Journal of Econometrics 8 (1978) 127–158. © North-Holland Publishing Company

ESTIMATION OF FUNCTIONS OF POPULATION MEANS AND REGRESSION COEFFICIENTS INCLUDING STRUCTURAL COEFFICIENTS

A Minimum Expected Loss (MELO) Approach*

Arnold ZELLNER

Graduate School of Business, University of Chicago, Chicago, IL60637, USA

Received November 1975, final version received May 1977

Parameter estimates that minimize the posterior expectation of generalized quadratic loss functions are derived for a wide range of estimation problems encountered in econometrics and statistics including estimation of structural coefficients of linear structural econometric models and reciprocals and ratios of population means and regression or reduced form coefficients. These MELO estimates are simple in form and possess finite sampling moments and risk in contrast to other estimators, including maximum likelihood estimators, that possess no finite moments and have infinite risk relative to quadratic loss functions. Additional finite and large sample properties of MELO estimates are derived and discussed.

1. Introduction

In this paper, problems of estimating (a) reciprocals and ratios of population means and regression coefficients and (b) structural coefficients of linear structural econometric models are considered. For each problem in (a) and (b) formulae defining parameter estimates that Minimize Posterior Expected LOss, termed MELO estimates, are derived for specific loss functions and general forms for prior distributions and likelihood functions. Below these general results are specialized to provide explicit MELO estimates of parameters in normal likelihood functions when diffuse prior distributions are employed. When these MELO estimates are viewed as estimators, it is found that they have at least a finite second moment and thus finite risk relative to quadratic and quite a few other loss functions. In contrast, maximum likelihood (ML) estimators for the problems in (a) when likelihood functions are based on normal data distributions do not possess finite moments and have infinite risk relative to quadratic and other loss functions.

*Research financed by NSF Grants GS-40033 and SOC 7305547 and by income from the H.G.B. Alexander Endowment Fund, Graduate School of Business, University of Chicago. Many of the results in this paper were reported to the 9th NBER–NSF Seminar on Bayesian Inference in Econometrics, held at the University of Montreal, October 1974, and to the Indian Econometric Conference, New Delhi, January 1975.

With respect to problems in (b), it is well-known that many widely used estimators such as 2SLS, LIML, 3SLS, FIML, etc., can fail to possess finite moments[1] and thus can have infinite risk relative to quadratic and other loss functions. Further, it is the case that for problems in (a) and (b), means of posterior distributions, based on diffuse or natural conjugate prior distributions, often fail to exist. Thus, in these cases, the usual Bayesian point estimate, the posterior mean, is not available.

The plan of the paper is as follows. In section 2, problems of estimating the reciprocal of a population mean, the reciprocal of a regression coefficient and ratios of population means and regression coefficients are analyzed. The analysis of section 3 involves deriving MELO estimators for structural coefficients of linear structural econometric models. In section 4 further consideration is given to the properties of the MELO estimates and estimators derived in sections 2 and 3. Last, in section 5 some concluding remarks are presented.

2. MELO estimation of reciprocals and ratios of population means and regression coefficients

2.1. *Reciprocals of population means and regression coefficients*

Let our model for the observations be

$$y_i = \mu + u_i, \qquad i = 1, 2, \ldots, n, \tag{2.1}$$

where y_i is the ith observation, μ is the common mean of the observations, and u_i is the ith disturbance or error term. Our problem is to estimate $\theta = 1/\mu$.

The loss function that we shall employ is the following relative squared error loss function:[2]

$$L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2/\theta^2, \tag{2.2}$$

where $\hat{\theta} = \hat{\theta}(y)$ is an estimate of θ. Note that for the relative squared error loss function, a *given* absolute error, $|\theta - \hat{\theta}|$, is considered to be more serious when the true value of θ is small than when it is large whereas with a squared error loss function, $(\theta - \hat{\theta})^2$, the same loss is experienced for a given absolute error whatever the magnitude of the true value of θ. Further, the

[1] See, e.g., Anderson–Sawa (1973), Bergstrom (1962), Hatanaka (1973), Sawa (1972).

[2] Relative squared error loss functions have been previously employed and discussed in the literature; see, e.g., De Groot (1970, p. 226) and Ferguson (1967, pp. 47, 26, 28). In Zellner–Vandaele (1975) it is shown that a Stein-like estimator for the mean of a multivariate normal pdf can be generated using a generalized relative squared error loss function and a diffuse, improper prior pdf.

loss function in (2.2) can be obtained as follows. From $\theta = 1/\mu$, we have $1-\theta\mu=0$. Letting $\varepsilon = 1-\hat{\theta}\mu$, $\varepsilon^2 = (1-\hat{\theta}\mu)^2 = (\theta-\hat{\theta})^2/\theta^2$. The quantity ε measures the extent to which the relation $1-\theta\mu = 0$ is in error when an estimate $\hat{\theta}$ is inserted for θ.

Given that we have a posterior pdf for μ, $p(\mu|y, PI)$ that possesses finite first and second moments, we can compute posterior expected loss by expressing (2.2) as $L = (\theta-\hat{\theta})^2/\theta^2 = 1-2\mu\hat{\theta}+\mu^2\hat{\theta}^2$ and using the posterior pdf for μ to obtain

$$EL = 1-2\hat{\theta}E\mu+\hat{\theta}^2E\mu^2, \tag{2.3}$$

where $E \equiv$ posterior expectation. The value of $\hat{\theta}$ that minimizes (2.3), the MELO estimate, denoted by $\hat{\theta}^*$ is

$$\hat{\theta}^* = \frac{E\mu}{E\mu^2} = \frac{\bar{\mu}}{\bar{\mu}^2+\operatorname{var}\mu} = \frac{1}{\bar{\mu}}\;\frac{1}{1+\operatorname{var}\mu/\bar{\mu}^2}, \tag{2.4}$$

where $\bar{\mu} \equiv E\mu$, the posterior mean, and $\operatorname{var}\mu \equiv$ the posterior variance of μ.[3] Thus it is seen that the MELO estimate, $\hat{\theta}^*$, is equal to the reciprocal of the posterior mean, $1/\bar{\mu}$, times a factor $1/(1+\operatorname{var}\mu/\bar{\mu}^2)$ that has a value between zero and one. This factor depends on the squared coefficient of variation of the posterior pdf for μ. When the posterior pdf for μ is very sharp, as would be the case when the sample size n is large, $\hat{\theta}^*$ is close to $1/\bar{\mu}$. However, when $\operatorname{var}\mu/\bar{\mu}^2$ is large as would be the case when $\operatorname{var}\mu$ is large and/or $\bar{\mu}^2$ is small, $|\hat{\theta}^*| < 1/|\bar{\mu}|$. Thus the factor $1/(1+\operatorname{var}\mu/\bar{\mu}^2)$ can be viewed as acting to 'shrink' $1/\bar{\mu}$. Generally the estimate $\hat{\theta}^*$ in (2.4) will be employed when $|\bar{\mu}|/\sigma_\mu > 1$, a condition giving evidence that a value of μ in close proximity to zero is not very probable and also resulting in $\partial|\hat{\theta}^*|/\partial|\bar{\mu}| < 0$, a reasonable condition. When $|\bar{\mu}|/\sigma_\mu < 1$, the posterior pdf for θ will often show pronounced bimodality and the problem of obtaining a point estimate for θ is more complicated. In many cases prior information about the algebraic sign of μ, e.g. $\mu > 0$, is sufficient to provide a unimodal posterior pdf for θ – see the appendix for details.

We now turn to the evaluation of $\hat{\theta}^*$ in (2.4) in some special cases. Suppose that the y_i's in (2.1) are normally and independently distributed, each with mean μ and common known variance σ_0^2. Then the likelihood function is given by $l(\mu|y, \sigma_0^2) \propto \exp\{-n(\mu-\bar{y})^2/2\sigma_0^2\}$, where '$\propto$' denotes proportionality and $\bar{y} = \sum_{i=1}^n y_i/n$, the sample mean. If we employ a diffuse prior pdf for $\mu, p(\mu) \propto$ const., $-\infty < \mu < \infty$, the posterior pdf for μ is in the following normal form, $p(\mu|y, \sigma_0^2) \propto \exp\{-n(\mu-\bar{y})^2/2\sigma_0^2\}$, with posterior

[3]For the loss functions, $L(\theta, \hat{\theta}) = (\theta-\hat{\theta})^2/\theta^{2r}$, $r = 1, 2, \ldots$, where $\theta = 1/\mu$, the MELO estimate for θ is $\hat{\theta}^* = E\mu^{2r-1}/E\mu^r$, $r = 1, 2, \ldots$, where $E \equiv$ posterior expectation operator.

mean, $\bar{\mu} = \bar{y}$, and posterior variance, $\operatorname{var}\mu = \sigma_0^2/n$.[4] Thus, for this set of assumptions, the MELO estimate $\hat{\theta}^*$ in (2.4) is given by

$$\hat{\theta}^* = \frac{1}{\bar{y}}\left(\frac{1}{1+\sigma_0^2/n\bar{y}^2}\right), \tag{2.5}$$

which viewed as an estimator has finite moments (see section 4) and hence bounded risk relative to the loss function in (2.2). Further, $\tilde{\theta} = 1/\bar{y}$ is the ML estimator for $\theta = 1/\mu$. Since $\tilde{\theta}$ is the reciprocal of a normal variable, $\bar{y}$, its moments do not exist and it has infinite risk relative to a quadratic loss function, the loss function in (2.2), and many other loss functions.

When the observations in (2.1) are normally and independently distributed, each with mean μ and variance σ^2 and both μ and σ^2 have unknown values, we shall analyze the problem of estimating $\theta = 1/\mu$ under the assumption that our prior information about parameters' values is diffuse.[5] That is, we shall employ a standard diffuse prior pdf, namely, $p(\mu,\sigma) \propto 1/\sigma$, $-\infty < \mu < \infty$ and $0 < \sigma < \infty$. When this prior pdf is combined with the normal likelihood function to obtain the joint posterior pdf for μ and σ and then σ is integrated out, the marginal posterior pdf for μ is well-known to be in the following univariate Student-t form,

$$p(\mu|\underline{y}, PI) \propto \{\nu s^2 + n(\mu-\bar{y})^2\}^{-(\nu+1)/2},$$

where $\nu = n-1$ and $\nu s^2 = \sum_{i=1}^{n}(y_i-\bar{y})^2$. For this posterior pdf, we have $E\mu = \bar{y}$ and $\operatorname{var}\mu = \nu s^2/n(\nu-2)$ for $\nu > 2$. Thus from (2.4), the MELO estimate for $\theta = 1/\mu$, is given by

$$\hat{\theta}^* = \frac{1}{\bar{y}}\left(\frac{1}{1+\nu s^2/n(\nu-2)\bar{y}^2}\right), \qquad \nu = n-1 > 2. \tag{2.6}$$

In (2.6), the MELO estimate $\hat{\theta}^*$ is in the form of a product of the reciprocal of the sample mean, $1/\bar{y}$, times a shrinking factor that has a value between zero and one.[6] Further (2.6) viewed as an estimator has finite moments and hence bounded risk. Additional properties of the estimators in (2.5) and (2.6) will be established in section 4. Among these is the property that these

[4]Since μ has a normal posterior pdf, the posterior mean of $\theta = 1/\mu$ does not exist. Further, if a proper natural conjugate prior pdf for μ were employed, the posterior pdf would be normal and the posterior mean of $\theta = 1/\mu$ would still fail to exist. However, (2.4) can be readily evaluated with either a diffuse or a natural conjugate prior distribution.

