test for generalized linear models.
4) 'IPWsurvival' Package by Y. Foucher has propensity score based adjusted survival curves and their log-rank statistic, where the adjustment acknowledges the presence of confounding factors. The log-rank test based on inverse probability weighting (IPW) is implemented. www.r-project.org, www.labcom-risca.com.

5) 'mr.raps' package by Qingyuan Zhao implements methods for two-sample Mendelian randomization with summary statistics by using Robust Adjusted Profile Score (RAPS).

6) 'VGAM' Package by Thomas Yee has Vector Generalized Linear and Additive Models (VGAM) and reduced rank regressions inspired by Rao and used in Econometrics. See https://www.stat.auckland.ac.nz/~yee/VGAM.

7) 'SCGLR' Package by Guillaume Cornu extends the Fisher-Rao scoring algorithm to combine Partial Least Squares (PLS) regression with a Generalized Linear Model (GLM). It allows joint modeling of variables from different exponential family distributions, searching for common PLS-type components.

8) 'robustrao' Package by Maria del Carmen Calatrava Moreno provides R code to compute the Rao-Stirling diversity index (Porter and Rafols, 2009) and its extension to acknowledge missing data.

9) 'RSarules' Package by Xiaoying Sun has random sampling association rules from a transaction dataset. It implements the Gibbs sampling algorithm by Qian et al. (2016) .

10) 'StatCharrms' Package by Joe Swintek has repeated measures and multi-generation studies. It has Rao-inspired weighted ANOVA, mixed-effects ANOVA, repeated measures ANOVA, and the Dunnett test. https://CRAN.R-project.org/package=StatCharrms

11) 'Select' Package by Daniel Laughlin determines species probabilities based on functional traits. It optimizes Rao's Q, a closed-form functional diversity index that incorporates species abundances, subject to other linear constraints. This framework optimizes both functional diversity and entropy simultaneously.

12) 'Iboot' Package by Nicola Lunardon has iterated bootstrap tests and confidence sets. The package implements a general algorithm to obtain iterated bootstrap tests and confidence sets for a p-dimensional parameter based on the un-studentized version of the Rao statistic.

13) 'minque' Package by Jixiang Wu is designed (i) to construct a user-defined linear mixed model, (ii) to employ minimum norm quadratic unbiased estimation (MINQUE) for variance component estimation and random effect prediction; and (iii) to employ jackknife resampling tests. The package has an application to Maize variety trial with two years and multi-locations in China.

14) 'cpca' Package by Andrey Ziyatdinov is for Principal Component Analysis (PCA), and Trendafilov's stepwise estimation of common principal components. See https://github.com/variani/cpca.

15) 'FactoMineR' package by Francois Husson implements using principal component analysis (PCA) when variables are quantitative, correspondence analysis (CA), and multiple correspondence analysis (MCA) when variables are categorical, Multiple Factor Analysis when variables are structured in groups. See http://factominer.free.fr.

16) 'svdvis' Package by Neo Christopher Chung provides Singular Value Decomposition (SVD) visualization useful for principal component analysis (PCA), factor analysis (FA) and related methods.

17) 'RMThreshold' Package by Uwe Menzel determines an objective threshold for signal-noise separation in random matrices by using eigenvalue spectrum analy-

sis. The algorithm unravels the modular structure of a matrix - or the corresponding network.

18) 'CEC' Package by Konrad Kamieniecki does Cross-Entropy Clustering (CEC) to divide the data into Gaussian type clusters. It allows the simultaneous use of various types of Gaussian mixture models. See https://github.com/azureblue/cec.

19) 'CEoptim' Package by Benoit Liquet is a Cross-Entropy R Package for Optimization.

20) 'afCEC' Package by Krzysztof Byrski uses Rao's cross-entropy to construct a function for cross-entropy clustering done by partitioning n-dimensional data into clusters. It is preceded by finding the parameters of the mixed generalized multivariate normal distribution, that optimally approximates the scattering of the data in the n-dimensional space. See https://github.com/GeigenPrinzipal/afCEC

21) 'pcaL1' Package by Paul Brooks is designed for L1-Norm principal components analysis methods pioneered by Rao in the 1950s. See http://www.optimization-online.org/DB_HTML/2012/04/3436.html and http://www.coin-or.org.