[5]The analysis can easily be extended to the case of a proper natural conjugate prior distribution for μ and σ or to certain priors that impose restrictions on the range of μ, e.g. $0 < \mu < \infty$.

[6]The condition $|\bar{\mu}|/\sigma_\mu > 1$, mentioned above, here specializes to $|\bar{y}|/(s/n^{1/2}) > (\nu/(\nu-2))^{1/2}$, obviously related to a sampling theory pre-test of the hypothesis $\mu = 0$.

estimators have the same *large sample* normal distribution as the ML estimator and hence are consistent and asymptotically efficient.

The problem of estimating the reciprocal of a regression coefficient often arises in practice. For example, in the simple Haavelmo consumption model, the reciprocal of the 'multiplier' is equal to the marginal propensity to save. Similarly in a simple quantity theory model the reciprocal of the velocity coefficient is a parameter of the money demand function. If our model for the observations is a simple regression model, $y_i = \pi x_i + u_i$, $i = 1, 2, \ldots, n$, where x_i is an independent variable and we wish to estimate $\theta = 1/\pi$ using the loss function in (2.2), we first derive the posterior pdf for π using whatever prior distribution and likelihood function that are considered appropriate. Let this posterior pdf be denoted by $p(\pi|\underline{y}, PI)$ and assume that it possesses finite first and second moments. Then using $\theta = 1/\pi$, the posterior expectation of the loss function in (2.2) can be expressed as $EL = 1 - 2\hat{\theta}E\pi + \hat{\theta}^2 E\pi^2$ and the MELO estimate for $\theta = 1/\pi$ is given by

$$\hat{\theta}^* = \frac{1}{\bar{\pi}}\left(\frac{1}{1 + \operatorname{var}\pi/\bar{\pi}^2}\right), \tag{2.7}$$

where $\bar{\pi} = E\pi$ and $\operatorname{var}\pi$ denote the posterior mean and variance of π, respectively. The MELO estimate in (2.7) is seen to be equal to the product of the reciprocal of the posterior mean of π, $1/\bar{\pi}$, and a 'shrinking' factor that has a value between zero and one.

To evaluate (2.7) for a particular case, assume that the data process, $y_i = \pi x_i + u_i$ is a simple normal regression, that is the u_i's are normally and independently distributed, each with zero mean and variance σ^2. If our prior information regarding the values of π and σ is vague, we can employ the following diffuse prior pdf, $p(\pi, \sigma) \propto 1/\sigma$, $-\infty < \pi < \infty$ and $0 < \sigma < \infty$.[7] The joint posterior pdf for π and σ is then given by

$$p(\pi, \sigma|\underline{y}, PI) \propto \sigma^{-(n+1)} \exp\{-[vs^2 + (\pi - \hat{\pi})^2 m_{xx}]/2\sigma^2\},$$

where $v = n - 1$, $\hat{\pi} = \sum_{i=1}^n x_i y_i / m_{xx}$, $m_{xx} = \sum_{i=1}^n x_i^2$, and $vs^2 = \sum_{i=1}^n (y_i - \hat{\pi} x_i)^2$. The marginal posterior pdf for π is then

$$p(\pi|\underline{y}, PI) \propto \{vs^2 + (\pi - \hat{\pi})^2 m_{xx}\}^{-(v+1)/2},$$

with posterior mean $\bar{\pi} = \hat{\pi}$ and $\operatorname{var}\pi \doteq vs^2/(v-2)m_{xx}$ for $v > 2$. Thus for

[7]The analysis can easily be extended to the case of a proper natural conjugate prior pdf for π and σ or to certain priors that limit the range of π, e.g., $0 < \pi < \infty$.

these assumptions, (2.7) becomes

$$\hat{\theta}^* = \frac{1}{\hat{\pi}}\left(\frac{1}{1+vs^2/(v-2)m_{xx}\hat{\pi}^2}\right), \qquad v = n-1 > 2. \tag{2.8}$$

It is seen that the MELO estimate in (2.8) is the product of $1/\hat{\pi}$, the ML estimator, and a 'shrinking' factor that has a value between zero and one.[8] It is well-known that the ML estimator $\tilde{\theta} = 1/\hat{\pi}$ does not possess finite moments, see e.g. Bergstrom (1962), and has infinite risk relative to quadratic and other loss functions. On the other hand, as shown in section 4, the MELO estimator in (2.8) has finite moments and bounded risk relative to quadratic and other loss functions. Also, as the sample size gets large, the MELO estimator in (2.8) and the ML estimator have the same large sample normal distribution.

2.2. *Ratios of regression coefficients and population means*

If our problem is to estimate the ratio of two parameters, say the ratio of two multiple regression coefficients, $\theta = \beta_1/\beta_2$, where β_1 and β_2 are multiple regression coefficients, we have $\beta_1 - \beta_2\theta = 0$. If an estimate of θ is inserted in this last expression, $\varepsilon = \beta_1 - \beta_2\hat{\theta}$, where ε measures the extent to which the 'restriction' $\beta_1 - \beta_2\theta = 0$ fails to be satisfied. Letting our loss L be $L = \varepsilon^2 = (\beta_1 - \beta_2\hat{\theta})^2$, we have

$$L = (\beta_1 - \beta_2\hat{\theta})^2 = \beta_2^2(\theta - \hat{\theta})^2. \tag{2.9}$$

In (2.9) loss is given by the product of β_2^2 and $(\theta - \hat{\theta})^2$, a generalized quadratic loss function. Note that for a *given* value of $|\theta - \hat{\theta}|$, L is larger when β_2^2 is large than when it is small. The larger β_2^2, ceteris paribus, the smaller $|\theta|$ and (2.9) provides a greater loss for a given absolute error $|\theta - \hat{\theta}|$. This contrasts with the implications of a squared error loss function, $(\theta - \hat{\theta})^2$ that provides the same loss for a given absolute error no matter what the value of θ. Thus to a large extent (2.9) behaves in a fashion similar to the relative squared error loss function introduced above.

Given that we have a posterior pdf for β_1 and β_2 that possesses finite first and second moments, we can compute posterior expected loss and obtain a MELO estimate for $\theta = \beta_1/\beta_2$ as follows:

$$EL = E(\beta_1 - \beta_2\hat{\theta})^2 = E\beta_1^2 - 2\hat{\theta}E\beta_1\beta_2 + \hat{\theta}^2 E\beta_2^2, \tag{2.10}$$

[8]Here the condition for $\partial|\hat{\theta}^*|/\partial|\hat{\pi}| < 0$ is $|\hat{\pi}|/sm_{xx}^{-1/2} > ((v-2)/v)^{1/2}$ obviously related to a sampling theory pre-test of the hypothesis $\pi = 0$. See Zellner (1976b, p. 622) for a discussion of pre-testing.

where E denotes the expectation with respect to the posterior pdf for β_1 and β_2. From (2.10), the $\hat{\theta}$ that minimizes EL, denoted by $\hat{\theta}^*$, is given by

$$\begin{aligned}\hat{\theta}^* &= E\beta_1\beta_2/E\beta_2^2 \\ &= [\bar{\beta}_1\bar{\beta}_2+\operatorname{cov}(\beta_1,\beta_2)]/[\bar{\beta}_2^2+\operatorname{var}\beta_2] \\ &= (\bar{\beta}_1/\bar{\beta}_2)[1+\operatorname{cov}(\beta_1,\beta_2)/\bar{\beta}_1\bar{\beta}_2]/(1+\operatorname{var}\beta_2/\bar{\beta}_2^2),\end{aligned} \tag{2.11}$$

where $\bar{\beta}_1$ and $\bar{\beta}_2$ are posterior means, $\operatorname{var}\beta_2$ is the posterior variance of β_2 and $\operatorname{cov}(\beta_1,\beta_2)$ is the posterior covariance of β_1 and β_2. Given a particular posterior pdf for β_1 and β_2, the MELO estimate in (2.11) can be evaluated as shown below.[9]

If instead of the ratio of multiple regression coefficients, we consider the ratio of two population means, $\theta=\mu_1/\mu_2$, we have $\mu_1-\mu_2\theta=0$ and $\varepsilon=\mu_1-\mu_2\hat{\theta}$, where $\hat{\theta}$ is an estimate of θ. Then just as above we take our loss to be

$$L=\varepsilon^2=(\mu_1-\mu_2\hat{\theta})^2=\mu_2^2(\theta-\hat{\theta})^2.$$

Given a posterior pdf for μ_1 and μ_2 that has finite first and second moments, we have $EL=E\mu_1^2-2\hat{\theta}E\mu_1\mu_2+\hat{\theta}^2E\mu_2^2$ and thus the MELO estimate, $\hat{\theta}^*$, for $\theta=\mu_1/\mu_2$ is

$$\begin{aligned}\hat{\theta}^* &= E\mu_1\mu_2/E\mu_2^2 \\ &= (\bar{\mu}_1/\bar{\mu}_2)[1+\operatorname{cov}(\mu_1,\mu_2)/\bar{\mu}_1\bar{\mu}_2]/(1+\operatorname{var}\mu_2/\bar{\mu}_2^2),\end{aligned} \tag{2.12}$$

where $\bar{\mu}_1$ and $\bar{\mu}_2$ are posterior means for μ_1 and μ_2, respectively, and $\operatorname{cov}(\mu_1,\mu_2)$ and $\operatorname{var}\mu_2$ are the posterior covariance of μ_1 and μ_2 and the posterior variance of μ_2, respectively.

Ratios of parameters are often also encountered in analyses of simple simultaneous equation models. For example, let $y_{1t}=\gamma y_{2t}+u_{1t}$ and $y_{2t}=\beta x_t+u_{2t}$ be two structural equations. The reduced form equations for this system are $y_{1t}=\pi_1x_t+v_{1t}$ and $y_{2t}=\pi_2x_t+v_{2t}$ with $\pi_1=\beta\gamma$ and $\pi_2=\beta$. Thus the structural coefficient γ is given by $\gamma=\pi_1/\pi_2$, the ratio of two regression coefficients of a bivariate regression system, the reduced form equations. From $\gamma=\pi_1/\pi_2$, we have $\pi_1-\pi_2\gamma=0$, the restriction on π_1 and π_2. Let $\hat{\gamma}$ be an estimate of γ and let our loss be $L=\varepsilon^2=(\pi_1-\pi_2\hat{\gamma})^2=\pi_2^2(\gamma-\hat{\gamma})^2$. Then

[9]For $\bar{\beta}_1,\bar{\beta}_2>0$, $\partial\hat{\theta}^*/\partial\bar{\beta}_1>0$ and $\partial\hat{\theta}^*/\partial\bar{\beta}_2<0$ if $(\bar{\beta}_1/\bar{\beta}_2)/2<\hat{\theta}^*$, with $\hat{\theta}^*$ given in (2.11). This last inequality is equivalent to $\frac{1}{2}<[1+\operatorname{cov}(\beta_1,\beta_2)/\bar{\beta}_1\bar{\beta}_2]/[1+\operatorname{var}\beta_2/\bar{\beta}_2^2]$, a condition on the factor that multiplies $\bar{\beta}_1\bar{\beta}_2$ in the second line of (2.11) and on the posterior pdf for β_1 and β_2. If $\operatorname{cov}(\beta_1,\beta_2)=0$, the condition reduces to $\operatorname{var}\beta_2/\bar{\beta}_2^2<1$ or $\|\bar{\beta}^2\|(\operatorname{var}\beta_2)^{1/2}>1$, a condition implying that β_2's value is probably not close to zero.