There are likely many more R packages which use, if not explicitly cite, Rao's work.

LM test in Econometrics is Rao's Score Test

One is often interested in testing a null hypothesis involving k restrictions as functions of parameters:

$$H_0 : f_j(\beta) = 0, (j = 1, \dots, k). \tag{19}$$

Aitchison and Silvey (1960) consider this test by setting up a Lagrangian involving the log-likelihood, $LL(\beta)$ as:

$$\tilde{L} = LL(\beta) + \Sigma_j^k \lambda_j f_j(\beta), \tag{20}$$

where the Lagrangian coefficients are λ_j and β has (p+1) parameters. Setting the derivative of $\tilde{L}$ with respect to β equal to zero. Let $\tilde{\lambda}$ denote a $k \times 1$ vector of Lagrange multipliers upon solving the first-order condition (FOC), and $\tilde{F}$ denote a $(p + 1) \times k$ matrix of partial derivatives $\partial f_j / \partial \beta$ evaluated a the solution of the FOC.

Now Lagrange Multiplier (LM) test statistic:

$$LM = \tilde{\lambda}' \tilde{F}' I_{inf}^{-1} \tilde{\lambda} \tilde{F} = g^{*\prime} I_{inf}^{-1} g^*, \tag{21}$$

which is a quadratic form involving the information matrix of Eq. (10) and which uses Rao's score vector stated above in Eq. (8) was published in 1948.

Note that the second equality in Eq. (21) above requires some simplification and is proved by Breusch and Pagan (1980). They remark that the latter form of the LM test statistic is Rao's efficient score test statistic and recommend using Rao's version, even though they continue to call it the LM test. Recall that both Eqs. (10), and (11) yield Fisher's Information matrix, and the researcher's convenience can dictate which one to use in a particular application. There is considerable literature

involving comparison of LM, likelihood ratio (LR) and Wald tests, sometimes called the 'holy trinity' of tests. It is shown that the three are asymptotically equivalent, that the LR test alone is unbiased and that simulations favor Rao's formulation of the LM test, which is devoid of explicit Lagrange multipliers. Rao (1991) reports papers where Rao's LM test works even better when the bias is removed and goes on to mention econometric references explicitly citing Rao.

Breusch and Pagan (1980) describe important applications of Rao's score test (LM) for a model involving a test for the liquidity trap in monetary policy. Durbin's h test for autocorrelated residuals when a lagged dependent variable is present, and Box-Pierce test for autocorrelations are shown to be "essentially LM" tests. They also recommend the score test for error components, non-spherical error covariances, Zellner's seemingly unrelated regression systems, and non-nested hypotheses testing.

When ML estimators are too difficult to calculate due to the presence of I_{inf} in the expression for the LM statistic based on Eq. (21), Section 4 of Breusch and Pagan (1980) suggests a pseudo-LM procedure using the following four steps: (1) first compute consistent estimates of the parameters under the null to get regression residuals $\bar{e}$. (2) Regress $\bar{e}$ on all scores g^* evaluated at the null so that the coefficient of determination is $R^* = g'^* I_{inf}^{-1} g^*$. (3) Regress $\bar{e}$ on scores g_2^* evaluated at the null on a subset of k coefficients (or coefficient functions) involved in the test, so that the coefficient of determination, $R^{**} = g_2'^* I_{inf}^{-1} g_2^*$. (4) Now $N(R^* - R^{**})$ is a $\chi^2(k)$ variable, allowing a test of the null. Many econometricians use this trick without being aware that it emanates from Rao's score test.

Breusch and Pagan (1980) also show that small sample distribution of the LM statistic can often be determined by numerical means. Thus Rao's statistic remains appealing and relevant in a modern computing environment.

Rao-Blackwellization and Sufficient Statistics

Econometric models are often used for policy decisions by politicians and other non-professionals, where it is convenient not to use the entire sample, but discuss only a relevant sample statistic. A formal justification for this common practice requires us to use only so-called "sufficient statistics," defined by Fisher as requiring that no other statistic from the same sample provides any additional information. Rao established the following result in 1945, showing a reduction in mean squared error (MSE).

Rao-Blackwell Result If t is an unbiased estimator of a parameter θ and if T is a sufficient statistic, then $t_* = E(t|T)$ is a function of T, unbiased for θ and $MSE(t_*) \leq MSE(t)$.

Its use in improving the asymptotic efficiency of estimators is referred to as Rao-Blackwellization in the literature since Blackwell had independently published a similar result a bit later in 1947. It can be shown that sample mean and variance are jointly sufficient statistics for corresponding population parameters. The above result implies that if a crude estimate of a function $f(\beta)$ is available, its conditional

expectation given a sufficient statistic gives an estimator which is optimal in the sense that its variance is 'small.'

Although I am not aware of any econometric application of Rao-Blackwell-ization, it remains worth exploring. For example, let us consider the maximum entropy bootstrap from Vinod and López-de-Lacalle (2009) implemented in the R package 'meboot.' It attempts to approximate a population ensemble Ω of time series such that the observed time series x_t is a realization from Ω. The construction of Ω is useful for the bootstrap inference for ubiquitous nonstationary time series in Econometrics.

One begins with the empirical cumulative distribution function (ecdf) of x_t, where ecdf is known to be a sufficient statistic and uses random numbers from a uniform density to construct a large number of incarnations of x_t denoted as Ω_i as i-th approximation to the true Ω. Note that 'meboot' approximately preserves time dependence properties of x_t such as autocorrelation coefficients of various orders and general shape of the spectral density of x_t without imposing any parametric constraints. Hence meboot permits computer-intensive bootstrap inference, without injecting misspecifications of the original economic model by converting commonly occurring nonstationary economic time series (e.g., GDP, Consumption) into stationary series by differencing or de-trending, just for statistical inference.

Each Ω_i satisfies the 'ergodic theorem' since ensemble average of all time series in it equals the time average $\bar{x}$. Hence Ω_i is an unbiased estimate of Ω. Since it is based on a sufficient statistic ecdf, the Rao-Blackwell result implies that conditional mean $E(\Omega_i|ecdf)$ denoted by $\bar{\Omega}$ satisfies $MSE(\bar{\Omega}) \leq MSE(\Omega_i)$. Fenga (2020) has recently used 'meboot' for forecasting covid-19 diffusion in Italy, relying on the ergodic theorem. If an econometric estimator is not already a function of a sufficient statistic, Rao has suggested a way to improve it.

Rao's U-Test for Additional Information

It is a common problem in empirical work having access to $(p + q)$ data series, whether the use of any additional q series is worthwhile. The usual variance ratio or F test used for this purpose in econometrics is a variant of Rao's U test from Rao (1948b).

Hotelling's canonical correlations have been applied in Econometrics for estimation of joint production functions with outputs wool and mutton and the usual capital and labor inputs, Vinod (1968) and Vinod (1976). Hotelling maximizes the correlation between a linear combination of outputs and inputs. However, Rao seems to maximize the 'within input' and 'within output' correlations plus 'between input' and 'between output' correlations.

Rao's Miscellaneous Contributions Relevant to Econometrics

Professor Rao's contributions relevant to Econometrics are too many to give an exhaustive survey. This section lists some of them not yet covered.

1) In the new world of Big Data, Rao's tools for dimension reduction based on separation theorems described in his articles in the *Journal of Multivariate Analysis* should become more and more relevant.

2) Multivariate analysis of variance (MANOVA) was first conceived by Rao as a multivariate analysis of dispersion.

3) Edgeworth expansions for a vector of variables in non-regular cases described in Bai and Rao (1991) have potential applicability in the bootstrap literature.

4) Kotz and Johnson have edited a book entitled "Breakthroughs in Statistics: 1889–1990," which includes a reprint of Rao (1948a). Calling it a breakthrough is particularly appropriate for economists, since it had considerable influence on asymptotic econometrics.

5) Rao's new probability distributions called weighted distributions are applied in Econometrics in Vinod (1991) for studying unemployment statistics and Okun's law.

6) Maximum entropy concepts commonly used in Econometrics owe a great deal to the characterization of probability distributions developed by Rao and coauthors from Russia in Kagan et al. (1973).

7) Rao's quadratic entropy (Rao 1991, p. 58) was motivated by trying to understand diversity between two groups by considering the average distance between two randomly chosen points. Rao's cross-entropy generalizes Kulback-Liebler 'directed divergence.' Financial economists who want to diversify their portfolios can use these divergence measures.