$EL = E\pi_1^2 - 2\hat{\gamma}E\pi_1\pi_2 + \hat{\gamma}^2\pi_2^2$ and the MELO estimate for $\gamma = \pi_1/\pi_2$ is

$$\begin{aligned}\hat{\gamma}^* &= E\pi_1\pi_2/E\pi_2^2 \\ &= (\bar{\pi}_1/\bar{\pi}_2)[1+\mathrm{cov}(\pi_1,\pi_2)/\bar{\pi}_1\bar{\pi}_2]/(1+\mathrm{var}\,\pi_2/\bar{\pi}_2^2),\end{aligned} \tag{2.13}$$

where $\bar{\pi}_1$ and $\bar{\pi}_2$ are posterior means for π_1 and π_2, respectively, and $\mathrm{cov}(\pi_1,\pi_2)$ and $\mathrm{var}\,\pi_2$ are the posterior covariance of π_1 and π_2 and the posterior variance of π_2, respectively.

To illustrate applications of the above analysis, we first consider a normal multiple regression problem, $\underline{y} = X\underline{\beta} + \underline{u}$, where $\underline{y}$ is an $n\times 1$ vector, X is a given $n\times k$ matrix with rank k, $\underline{\beta}$ is a $k\times 1$ vector of regression coefficients with unknown values, and $\underline{u}$ is an $n\times 1$ disturbance vector. It is assumed that the elements of $\underline{u}$ have been independently drawn from a normal distribution with zero mean and variance σ^2 whose value is not known. Under these assumptions, the likelihood function is

$$l(\underline{\beta},\sigma|\underline{y},X) \propto \sigma^{-n}\exp\{-[\nu s^2 + (\underline{\beta}-\hat{\underline{\beta}})'X'X(\underline{\beta}-\hat{\underline{\beta}})]/2\sigma^2\},$$

where $\nu = n-k$, $\hat{\underline{\beta}} = (X'X)^{-1}X'\underline{y}$ and $\nu s^2 = (\underline{y}-X\hat{\underline{\beta}})'(\underline{y}-X\hat{\underline{\beta}})$. If our prior information about the parameters' values is vague[10] and we represent it by the usual following diffuse prior, $p(\underline{\beta},\sigma)\propto 1/\sigma$, $0<\sigma<\infty$ and $-\infty<\beta_i<\infty$, $i = 1,2,\dots,k$, the joint posterior pdf for $\underline{\beta}$ and σ is

$$p(\underline{\beta},\sigma|\underline{y},X,PI) \propto \sigma^{-(n+1)}\exp\{-[\nu s^2 + \underline{\beta}-\hat{\underline{\beta}})'X'X(\underline{\beta}-\hat{\underline{\beta}})]/2\sigma^2\}.$$

On integrating this last expression with respect to σ, $0<\sigma<\infty$, the marginal posterior pdf for $\underline{\beta}$ is in the following well-known multivariate Student-t form,

$$p(\underline{\beta}|\underline{y},X,PI) \propto \{\nu s^2 + (\underline{\beta}-\hat{\underline{\beta}})'X'X(\underline{\beta}-\hat{\underline{\beta}})\}^{-(\nu+k)/2}.$$

Then the posterior mean and covariance matrix for $\underline{\beta}$ are $E\underline{\beta} = \hat{\underline{\beta}}$ and $V(\underline{\beta}) = (X'X)^{-1}\nu s^2/(\nu-2)$, for $\nu>2$. From these results we know all that is needed to evaluate the MELO estimate in (2.11) for $\theta = \beta_1/\beta_2$, where β_1 and β_2 are the first and second elements of the regression coefficient vector $\underline{\beta}$. Thus, for this problem, (2.11) specializes as follows:

$$\hat{\theta}^* = (\hat{\beta}_1/\hat{\beta}_2)[1+m^{12}\bar{s}^2/\hat{\beta}_1\hat{\beta}_2]/(1+m^{22}\bar{s}^2/\hat{\beta}_2^2), \qquad \nu = n-k>2, \tag{2.14}$$

[10] The analysis can easily be extended to the case in which a proper natural conjugate prior distribution is employed.

where $\hat{\beta}_1$ and $\hat{\beta}_2$ are the first two elements of $\underline{\hat{\beta}} = (X'X)^{-1}X'\underline{y}$, m^{ij} is the (i,j)th element of $(X'X)^{-1}$ and

$$\bar{s}^2 = vs^2/(v-2) = (\underline{y}-X\underline{\hat{\beta}})'(\underline{y}-X\underline{\hat{\beta}})/(v-2).$$

It is seen that the MELO estimator $\hat{\theta}^*$ for $\theta = \beta_1/\beta_2$ in (2.14) is equal to $\hat{\beta}_1/\hat{\beta}_2$, the ML estimator, times a factor that depends on the relative second-order posterior moments of β_1 and β_2. It is well-known that the moments of the ML estimator $\tilde{\theta} = \hat{\beta}_1/\hat{\beta}_2$ do not exist whereas the moments of the MELO estimator in (2.14) do exist. Thus the MELO and ML estimators have very different finite sample properties; however as the sample size gets large, their large sample distributions become identical – see section 4.

The problem of estimating the ratio of two population means, $\theta = \mu_1/\mu_2$, can be solved using analysis similar to that employed above. As a specific example, let $y_{1i} = \mu_1 + u_{1i}$ and $y_{2i} = \mu_2 + u_{2i}$, $i = 1,2,\ldots,n$, where μ_1 and μ_2 are population means, y_{1i} and y_{2i} are observations and u_{1i} and u_{2i} are disturbance terms. Here we shall assume that the u_{1i}'s and u_{2i}'s are normally and independently distributed with zero means, that the u_{1i}'s have common variance σ_1^2 and u_{2i}'s have common variance σ_2^2. Under these assumptions, the likelihood function is

$$l(\mu_1,\mu_2,\sigma_1,\sigma_2|\underline{y}_1,\underline{y}_2) \propto \sigma_1^{-n}\sigma_2^{-n}\exp\{-[vs_1^2+n(\mu_1-\bar{y}_1)^2]/2\sigma_1^2\}$$
$$\times \exp\{-[vs_2^2+n(\mu_2-\bar{y}_2)^2]/2\sigma_2^2\},$$

where $v = n-1$, $\bar{y}_1 = \sum_{i=1}^n y_{1i}/n$, $\bar{y}_2 = \sum_{i=1}^n y_{2i}/n$, $vs_1^2 = \sum_{i=1}^n (y_{1i}-\bar{y}_1)^2$ and $vs_2^2 = \sum_{i=1}^n (y_{2i}-\bar{y}_2)^2$.

If we employ the following diffuse prior pdf, $p(\mu_1,\mu_2,\sigma_1,\sigma_2) \propto 1/\sigma_1\sigma_2$, $0<\sigma_i<\infty$ and $-\infty<\mu_i<\infty$, $i=1,2$, the marginal posterior pdf for μ_1 and μ_2 is given by

$$p(\mu_1,\mu_2|\underline{y}_1,\underline{y}_2,PI) \propto \{vs_1^2+n(\mu_1-\bar{y}_1)^2\}^{-(v+1)/2}$$
$$\times \{vs_2^2+n(\mu_2-\bar{y}_2)^2\}^{-(v+1)/2},$$

with posterior means, $E\mu_1 = \bar{y}_1$ and $E\mu_2 = \bar{y}_2$ and posterior variances, $\operatorname{var}\mu_1 = vs_1^2/n(v-2)$ and $\operatorname{var}\mu_2 = vs_2^2/n(v-2)$ for $v>2$. For this problem the posterior covariance of μ_1 and μ_2 is zero. Thus the MELO estimator for $\theta = \mu_1/\mu_2$ is given from (2.12) as follows:

$$\hat{\theta}^* = (\bar{y}_1/\bar{y}_2)\{1/(1+vs_2^2/n(v-2)\bar{y}_2^2)\}, \qquad v = n-1 > 2. \tag{2.15}$$

It is seen that (2.15) is in the form of a product of the ML estimator, $\tilde{\theta} = \bar{y}_1/\bar{y}_2$, times a factor that depends on the coefficient of variation of the posterior pdf for μ_2. While the ML estimator $\tilde{\theta} = \bar{y}_1/\bar{y}_2$ does not possess finite moments, the MELO estimator in (2.15) has finite moments and hence bounded risk relative to quadratic and other loss functions.

As a last case, consider the simple bivariate normal regression system, $y_{1t} = \pi_1 x_t + v_{1t}$ and $y_{2t} = \pi_1 x_t + v_{1t}$, $t = 1,2,\dots,T$, where the pairs of error terms (v_{1t}, v_{2t}) are normally and independently distributed, each with zero mean vector and common 2×2 pds covariance matrix Ω. Given that we employ a diffuse prior pdf for π_1, π_2 and the distinct elements of Ω, $p(\pi_1,\pi_2,\Omega) \propto |\Omega|^{-1}$, it is known that the marginal posterior pdf for π_1 and π_2 is in the bivariate Student-t form with posterior mean $\hat{\underline{\pi}}' = (\hat{\pi}_1, \hat{\pi}_2)$ where

$$\hat{\pi}_i = \sum_{t=1}^{T} x_t y_{it} / \sum x_t^2, \qquad i = 1,2,$$

and posterior covariance matrix $\bar{S}/m_{xx}$, where $m_{xx} = \sum x_t^2$ and $\bar{S}$ is a symmetric 2×2 matrix with typical element,

$$\bar{s}_{ij} = \sum (y_{it} - x_t\hat{\pi}_i)(y_{jt} - x_t\hat{\pi}_j)/(v-2), \qquad i,j = 1,2,$$

with $v = n-1 > 2$. On inserting these posterior moments in (2.13), the MELO estimate is

$$\hat{\gamma}^* = (\hat{\pi}_1/\hat{\pi}_2)(1+\bar{s}_{12}/m_{xx}\hat{\pi}_1\hat{\pi}_2)/(1+\bar{s}_{22}/m_{xx}\hat{\pi}_2^2). \tag{2.16}$$

Again it is the case that the MELO estimate $\hat{\gamma}^*$ is the product of the ML estimate, $\hat{\pi}_1/\hat{\pi}_2$, times a 'correction' factor.

For the reader's convenience, some of the results in this section are presented in tabular form in table 1.

3. MELO estimates for structural coefficients

In this section we derive MELO estimates for structural coefficients of linear structural econometric models. Initially we consider a single structural equation and then go on to consider joint estimation of sets of structural equations' parameters.

3.1. 'Single equation' analysis

Let a structural equation of a model, say the first equation that is assumed

Table 1
Tabular summary of MELO estimates.

Problem	Loss function	MELO estimate[a]	Eq. in text
1. Reciprocal mean, $\theta = 1/\mu$	$L = ((\theta-\hat{\theta})/\theta)^2$	A. $\hat{\theta}^* = (1/\bar{\mu})(1/(1+\operatorname{var}\mu/\bar{\mu}^2))$	(2.4)
		B. $\hat{\theta}^* = (1/\bar{y})(1/[1+\nu s^2/n(\nu-2)])$	(2.6)
2. Reciprocal of regression coefficient, $\theta = 1/\pi$	$L = ((\theta-\hat{\theta})/\theta)^2$	A. $\hat{\theta}^* = (1/\bar{\pi})(1/(1+\operatorname{var}\pi/\bar{\pi}^2))$	(2.7)
		B. $\hat{\theta}^* = (1/\hat{\pi})(1/[1+\nu s^2/(\nu-2)m_{xx}\hat{\pi}^2])$	(2.8)
2. Ratio of multiple regression coefficients, $\theta = \beta_1/\beta_2$	$L = \beta_2^2(\theta-\hat{\theta})^2$	A. $\hat{\theta}^* = (\bar{\beta}_1/\bar{\beta}_2)[1+\operatorname{cov}(\beta_1,\beta_2)/\bar{\beta}_1\bar{\beta}_2]/(1+\operatorname{var}\beta_1/\bar{\beta}_2^2)$	(2.11)
		B. $\hat{\theta}^* = (\hat{\beta}_1/\hat{\beta}_2)(1+m^{12}\bar{s}^2/\hat{\beta}_1\hat{\beta}_2)/(1+m^{22}\bar{s}^2/\hat{\beta}_2^2)$	(2.14)
4. Ratio of population means, $\theta = \mu_1/\mu_2$	$L = \mu_2^2(\theta-\hat{\theta})^2$	A. $\hat{\theta}^* = (\bar{\mu}_1/\bar{\mu}_2)[1+\operatorname{cov}(\mu_1,\mu_2)/\bar{\mu}_1\bar{\mu}_2]/(1+\operatorname{var}\mu_2/\bar{\mu}_2^2)$	(2.12)
		B. $\hat{\theta}^* = (\bar{y}_1/\bar{y}_2)/[1+\nu s_2^2/n(\nu-2)\bar{y}_2^2]$	(2.15)
5. Ratio of bivariate regression coefficients, $\gamma = \pi_1/\pi_2$	$L = \pi_2^2(\gamma-\hat{\gamma})^2$	A. $\hat{\gamma}^* = (\bar{\pi}_1/\bar{\pi}_2)[1+\operatorname{cov}(\pi_1,\pi_2)/\bar{\pi}_1\bar{\pi}_2]/(1+\operatorname{var}\pi_2/\bar{\pi}_2^2)$	(2.13)
		B. $\hat{\gamma}^* = (\hat{\pi}_1/\hat{\pi}_2)(1+\bar{s}_{12}/m_{xx}\hat{\pi}_1\hat{\pi}_2)/(1+\bar{s}_{22}/m_{xx}\hat{\pi}_2^2)$	(2.16)

[a]Case A is the general MELO estimate expressed in terms of posterior moments. Case B is a special case of Case A in which normal data processes and diffuse prior distributions are employed.

identified, be given by

$$\underset{n\times 1}{\underline{y}_1} = \underset{n\times m_1}{Y_1}\ \underset{m_1\times 1}{\underline{\gamma}_1} + \underset{n\times k_1}{X_1}\ \underset{k_1\times 1}{\underline{\beta}_1} + \underset{n\times 1}{\underline{u}_1}, \tag{3.1}$$

where $\underline{y}_1$ and Y_1 are an $n\times 1$ vector and an $n\times m_1$ matrix, respectively, of observations on $m = m_1+1$ endogenous variables, X_1 is an $n\times k_1$ matrix of rank k_1 of observations on k_1 predetermined variables, $\underline{\gamma}_1$ and $\underline{\beta}_1$ are $m_1\times 1$ and $k_1\times 1$ vectors of structural coefficients and $\underline{u}_1$ is an $n\times 1$ vector of structural disturbances. X_1 is a sub-matrix of the $n\times k$ matrix X of observations on all predetermined variables in the model, that is $X = (X_1 \vdots X_0)$ where X_0 is $n\times k_0$, $k = k_1+k_0$, and it is assumed that X has rank k.

The reduced form equations for $\underline{y}_1$ and Y_1 are given by

$$(\underset{n\times 1}{\underline{y}_1} \vdots \underset{n\times m_1}{Y_1}) = \underset{n\times k}{X}\ (\underset{k\times 1}{\underline{\pi}_1} \vdots \underset{k\times m_1}{\Pi_1}) + (\underset{n\times 1}{\underline{v}_1} \vdots \underset{n\times m_1}{V_1}), \tag{3.2}$$

where $\underline{\pi}_1$ is a $k\times 1$ vector and Π_1 is a $k\times m_1$ matrix of reduced form coefficients and $\underline{v}_1$ is an $n\times 1$ vector and V_1 is an $n\times m_1$ matrix of reduced form disturbances.

On multiplying both sides of (3.2) on the right by $(1 \vdots -\underline{\gamma}_1')'$ and comparing the result with (3.1), we obtain the well-known results

$$\underline{\pi}_1 = \Pi_1\underline{\gamma}_1 + \begin{pmatrix}\underline{\beta}_1\\ \underline{0}\end{pmatrix}, \tag{3.3a}$$

and

$$\underline{v}_1 - V_1\underline{\gamma}_1 = \underline{u}_1, \tag{3.3b}$$

where the zero vector in (3.3a) is $(k-k_1)\times 1$ with $k-k_1 \geqq m_1$ the necessary order condition for identification implied by the usual rank condition for identification of $\underline{\gamma}_1$ and $\underline{\beta}_1$. In (3.3a) we have the restrictions on the reduced form coefficients that involve the parameter vectors $\underline{\gamma}_1$ and $\underline{\beta}_1$ that we wish to estimate.

To formulate a loss function for the problem of estimating $\underline{\gamma}_1$ and $\underline{\beta}_1$ multiply both sides of (3.3a) on the left by $X = (X_1 \vdots X_0)$ to obtain $X\underline{\pi}_1 = X\Pi_1\underline{\gamma}_1 + X_1\underline{\beta}_1$, or $X\underline{\pi}_1 = \bar{Z}_1\underline{\delta}_1$, where $\bar{Z}_1 = (X\Pi_1 \vdots X_1)$ and $\underline{\delta}_1' = (\underline{\gamma}_1' \vdots \underline{\beta}_1')$. If $\hat{\underline{\delta}}_1$ is an estimate of $\underline{\delta}_1$, let $\underline{\varepsilon} = X\underline{\pi}_1 - \bar{Z}_1\hat{\underline{\delta}}_1$ and define the loss function as follows:

$$\begin{aligned} L = \underline{\varepsilon}'\underline{\varepsilon} &= (X\underline{\pi}_1 - \bar{Z}_1\hat{\underline{\delta}}_1)'(X\underline{\pi}_1 - \bar{Z}_1\hat{\underline{\delta}}_1) \\ &= (\underline{\delta}_1 - \hat{\underline{\delta}}_1)'\bar{Z}_1'\bar{Z}_1(\underline{\delta}_1 - \hat{\underline{\delta}}_1), \end{aligned} \tag{3.4}$$

where in going from the first to the second line of (3.4) $X\underline{\pi}_1 = \bar{Z}_1\underline{\delta}_1$ has been employed. Thus the loss function in (3.4) is quadratic in $\underline{\delta}_1 - \hat{\underline{\delta}}_1$ with a pds matrix, $\bar{Z}_1'\bar{Z}_1$, where $\bar{Z}_1 = (X\Pi_1 \vdots X_1)$.

Given a posterior pdf for $\Pi = (\underline{\pi}_1 \vdots \Pi_1)$ that possesses finite first and second moments, posterior expected loss, evaluated from the first line of (3.4) is

$$EL = E\underline{\pi}_1'X'X\underline{\pi}_1 - 2\hat{\underline{\delta}}_1'E\bar{Z}_1'X\underline{\pi}_1 + \hat{\underline{\delta}}_1'E\bar{Z}_1'\bar{Z}_1\hat{\underline{\delta}}_1, \tag{3.5}$$

and the value of $\hat{\underline{\delta}}_1, \hat{\underline{\delta}}_1^*$ that minimizes expected loss is

$$\begin{aligned} \hat{\underline{\delta}}_1^* &= (E\bar{Z}_1'\bar{Z}_1)^{-1}E\bar{Z}_1'X\underline{\pi}_1 \\ &= \begin{bmatrix} E\Pi_1'X'X\Pi_1 & E\Pi_1'X'X_1 \\ EX_1'X\Pi_1 & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} E\Pi_1'X'X\underline{\pi}_1 \\ EX_1'X\underline{\pi}_1 \end{bmatrix}. \end{aligned} \tag{3.6}$$

$\hat{\underline{\delta}}_1^*$ in (3.6) is the MELO estimate relative to the loss function in (3.4) and whatever posterior pdf is employed in evaluating the posterior moments in (3.6).[11]

As an explicit practical example illustrating application of (3.6), we write the reduced form system in (3.2) as

$$Y = X\Pi + V, \tag{3.7}$$

where $Y = (\underline{y}_1 \vdots Y_1)$, $\Pi = (\underline{\pi}_1 \vdots \Pi_1)$ and $V = (\underline{v}_1 \vdots V_1)$, and assume that the rows of V are normally and independently distributed, each with zero mean vector and common pds $m \times m$ covariance matrix Ω, with $m = m_1 + 1$. If we employ a diffuse prior pdf for Π and Ω, $p(\Pi, \Omega) \propto |\Omega|^{-(m+1)/2}$, it is well-known [see e.g. Zellner (1971a, p. 229)] that the marginal posterior pdf for Π is in the following matrix Student-t form:

$$p(\Pi|Y, PI) \propto |S + (\Pi - \hat{\Pi})'X'X(\Pi - \hat{\Pi})|^{-n/2}, \tag{3.8}$$

where $\hat{\Pi} = (X'X)^{-1}X'Y$, the posterior mean of Π, and

$$S = (Y - X\hat{\Pi})'(Y - X\hat{\Pi}) = \hat{V}'\hat{V}.$$

Letting

$$\Pi - \hat{\Pi} = (\underline{\pi}_1 - \hat{\underline{\pi}}_1, \underline{\pi}_2 - \hat{\underline{\pi}}_2, \ldots, \underline{\pi}_m - \hat{\underline{\pi}}_m),$$

[11] $\hat{\underline{\delta}}_1^*$ is an estimate of $\underline{\delta}_1$ that appears in $X\underline{\pi}_1 = \bar{Z}_1\underline{\delta}_1$, a set of n exact equations. In general if we are interested in estimating a vector $\underline{\theta}$ appearing in $\underline{\eta} = A\underline{\theta}$, where $\underline{\eta}$ is an $n \times 1$ vector and A is an $n \times k$ matrix of rank k, let $\hat{\underline{\theta}}$ be any estimate of $\underline{\theta}$. Then if our loss function is $L = (\underline{\eta} - A\hat{\underline{\theta}})'Q(\underline{\eta} - A\hat{\underline{\theta}}) = (\underline{\theta} - \hat{\underline{\theta}})'A'QA(\underline{\theta} - \hat{\underline{\theta}})$, where Q is an $n \times n$ pds matrix, the MELO estimate for θ is $\hat{\underline{\theta}}^* = (EA'QA)^{-1}EA'Q\underline{\eta}$ where $E \equiv$ posterior expectation operator.

and

$$(\underline{\pi}-\hat{\underline{\pi}})' = (\underline{\pi}_1'-\hat{\underline{\pi}}_1', \underline{\pi}_2'-\hat{\underline{\pi}}_2', \dots, \underline{\pi}_m'-\hat{\underline{\pi}}_m'),$$

the posterior covariance matrix for $\underline{\pi}$ is [see e.g. Box–Tiao (1973, p. 477)]

$$E(\underline{\pi}-\hat{\underline{\pi}})(\underline{\pi}-\hat{\underline{\pi}})' = \bar{S}\otimes(X'X)^{-1}, \tag{3.9}$$

where $\bar{S} = (Y-X\hat{\Pi})'(Y-X\hat{\Pi})/(v-2)$, with $v = n-k-(m-1) > 2$.