Final Remark

There is enough information here to justify treating Rao as eminently eligible to receive the Nobel prize in Economics. Dr. Rao's contributions have influenced a huge range of R packages. A sample of 21 such packages listed in this paper has been downloaded thousands of times every month. Dr. Rao obviously continues to be relevant for practicing researchers in all quantitative fields, including economics. Perhaps, awarding Dr. Rao the economics Nobel will further enhance the prestige of the award.

References

Aitchison, J., and S.D. Silvey. 1960. Maximum-likelihood Estimation of Parameters Subject to Restraints. *Annals of Mathematical Statistics* 29: 813–828.

Bai, Z.D., and C.R. Rao. 1991. Edgeworth Expansion of a Function of Sample Means. *Annals of Statistics* 19 (3): 1295–1315.

Bera, A. 2003. ET Interview: Professor C R Rao. *Econometric Theory* 19: 331–400.

Breusch, T.S., and A. Pagan. 1980. The Lagrange multiplier test and its application to model specification in econometrics. *Review of Economic Studies* 47: 239–253.

Fenga, L. 2020. Forecasting the CoViD-19 diffusion in Italy and the related occupancy of Intensive Care Units, Tech. rep., Italian National Institute of Statistics ISTAT, Rome, Italy 00184. https://www.medrxiv.org/content/10.1101/2020.03.30.20047894v1.full.pdf. Accessed 13 June 2020.

Godambe, V. P. 2001. Estimation of Median: Quasi-Likelihood and Optimum Estimating Functions, Working paper, Statistics Department, University of Waterloo. https://www.researchgate.net/publication/251344524_Estimation_of_Median_Quasi-Likelihood_and_Optimum_Estimating_Functions. Accessed 13 June 2020.

Hall, W.J., and D.J. Mathiason. 1990. On Large-Sample Estimation and Testing in Parametric Models. *International Statistical Review* 58 (1): 77–97.

Kagan, A.M., YuV Linnik, and C.R. Rao. 1973. *Characterization Problems in Mathematical Statistics*. New York: Wiley.

Kale, B.K. 1962. An Extension of the Cramer-Rao Inequality for Statistical Estimation Functions. *Skandinavisk Aktuarietidisk Rift* 45: 60–89.

Kumar, T. K., Vinod, H. D., and Deman, S. 2010. Professor C.R. Rao's Contributions to Econometrics, *SSRN eLibrary*. http://ssrn.com/abstract=1722743. Accessed 13 June 2020.

Newey, W.K. 1990. Semiparametric efficiency bounds. *Journal of Applied Econometrics* 5: 99–135.

Rao, C.R. 1948a. Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proceedings of the Cambridge Philosophical Society* 44: 50–57.

Rao, C.R. 1948b. Tests of Significance in Multivariate Analysis. *Biometrika* 35: 58–79.

Rao, C. R. 1991. *An Autobiographical Account of Research Work*, Montreal, Quebec, Canada, H3G 1M6: T. D. Dwivedi, Statistics '91 Canada, Concordia University.

Vinod, H.D. 1968. Econometrics of joint production. *Econometrica* 36: 322–336.

Vinod, H.D. 1976. Canonical ridge and econometrics of joint production. *Journal of Econometrics* 4: 147–166.

Vinod, H.D. 1991. Rao's Weighted Distribution in Econometrics: An Application to Unemployment Statistics and Okun's Law. *Journal of Quantitative Economics, New Series* 7 (2): 247–254.

Vinod, H.D. 2008. *Hands-on Intermediate Econometrics Using R: Templates for Extending Dozens of Practical Examples*. Hackensack, NJ: World Scientific. ISBN 10-981-281-885-5. https://www.worldscientific.com/worldscibooks/10.1142/6895. Accessed 13 June 2020.

Vinod, H.D., and J. López-de-Lacalle. 2009. Maximum Entropy Bootstrap for Time Series: The meboot R Package. *Journal of Statistical Software* 29: 1–19.

Vinod, H.D., and B.D. McCullough. 2003. Comments: Econometrics and software. *Journal of Economic Perspectives* 17 (1): 223–224.