Using the above results to evaluate the posterior expectations in (3.6), we have

$$\begin{aligned} E\Pi_1'X'X\Pi_1 &= E(\Pi_1-\hat{\Pi}_1+\hat{\Pi}_1)'X'X(\Pi_1-\hat{\Pi}_1+\hat{\Pi}_1) \\ &= \hat{\Pi}_1'X'X\hat{\Pi}_1 + E(\Pi_1-\hat{\Pi}_1)'X'X(\Pi_1-\hat{\Pi}_1) \\ &= \hat{\Pi}_1'X'X\hat{\Pi}_1 + k\bar{S}_{22}, \end{aligned} \tag{3.10}$$

and

$$\begin{aligned} E\Pi_1'X'X\underline{\pi}_1 &= E(\Pi_1-\hat{\Pi}_1+\hat{\Pi}_1)'X'X(\underline{\pi}_1-\hat{\underline{\pi}}_1+\hat{\underline{\pi}}_1) \\ &= \hat{\Pi}_1'X'X\hat{\underline{\pi}}_1 + E(\Pi_1-\hat{\Pi}_1)'X'X(\underline{\pi}_1-\hat{\underline{\pi}}_1) \\ &= \hat{\Pi}_1'X'X\hat{\underline{\pi}}_1 + k\bar{\underline{s}}_{12}, \end{aligned} \tag{3.11}$$

where the partitionings, $\Pi = (\underline{\pi}_1 \vdots \Pi_1)$, $\hat{\Pi} = (\hat{\underline{\pi}}_1 \vdots \hat{\Pi}_1)$ and

$$\bar{S} = (Y-X\hat{\Pi})'(Y-X\hat{\Pi})/(v-2) = \begin{pmatrix} \bar{s}_{11} & \bar{\underline{s}}_{12}' \\ \bar{\underline{s}}_{12} & \bar{S}_{22} \end{pmatrix}, \tag{3.12}$$

have been employed.[12] Substituting the results (3.10) and (3.11) into (3.6), we have

$$\hat{\underline{\delta}}_1^* = \begin{bmatrix} \hat{\Pi}_1'X'X\hat{\Pi}_1 + k\bar{S}_{22} & \hat{\Pi}_1'X'X_1 \\ X_1'X\hat{\Pi}_1 & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{\Pi}_1'X'X\hat{\underline{\pi}}_1 + k\bar{\underline{s}}_{12} \\ X_1'X\hat{\underline{\pi}}_1 \end{bmatrix}. \tag{3.13}$$

[12] In the derivation reported in (3.10), note that $E(\Pi_1-\hat{\Pi}_1)'X'X(\Pi_1-\hat{\Pi}_1) = \{E(\underline{\pi}_i-\hat{\underline{\pi}}_i)'X'X(\underline{\pi}_j-\hat{\underline{\pi}}_j)\} = \{\mathrm{tr}\,X'X\,\mathrm{cov}(\underline{\pi}_i,\underline{\pi}_j)\} = \{k\bar{s}_{ij}\} = k\bar{S}_{22}$. Similar operations yield $E(\Pi_1-\hat{\Pi}_1)'X'X(\underline{\pi}_1-\hat{\underline{\pi}}_1) = k\bar{\underline{s}}_{12}$, with $\bar{S}_{22}$ and $\bar{\underline{s}}_{12}$ defined in (3.12).

With $\underline{\hat{\delta}}_1^* = (\underline{\hat{\gamma}}_1'^*, \underline{\hat{\beta}}_1'^*)'$, we obtain the following expressions for $\underline{\hat{\gamma}}_1^*$ and $\underline{\hat{\beta}}_1^*$ from (3.13):

$$\underline{\hat{\gamma}}_1^* = (M_1 + k\bar{S}_{22})^{-1}(M_1\underline{\hat{\gamma}}_1 + k\bar{S}_{22}\underline{\hat{\hat{\gamma}}}_1), \tag{3.14}$$

where $M_1 = \hat{\Pi}_1' X[I - X_1(X_1'X_1)^{-1}X_1']X\hat{\Pi}_1$, $\underline{\hat{\gamma}}_1$ is the 2SLS estimator for $\underline{\gamma}_1$, $\underline{\hat{\hat{\gamma}}} = \bar{S}_{22}^{-1}\underline{\bar{s}}_{12} = (\hat{V}_1'\hat{V}_1)^{-1}\hat{V}_1\underline{\hat{v}}_1$, and

$$\underline{\hat{\beta}}_1^* = (X_1'X_1)^{-1}X_1'(X\underline{\hat{\pi}}_1 - X\hat{\Pi}_1\underline{\hat{\gamma}}_1^*). \tag{3.15}$$

From (3.14), it is seen that the MELO estimate for $\underline{\gamma}_1$ is a matrix weighted average of the 2SLS estimate, $\underline{\hat{\gamma}}_1$, and the estimate $\underline{\hat{\hat{\gamma}}}_1$ formed from reduced form residuals. Also (3.14) can be expressed as a matrix weighted average of the DLS and 2SLS estimates of $\underline{\gamma}_1$ – see Zellner (1976a).

Surprisingly, the estimate $\underline{\hat{\delta}}_1^*$ for $\underline{\delta}_1$ in (3.13) is in the form of a '$\mathscr{R}$-class' estimate with $\mathscr{R} = \mathscr{R}^*$ where $\mathscr{R}^* = 1 - k/(\nu - 2)$, with $\nu = n - k - (m-1) > 2$. Hence, the present analysis yields an optimal value for $\mathscr{R}$ that is less than one, the value that yields the 2SLS estimate.[13] Hatanaka (1973, pp. 12–14) has shown that in general the first two finite-sample moments of $\mathscr{R}$-class estimators exist when $\mathscr{R} < 1$ and other conditions, satisfied above, are met.[14] Thus $\underline{\hat{\delta}}_1^*$ in (3.13) has finite first and second moments and bounded risk relative to quadratic loss functions. Further it is to be noted that $\mathscr{R}^* \to 1$ and $n^{1/2}(\mathscr{R}^* - 1) \to 0$ as $n \to \infty$. Thus $\underline{\hat{\delta}}_1^*$ is consistent and asymptotically equivalent to other consistent and asymptotically normal $\mathscr{R}$-class estimators. For the reader's benefit, the following table provides values of $\mathscr{R}^*$ for various values of k, the number of variables in X and $\nu - 2 = n - k - (m-1) - 2$.

From the table it is interesting to observe that $\mathscr{R}^*$ is substantially below one in a variety of circumstances. Also when $k = \nu - 2$, $\mathscr{R}^* = 0$, which yields the 'direct' or 'ordinary' least squares estimator for $\underline{\delta}_1$. When $k > \nu - 2 = n - k - (m-1) - 2$ or $2k + m - 1 > n$, a condition under which the number of parameters is large relative to the sample size n, $\mathscr{R}^* < 0$.

We further note that the derivation of $\underline{\hat{\delta}}_1^*$ in (3.13) is appropriate when X contains lagged endogenous variables, given that initial values of such variables are taken as given in formulating the normal likelihood function for

[13]When $\mathscr{R} > 1$, Sawa (1972) showed that $\mathscr{R}$-class estimators for structural coefficients in an equation containing two endogenous variables do not possess finite moments of any order. See also Mariano (1973). In Nagar (1959) and Sawa (1972), 'approximately optimal' values of $\mathscr{R}$ that are functions of parameters with unknown values are presented.

[14]These other conditions, in terms of our notation, are: (i) $k - k_1 - m_1 \geqq 0$ (order condition for identification); (ii) $n - k - m_1 \geqq 0$; and (iii) $n - k_1 - m_1 \geqq 2$. With the assumption $\nu = n - k - (m-1) > 2$ satisfied, Hatanaka's conditions are also satisfied. Hatanaka further assumes that X does not contain lagged endogenous variables.

Table 2
Values of $\mathscr{R}^*$ for selected values of k and $\nu-2$.

		Values of $\nu-2$										
		1	2	3	4	5	6	10	20	50	...	∞
Values of k	1	0	1/2	2/3	3/4	4/5	5/6	9/10	19/20	49/50	...	1
	2	−1	0	1/3	1/2	3/5	2/3	4/5	9/10	24/25	...	1
	3	−2	−1/2	0	1/4	2/5	1/2	7/10	17/20	47/50	...	1
	4	−3	−1	−1/3	0	1/5	1/3	3/5	4/5	23/25	...	1
	5	−4	−3/2	−2/3	−1/5	0	1/6	1/2	3/4	9/10	...	1

the reduced form system, $Y = X\Pi + V$. In addition, it is possible to employ a broadened loss function $L = (\underline{\delta}_1 - \hat{\underline{\delta}}_1)'\bar{Z}_1' Q_1 \bar{Z}_1(\underline{\delta} - \hat{\underline{\delta}}_1)$, where Q_1 is a given pds matrix in place of (3.4).[15]

3.2. 'Full-system' analysis

We now turn to derive MELO estimates of structural coefficients, assumed identified, appearing in all equations of a linear simultaneous equation model. Let the αth structural equation be given by

$$\underline{y}_\alpha = YT_\alpha \underline{\gamma}_\alpha + X_\alpha \underline{\beta}_\alpha + \underline{u}_\alpha, \qquad \alpha = 1, 2, \ldots, g, \tag{3.16}$$

where $\underline{y}_\alpha$ is an $n \times 1$ vector of observations on an endogenous variable whose coefficient is equal to one, Y_α is an $n \times m_\alpha$ matrix of observations on m_α other endogenous variables appearing in the αth equation with $m_\alpha \times 1$ coefficient vector $\underline{\gamma}_\alpha$, X_α is an $n \times k_\alpha$ matrix of observations on k_α predetermined variables with $k_\alpha \times 1$ coefficient vector $\underline{\beta}_\alpha$, and $\underline{u}_\alpha$ is an $n \times 1$ disturbance vector. As usual, it is assumed that X_α is a submatrix of X, the $n \times k$ matrix of observations on all k predetermined variables that has rank k. Using the reduced form equations for $\underline{y}_\alpha$ and Y_α, $\underline{y}_\alpha = X\underline{\pi}_\alpha + \underline{v}_\alpha$ and $Y_\alpha = X\Pi_\alpha + V_\alpha$, it is possible to express (3.16) as

$$\underline{y}_\alpha = X\Pi_\alpha \underline{\gamma}_\alpha + X_\alpha \underline{\beta}_\alpha + \underline{v}_\alpha, \tag{3.17a}$$

or $\qquad \alpha = 1, 2, \ldots, g.$

$$X\underline{\pi}_\alpha = X\Pi_\alpha \underline{\gamma}_\alpha + X_\alpha \underline{\beta}_\alpha, \tag{3.17b}$$

From (3.17), we can write

$$\begin{pmatrix} X\underline{\pi}_1 \\ X\underline{\pi}_2 \\ \vdots \\ X\underline{\pi}_g \end{pmatrix} = \begin{pmatrix} X\Pi_1 & 0 & \cdots & 0 \\ 0 & X\Pi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X\Pi_g \end{pmatrix} \begin{pmatrix} \underline{\gamma}_1 \\ \underline{\gamma}_2 \\ \vdots \\ \underline{\gamma}_g \end{pmatrix}$$

$$+ \begin{pmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & X_g \end{pmatrix} \begin{pmatrix} \underline{\beta}_1 \\ \underline{\beta}_2 \\ \vdots \\ \underline{\beta}_g \end{pmatrix}, \tag{3.18a}$$

[15] In Zellner–Park (1975), the MELO estimate $\hat{\underline{\delta}}_1^*$ in (3.13) is computed for equations of Klein's Model I and of the Girshick–Haavelmo demand and supply model for food and compared with DLS and 2SLS estimates. The MELO estimates tend to lie between the DLS and 2SLS estimates.

or

$$\underline{\bar{w}} = \bar{W}\underline{\gamma}+X\underline{\beta}, \tag{3.18b}$$

or

$$\bar{w} = \bar{Z}\delta. \tag{3.18c}$$

where $\underline{\bar{w}}$ denotes the partitioned vector on the l.h.s. of (3.18a), $\bar{W}$ and $\bar{X}$ the first and second block diagonal matrices on the r.h.s. of (3.18a) $\underline{\gamma}' = (\gamma_1', \gamma_2', \ldots, \gamma_g')$, $\underline{\beta}' = (\underline{\beta}_1', \underline{\beta}_2', \ldots, \underline{\beta}_g')$, $\bar{Z} = (\bar{W} \vdots \bar{X})$ and $\delta' = (\gamma' \vdots \beta')$.

As our loss function for estimating the parameter vector $\underline{\delta}$ in (3.18), we let $\underline{\varepsilon} = \underline{\bar{w}} - \bar{Z}\underline{\hat{\delta}}$ where $\underline{\hat{\delta}}$ is an estimate of $\underline{\delta}$ and form

$$L = (\underline{\bar{w}} - \bar{Z}\underline{\hat{\delta}})'Q(\underline{\bar{w}} - \bar{Z}\underline{\hat{\delta}})$$

$$= (\underline{\delta} - \underline{\hat{\delta}})'\bar{Z}'Q\bar{Z}(\underline{\delta} - \underline{\hat{\delta}}), \tag{3.19}$$

where Q is a pds matrix that may have elements that are parameters with unknown values. Given a posterior pdf for the reduced form coefficients and elements of Q, we compute the posterior expectation of (3.19) and find the minimizing value of $\underline{\delta}$ as follows where E is the posterior expectation operator:

$$EL = E\underline{\bar{w}}'Q\underline{\bar{w}} - 2\underline{\hat{\delta}}'E\bar{Z}'Q\underline{\bar{w}} + \underline{\hat{\delta}}'\bar{Z}Q\bar{Z}\underline{\hat{\delta}}. \tag{3.20}$$

By differentiating (3.20) with respect to $\underline{\hat{\delta}}$, we find the minimum expected loss estimate to be

$$\underline{\hat{\delta}}^* = (E\bar{Z}'Q\bar{Z})^{-1}E\bar{Z}'Q\underline{\bar{w}}. \tag{3.21}$$

The MELO estimate $\underline{\hat{\delta}}^*$ is a general solution for our problem of estimating the structural coefficient vector $\underline{\delta}$ given the loss function in (3.19) and a posterior pdf for the reduced form coefficients and the elements Q. Below we shall consider the case $Q = \Omega_g^{-1} \otimes I_n$ where Ω_g is the $g \times g$ pds reduced form disturbance covariance matrix.

To evaluate (3.21) with $Q = \Omega_g^{-1} \otimes I_n$, consider the reduced form system, $Y = X\Pi + V$, where $Y = (\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_g)$ is an $n \times g$ matrix of observations on all endogenous variables in the system, X is an $n \times k$ matrix of rank k of observations on all predetermined variables, Π is a $k \times g$ matrix of reduced form coefficients and V is an $n \times g$ matrix of disturbances. It is assumed that the n rows of V have been independently drawn from a g-dimensional normal distribution with zero mean vector and $g \times g$ pds covariance matrix Ω_g. If we employ the following diffuse prior pdf for Π and Ω_g,

$p(\Pi,\Omega_g) \propto |\Omega_g|^{-(g+1)/2}$, the joint posterior pdf for Π and Ω_g is given by [see e.g. Zellner (1971a, p. 227)]

$$p(\Pi,\Omega_g|Y,PI) \propto |\Omega_g|^{-(n+g+1)/2} \times \exp\{-\tfrac{1}{2}\operatorname{tr}[\hat{\Omega}_g+(\Pi-\hat{\Pi})'X'X(\Pi-\hat{\Pi})]\Omega_g^{-1}\}, \tag{3.22}$$

where $\hat{\Pi} = (X'X)^{-1}X'Y$ and $\hat{\Omega}_g = (Y-X\hat{\Pi})'(Y-X\hat{\Pi})$. From (3.22), it is seen that the conditional posterior pdf for Π given Ω_g is in the multivariate normal form with mean $\hat{\Pi}$ and covariance matrix $\Omega_g \otimes (X'X)^{-1}$. This fact will be employed to evaluate the posterior expectations in (3.16) conditional on Ω_g. That is, with respect to $E\bar{Z}'Q\bar{Z}$ on the r.h.s. of (3.16), the (α,l)th submatrix of $\bar{Z}'Q\bar{Z} = \bar{Z}'(\Omega_g^{-1}\otimes I_n)\bar{Z}$ is $\bar{Z}_\alpha'\bar{Z}_l\omega^{\alpha l}$, where $\omega^{\alpha l}$ is the (α,l)th element of Ω_g^{-1}, $\bar{Z}_\alpha = (X\Pi_\alpha \vdots X_\alpha)$ and $\bar{Z}_l = (X\Pi_l \vdots X_l)$. Then the conditional posterior mean of the (α,l)th submatrix is given by

$$E\bar{Z}_\alpha'\bar{Z}_l\omega^{\alpha l} = [\hat{Z}_\alpha'\hat{Z}_l + E(\bar{Z}_\alpha-\hat{Z}_\alpha)'(\bar{Z}_l-\hat{Z}_l)\omega^{\alpha l}], \tag{3.23}$$

where E is the conditional (given Ω_g) posterior expectation operator, $\bar{Z}_\alpha = (X\hat{\Pi}_\alpha \vdots X_\alpha)$ and $\hat{Z}_l = (X\hat{\Pi}_l \vdots X_l)$ with $\hat{\Pi}_\alpha = (X'X)^{-1}X'Y_\alpha$ and $\hat{\Pi}_l = (X'X)^{-1}X'Y_l$ the posterior means of Π_α and Π_l, respectively. Further, the vector $E\bar{Z}'Q\underline{\bar{w}}$ on the r.h.s. of (3.21) is

$$E\bar{Z}'(\Omega_g^{-1}\otimes I_n)\underline{\bar{w}} = E\left\{\sum_{l=1}^{g} \bar{Z}_\alpha'\underline{\bar{w}}_l\omega^{\alpha l}\right\}, \qquad \alpha = 1,2,\dots,g,$$

$$= \left\{\sum_{l=1}^{g}[\hat{Z}_\alpha'\hat{w}_l + E(\bar{Z}_\alpha-\hat{Z}_\alpha)'(\underline{\bar{w}}_l-\underline{\hat{w}}_l)]\omega^{\alpha l}\right\}, \tag{3.24}$$

where $\underline{\hat{w}}_l = X\underline{\hat{\pi}}_l$, with $\hat{\pi}_l = (X'X)^{-1}X'\underline{y}_l$. Thus the optimal estimate $\underline{\hat{\delta}}^*$ given Ω_g is

$$\underline{\hat{\delta}}^* = \{[\hat{Z}_\alpha'\hat{Z}_l + E(\bar{Z}_\alpha-\hat{Z}_\alpha)'(\bar{Z}_l-\hat{Z}_l)]\omega^{\alpha l}\}^{-1} \times\left\{\sum_{l=1}^{g}[\hat{Z}_\alpha'\underline{\hat{w}}_l + E(\bar{Z}_\alpha-\hat{Z}_\alpha)'(\underline{\bar{w}}_l-\underline{\hat{w}}_l)]\omega^{\alpha l}\right\}, \tag{3.25}$$

where the entry in the first pair of curly braces is a typical submatrix, $\alpha,l=1,2,\dots,g$, and in the second pair, a typical element of a vector, $\alpha=1,2,\dots g$. Given that the conditional posterior pdf for the reduced form coefficients, Π, given Ω_g is multivariate normal with mean $\hat{\Pi} = (X'X)^{-1}X'Y$ and covariance matrix $\Omega_g \otimes (X'X)^{-1}$, the conditional posterior expectations in (3.25) can be evaluated to give

$$\underline{\hat{\delta}}^* = \left\{\left[\begin{pmatrix} \hat{Y}_\alpha' \hat{Y}_l & \hat{Y}_\alpha' X_l \\ X_\alpha' \hat{Y}_l & X_\alpha' X_l \end{pmatrix} + k\begin{pmatrix} \Omega_{\alpha l} & 0 \\ 0 & 0 \end{pmatrix}\right]\omega^{\alpha l}\right\}^{-1}$$

$$\times\left\{\sum_{l=1}^{g}\left[\begin{pmatrix} \hat{Y}_\alpha' \\ X_\alpha' \end{pmatrix}\underline{y}_l + k\begin{pmatrix} \underline{\omega}_{\alpha l} \\ \underline{0} \end{pmatrix}\omega^{\alpha l}\right]\right\}, \tag{3.26}$$

where $\hat{Y}_\alpha = X\hat{\Pi}_\alpha$, $\hat{Y}_l = X\hat{\Pi}_l$, $\omega^{\alpha l}$ is the (α, l)th element of Ω_g^{-1}, $\Omega_{\alpha l}$ is a submatrix of Ω_g that is equal to the sampling covariance between corresponding rows of $V_\alpha = Y_\alpha - X\Pi_\alpha$ and $V_l = Y_l - X\Pi_l$, and $\underline{\omega}_{\alpha l}$ is a vector of sampling covariances between elements of $\underline{v}_\alpha = \underline{y}_\alpha - X\underline{\pi}_\alpha$ and corresponding rows of $V_l = Y - X\Pi_l$.

From (3.22), the marginal posterior pdf for the distinct elements of Ω_g is well-known to be in the form of an inverted Wishart pdf – see e.g. Zellner (1971a, p. 227), and this result will be employed in future work to evaluate the following posterior expectations $E\omega^{\alpha l}$, $E\Omega_{\alpha l}$ and $E\underline{\omega}_{\alpha l}\omega^{\alpha l}$ that are needed to provide a MELO estimate of $\underline{\delta}$ that is not conditional upon Ω_g being given.[16]

4. Properties of MELO estimates and estimators

In this section, selected properties of MELO estimates and estimators presented above will be established. We first consider properties of MELO estimates and then go on to consider large and finite sample properties of the MELO estimators for the estimation problems considered in previous sections.[17]

4.1. Properties of MELO estimates

For the MELO estimates presented in section 2, it is direct to establish that as the sample size grows large, each MELO estimate approaches its corresponding maximum likelihood (ML) estimate. Also, this property holds for the MELO structural coefficient estimates in section 3. Thus, when non-dogmatic prior distributions are employed, as the sample size grows, the MELO estimates approach corresponding ML estimates. In this connection, it is well-known that when the sample size grows, in general posterior distributions derived from non-dogmatic prior distributions and well-behaved

[16]Of course sample estimates of $\Omega_{\alpha l}, \omega^{\alpha l}$, and $\underline{\omega}_{\alpha l}$ can be inserted in (3.26) to give an approximation to the unconditional MELO estimate of $\underline{\delta}$.

[17]See Hill (1975) for a thoughtful discussion of the relevance of sampling properties of Bayesian estimators. In Zellner–Park (1975) further analysis of the finite sample properties of MELO estimators is presented.

likelihood functions, approach a normal distribution's form centered at the ML estimate.[18]

In very small sample situations, it is the case that posterior distributions for the problems considered in previous sections can have more than one pronounced mode.[19] Examination of posterior distributions' shapes will indicate to an investigator whether more than one important mode is present. In such cases, it is necessary to be extremely careful and thoughtful in choosing a point estimate. In the appendix to this paper, a problem of point estimation when a posterior distribution has more than one pronounced model is analyzed.

In connection with estimation of $\theta = 1/\mu$, the reciprocal of a population mean, it is apparent from (2.5) and (2.6) that the MELO estimate, $\hat{\theta}^*$, satisfies $|\hat{\theta}^*| < |\tilde{\theta}|$ where $\tilde{\theta} = 1/\bar{y}$, the ML estimate. For (2.5) and (2.6), $\hat{\theta}^*/\tilde{\theta} = 1/(1+Z^2)$ where Z^2 is the square of the coefficient of variation of the posterior pdf for μ. Thus the extent to which $\tilde{\theta}^*$ differs from $\tilde{\theta}$ depends just on the relative precision of the posterior pdf for μ. Similar observations apply to (2.8), the MELO estimate for a reciprocal of a regression coefficient.

With respect to estimation of ratios of parameters, the relation of the MELO estimates in (2.12) and (2.13) to ML estimates is readily apparent. E.g. for (2.12), we have $\hat{\theta}^* = \tilde{\theta}_{\mathrm{ML}}[1+\mathrm{cov}(\beta_1,\beta_2)/\hat{\beta}_1\hat{\beta}_2]/[1+\mathrm{var}\,\beta_2/\hat{\beta}_2^2]$, where $\tilde{\theta}_{\mathrm{ML}} = \hat{\beta}_1/\hat{\beta}_2$, the ML estimate, $\mathrm{cov}(\beta_1,\beta_2) = m^{12}\bar{s}^2$ is the posterior variance of β_1 and β_2, and $\mathrm{var}\,\beta_2 = m^{22}\bar{s}^2$ is the posterior variance of β_2. Then, with $\rho_{12} = \mathrm{cov}(\beta_1,\beta_2)/(\mathrm{var}\,\beta_1\ \mathrm{var}\,\beta_2)^{1/2}$, $\hat{\theta}^*/\tilde{\theta}_{\mathrm{ML}} = (1+\rho_{12}\phi_1\phi_2)/(1-\phi_2^2)$, where ϕ_1 and ϕ_2 are the posterior coefficients of variation for β_1 and β_2, respectively. Thus, depending on the values of ρ_{12}, ϕ_1 and ϕ_2, the ratio $\hat{\theta}^*/\tilde{\theta}_{\mathrm{ML}}$ can assume the values shown below:

	Condition on ρ_{12}, ϕ_1 and ϕ_2	Range of ratio $\hat{\theta}^*/\tilde{\theta}_{\mathrm{ML}}$
(a)	$\rho_{12} = 0, 0 < \phi_2^2, \phi_1^2 < \infty$	$0 < \hat{\theta}^*/\tilde{\theta}_{\mathrm{ML}} < 1$
(b)	$\phi_2^2 \leqq \rho_{12}\phi_1\phi_2 < \infty$	$1 \leqq \hat{\theta}^*/\tilde{\theta}_{\mathrm{ML}} < \infty$
(c)	$-1 < \rho_{12}\phi_1\phi_2 < \phi_2^2$	$0 < \hat{\theta}^*/\tilde{\theta}_{\mathrm{ML}} < 1$
(d)	$-\infty < \rho_{12}\phi_1\phi_2 \leqq -1$	$-\infty < \hat{\theta}^*/\tilde{\theta}_{\mathrm{ML}} \leqq 0$

When $\rho_{12} = 0$, line (a) of the table indicates $0 < \hat{\theta}^*/\tilde{\theta}_{\mathrm{ML}} < 1$. From line (b) of the table, it is seen that $\hat{\theta}^*$ is 'expanded' relative to $\tilde{\theta}_{\mathrm{ML}}$ when $\phi_2^2 < \rho_{12}\phi_1\phi_2 < \infty$. This latter condition can be satisfied when $\hat{\beta}_1$ and $\hat{\beta}_2$ have the same algebraic sign and $\rho_{12} > 0$ or $\hat{\beta}_1$ and $\hat{\beta}_2$ have opposite algebraic signs and $\rho_{12} < 0$. In either case, we have $|\phi_2| < |\rho_{12}||\phi_1|$, that is the relative variability

[18]See e.g. Jeffreys (1967), Le Cam (1958), Lindley (1961), Zellner (1971a).

[19]This is also true for the sampling distributions of ML estimators, e.g. $\hat{\theta}_{\mathrm{ML}} = 1/\bar{y}$, where $\bar{y}$ is a sample mean, or $\hat{\theta}_{\mathrm{ML}} = \hat{\beta}_1/\hat{\beta}_2$, where $\hat{\beta}_1$ and $\hat{\beta}_2$ are ML regression coefficient estimators.

of the posterior pdf for β_2 is less than that for β_1. Thus, heuristically, adjusting $\hat{\beta}_1$ *upward* relative to $\hat{\beta}_2$ guards against an underestimate of the ratio β_1/β_2. Under the condition of line (c), $0 < \hat{\theta}^*/\tilde{\theta}_{\mathrm{ML}} < 1$, there is a 'shrinking' of the ML estimate. Last, in line (d), with $-\infty < \rho_{12}\phi_1\phi_2 < -1$, $\tilde{\theta}^*$ and $\tilde{\theta}_{\mathrm{ML}}$ can have different algebraic signs. In this case, if ϕ_1 and ϕ_2 have the same algebraic sign, then $\rho_{12} < 0$, and the coefficients of variation ϕ_1 and ϕ_2 are rather large. Thus given the negative posterior correlation of β_1 and β_2, there is a high probability of obtaining the 'wrong' algebraic sign in estimating β_1/β_2 by use of the ML estimate and θ^* has an algebraic sign different from $\hat{\theta}_{\mathrm{ML}}$. However, this will just occur for large relative variability of one or both of the marginal posterior pdfs for β_1 and β_2.

For the MELO structural coefficient estimates in (3.6), (3.13), (3.21) and (3.26), as the sample size grows these estimates approach ML estimates since they are functions of posterior means that approach ML estimates under general conditions.

In small samples, (3.14) indicates that a particular MELO estimate of the coefficient vector $\underline{\gamma}_1$ can be expressed as a matrix weighted average of the 2SLS estimate and $\hat{\underline{\gamma}}_1 = (\hat{V}_1'\hat{V}_1)^{-1}\hat{V}_1'\hat{\underline{v}}_1$, or of the 2SLS and DLS estimates [Zellner (1976a)]. Properties of matrix weighted averages of two vectors can be studied conveniently using the curve décolletage [Dickey (1975) and Leamer (1973)]. The small sample problem of multi-modal posterior distributions and point estimation for structural coefficients will not be treated herein but deserves attention.

4.2. *Large sample properties of MELO estimators*

When the MELO estimates presented above are viewed as estimators, it is the case that for each problem considered the MELO estimator has the same asymptotic distribution as the ML estimator.

The MELO estimator for the reciprocal of a population mean, $\theta = 1/\mu$, is given in (2.4). From (2.4), we can write

$$n^{1/2}(\hat{\theta}^* - \theta) = n^{1/2}(1/\bar{\mu} - \theta) - (n^{1/2}/\bar{\mu})(\operatorname{var}\mu/\bar{\mu}^2)(1 + \operatorname{var}\mu/\bar{\mu}^2), \quad (4.1)$$

where $\hat{\theta}^*$ is the MELO estimator (2.4), n is the sample size, $\bar{\mu}$ is the mean and $\operatorname{var}\mu$ is the variance of the posterior pdf for μ. The probability limit (plim) of the second term on the r.h.s. of (4.1) is zero, given that $\operatorname{plim}\bar{\mu} \neq 0$. Thus $n^{1/2}(\hat{\theta}^* - \theta)$ and $n^{1/2}(1/\bar{\mu} - \theta)$ have the same asymptotic distribution by use of the convergence result in Cramer (1946, p. 254). Further, with $\bar{\mu} = \bar{y} + \Delta$, where $\bar{y}$ is the sample mean and

$$\Delta = \bar{\mu} - \bar{y}, \qquad 1/\bar{\mu} = 1/\bar{y} - \Delta/\bar{y}(1 + \Delta/\bar{y}),$$

and

$$n^{1/2}(1/\bar{\mu}-\theta) = n^{1/2}(1/\bar{y}-\theta)-n^{1/2}\Delta/\bar{y}(1+\Delta/\bar{y}).$$

Given that $\operatorname{plim}\bar{y} \neq 0$ and $\operatorname{plim} n^{1/2}\Delta = 0$, $n^{1/2}(1/\bar{\mu}-\theta)$, $n^{1/2}(1/\bar{y}-\theta)$, and $n^{1/2}(\hat{\theta}^*-\theta)$ all have the same asymptotic normal distribution under usual central limit theorem conditions – see e.g. Dhrymes (1974) and Theil (1971).

For the MELO reciprocal mean estimator in (2.6), we have

$$n^{1/2}(\hat{\theta}^*-\theta) = n^{1/2}(1/\bar{y}-\theta)-(n^{1/2}/\bar{y})(\bar{s}^2/n\bar{y}^2)/(1+\bar{s}^2/n\bar{y}^2),$$

$$\bar{s}^2 = vs^2/(v-2), \qquad vs^2 = \sum_{i=1}^{n}(y_i-\bar{y})^2,$$

and

$$v = n-1 > 2.$$

Then given $\operatorname{plim}\bar{y} \neq 0$ and applying the convergence result in Cramer (1946, p. 254), $n^{1/2}(\hat{\theta}^*-\theta)$ and $n^{1/2}(1/\bar{y}-\theta)$ have the same asymptotic distribution that will be normal under usual conditions of central limit theorems. Since proofs that the asymptotic distributions of the MELO estimators in (2.7) and (2.8) and of the ML estimator $1/\hat{\pi}$ are the same follow along similar lines as those presented above, they will not be presented.

With respect to the MELO estimator for the ratio of two parameters, e.g. (2.11) that is the estimator for the ratio of two regression coefficients, we have

$$n^{1/2}(\hat{\theta}^*-\theta) = n^{1/2}(\bar{\beta}_1/\bar{\beta}_2-\theta)+(n^{1/2}\bar{\beta}_1/\bar{\beta}_2)$$

$$\times\left(\frac{\operatorname{cov}(\beta_1,\beta_2)/\bar{\beta}_1\bar{\beta}_2-\operatorname{var}\beta_2/\bar{\beta}_2^2}{1+\operatorname{var}\beta_2/\bar{\beta}_2^2}\right), \tag{4.2}$$

where $\hat{\theta}^*$ is shown in (2.11), $\theta = \beta_1/\beta_2$, $\bar{\beta}_1$ and $\bar{\beta}_2$ are posterior means, and $\operatorname{cov}(\beta_1,\beta_2)$ and $\operatorname{var}\beta_2$ are the posterior covariance of β_1 and β_2 and the posterior variance of β_2, respectively. Given that $\operatorname{plim}\bar{\beta}_2 \neq 0$ and that $\operatorname{plim} n^{1/2}\operatorname{cov}(\beta_1,\beta_2) = \operatorname{plim} n^{1/2}\operatorname{var}\beta_2 = 0$, $n^{1/2}(\hat{\theta}^*-\theta)$ and $n^{1/2}(\bar{\beta}_1/\bar{\beta}_2-\theta)$ have the same asymptotic distribution that is the same as the asymptotic normal distribution for $n^{1/2}(\hat{\beta}_1/\hat{\beta}_2-\theta)$, where $\hat{\beta}_1/\hat{\beta}_2$ is the ML estimator for θ.[20]

[20]The proof that $n^{1/2}(\hat{\theta}^*-\theta)$, $n^{1/2}(\bar{\beta}_1/\bar{\beta}_2-\theta)$ and $n^{1/2}(\hat{\beta}_1/\hat{\beta}_2-\theta)$ all have the same asymptotic normal distribution, given $\operatorname{plim}\bar{\beta}_2 = \operatorname{plim}\hat{\beta}_2 \neq 0$, is direct and thus is not presented.

As regards the structural coefficient estimator $\hat{\underline{\delta}}_1^*$ in (3.6), we have

$$\begin{aligned}\hat{\underline{\delta}}_1^* &= (\hat{Z}_1'\hat{Z}_1 + \Delta_1)^{-1}(\hat{Z}_1'X\hat{\underline{\pi}}_1 + \underline{\Delta}_2)\\ &= [I + (\hat{Z}_1'\hat{Z}_1)^{-1}\Delta_1]^{-1}[(\hat{Z}_1'\hat{Z}_1)^{-1}\hat{Z}_1'X\hat{\underline{\pi}}_1 + (\hat{Z}_1'\hat{Z}_1)^{-1}\underline{\Delta}_2]\\ &= \hat{\underline{\delta}}_1 - [I + (\hat{Z}_1'\hat{Z}_1)^{-1}\Delta_1]^{-1}(\hat{Z}_1'\hat{Z}_1)^{-1}[\Delta_1\hat{\underline{\delta}}_1 - \underline{\Delta}_2],\end{aligned} \tag{4.3}$$

where $\hat{\underline{\delta}}_1 = (\hat{Z}_1'\hat{Z}_1)^{-1}\hat{Z}_1'X\hat{\underline{\pi}}_1$ is the 2SLS estimator, $\hat{Z}_1 = (\hat{Y}_1 \vdots X_1)$ with $\hat{Y}_1 = X\hat{\Pi}_1$ where $\hat{\Pi}_1 = (X'X)^{-1}X'Y_1$, $\Delta_1 = E\bar{Z}_1'\bar{Z}_1 - \hat{Z}_1'\hat{Z}_1$, and $\underline{\Delta}_2 = E\bar{Z}_1'X\underline{\pi}_1 - \hat{Z}_1'X\hat{\underline{\pi}}_1$, with E the posterior expectation operator and $\bar{Z}_1 = (X\Pi_1 \vdots X_1)$. Then $n^{1/2}(\hat{\underline{\delta}}_1^* - \underline{\delta}_1) = n^{1/2}(\hat{\underline{\delta}}_1 - \underline{\delta}_1) + n^{1/2}\Delta_3$, where Δ_3 represents the second term on the r.h.s. of the third line of (4.3). Since plim $n^{1/2}\Delta_3 = 0$,[21] $n^{1/2}(\hat{\underline{\delta}}_1^* - \underline{\delta}_1)$ and $n^{1/2}(\hat{\underline{\delta}}_1 - \underline{\delta}_1)$ have the same asymptotic normal distribution.[22] Further since the 2SLS and limited-information ML estimators have the same asymptotic normal distribution, this is also the asymptotic distribution of $n^{1/2}(\hat{\underline{\delta}}_1^* - \underline{\delta}_1)$. With respect to the particular structural coefficient estimator in (3.13), note that it is an $\mathscr{R}$-class estimator with $\mathscr{R} = \mathscr{R}^* = 1 - k/(n-k-m+1)$. Thus $\operatorname{plim} n^{1/2}(\mathscr{R}^* - 1) = 0$ and under this condition it is known that $\mathscr{R}$-class estimators, including the limited information ML estimator, all have the same asymptotic normal distribution.[23]

4.3. *Some finite sample properties of MELO estimators*

In this section some finite sample properties of the MELO estimators, presented above, are established.[24] In particular, the existence of the moments of certain MELO estimators will be demonstrated. Also, for the reciprocal mean problem, an exact expression for the MELO estimator's average risk is derived. Finally, considerations bearing on the admissibility of certain MELO estimators are presented.

The MELO estimator (2.5) for $\theta = 1/\mu$, with the value of σ^2 assumed known, is $\hat{\theta}^* = \bar{y}/(\bar{y}^2 + \sigma_0^2/n) = (n^{1/2}/\sigma_0)Z/(1+Z^2)$, where $Z = n^{1/2}\bar{y}/\sigma_0$. Then it is easily established that $|\hat{\theta}^*| \leqq n^{1/2}/2\sigma_0$. Thus the rth absolute moment of $\hat{\theta}^*$ is

$$E|\hat{\theta}^*|^r = \int_{-\infty}^{\infty} |\hat{\theta}^+|^r p(\bar{y}|\mu, \sigma_0/n^{1/2})\,\mathrm{d}\bar{y} < (n^{1/2}(2\sigma_0)^r < \infty,$$

[21]Note that $\operatorname{plim} n^{1/2}(\hat{Z}_1'\hat{Z}_1)^{-1}\Delta_1\hat{\underline{\delta}}_1 = \operatorname{plim} n^{1/2}(\hat{Z}_1'\hat{Z}_1)^{-1}\underline{\Delta}_2 = 0$ assuming that $\operatorname{plim} \hat{Z}_1'\hat{Z}_1/n = M$, a pds matrix and since $\operatorname{plim} \Delta_1/n^{1/2} = 0$ and $\operatorname{plim} \underline{\Delta}_2/n^{1/2} = \underline{0}$.

[22]For a proof that $n^{1/2}(\hat{\underline{\delta}}_1 - \underline{\delta}_1)$ has an asymptotic normal distribution, see e.g. Dhrymes (1974) and Theil (1971).

[23]See, e.g., Theil (1971, pp. 505–506).

[24]Further properties including exact finite sample distributions, exact and approximate risk functions, etc., are presented in Zellner–Park (1975).

for $0 < \sigma_0/n^{1/2} < \infty$, where $p(\bar{y}|\mu, \sigma_0/n^{1/2})$ is the normal pdf for $\bar{y}$ with mean μ and standard deviation $\sigma_0/n^{1/2}$. Since the absolute moments of $\hat{\theta}^*$ exist and are finite, $E(\hat{\theta}^*)^r < \infty$, $r = 1, 2, \ldots$. Further, the risk of $\hat{\theta}^*$ relative to the loss function, $L = ((\theta - \hat{\theta}^*)/\theta)^2 = (1 - \mu\hat{\theta}^*)^2$ is finite for $0 < \sigma_0/n^{1/2} < \infty$ and $-\infty < \mu < \infty$.[25]

In estimating $\theta = 1/\mu$ when σ^2 has an unknown value, the MELO estimator, presented in (2.6), is $\hat{\theta}^* = \bar{y}/[\bar{y}^2 + vs^2/n(v-2)] = (1/\bar{c})\bar{z}/(1+\bar{z}^2)$, where $\bar{c}^2 = vs^2/n(v-2)$, $v = n-1 > 2$, and $\bar{z} = \bar{y}/\bar{c}$. Then $|\hat{\theta}^*| \leqq 1/2\bar{c}$ and

$$E|\hat{\theta}^*|^r = \int_0^{\infty} \int_{-\infty}^{\infty} |\hat{\theta}^*|^r p(\bar{y}|\mu, \sigma/n^{1/2}) g(x|v)\,\mathrm{d}\bar{y}\,\mathrm{d}x$$
$$< \left(\frac{v-2}{2\sigma^2/n}\right)^{r/2} \int_0^{\infty} x^{-r/2} g(x|v)\,\mathrm{d}x < \infty,$$

for $0 < \sigma/n^{1/2} < \infty$, where $x = vs^2/\sigma^2$ has a χ^2 pdf with $v = n-1$ degrees of freedom, denoted by $g(x|v)$. Thus $E|\hat{\theta}^*|^r$ and $E(\hat{\theta}^*)^r$, $r = 1, 2, \ldots$, exist and are finite. This implies that the risk of $\hat{\theta}^* = \bar{y}/[\bar{y}^2 + vs^2/n(v-2)]$, $v > 2$, relative to $L = (\theta - \hat{\theta}^*)^2/\theta^2 = (1 - \mu\hat{\theta}^*)^2$ exists and is finite for $0 < \sigma/n^{1/2} < \infty$ and $-\infty < \mu < \infty$.

To obtain the average risk (AR) of the estimator $\hat{\theta}^* = \bar{y}/(\bar{y}^2 + \sigma_0^2/n)$, the following integral was evaluated:

$$AR = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (1 - \mu\hat{\theta}^*)^2 p(\bar{y}|\mu, \sigma_0/n^{1/2})\,\mathrm{d}\bar{y}\,\mathrm{d}\mu = \pi\sigma_0/n^{1/2}, \tag{4.4}$$

where the improper prior pdf for μ, $p(\mu)\,\mathrm{d}\mu \propto \mathrm{d}\mu$, $-\infty < \mu < \infty$ has been employed, $p(\bar{y}|\mu, \sigma_0/n^{1/2})$ is the normal pdf for $\bar{y}$ with mean μ and variance σ_0^2/n, and $\pi = 3.14159$.[26]

[25]The risk of $\hat{\theta}^*$ relative to the quadratic loss function, $L = (\hat{\theta}^* - \theta)^2$ is finite for $0 < |\mu| < \infty$ and $0 < \sigma_0/n^{1/2} < \infty$.

[26]To evaluate (4.4), write $\hat{\theta}^* = \bar{y}/(\bar{y}^2 + \sigma_0^2/n) = (n^{1/2}/\sigma_0)(w + \bar{\mu})/[1 + (w\bar{\mu})^2]$, where $w = n^{1/2}(\bar{y} - \mu)/\sigma_0$ is $N(0,1)$ and $\bar{\mu} = n^{1/2}\mu/\sigma_0$. Then $L = (1 - \mu\hat{\theta}^*)^2 = \{[1 + w(w + \bar{\mu})]/[1 + (w + \bar{\mu})^2]\}^2$ and (4.4) becomes

$$AR = \sigma_0/(2\pi n)^{1/2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \{[1 + w(w + \bar{\mu})]/[1 + (w + \bar{\mu})^2]\}^2\, \mathrm{d}^{-w^2/2}\mathrm{d}w\,\mathrm{d}\bar{\mu},$$

where $\mathrm{d}\mu = (\sigma_0/n^{1/2})\mathrm{d}\bar{\mu}$ has been employed. Now make the following change of variables, $x = w + \mu$ and $w = w$, that has Jacobian equal to one, to obtain

$$AR = \sigma_0/(2\pi n)^{1/2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [(1 + wx)/(1 + x^2)]\,\mathrm{e}^{-w^2/2}\mathrm{d}w\,\mathrm{d}x$$
$$= \sigma_0/n^{1/2} \int_{-\infty}^{\infty} [1/(1 + x^2)]\,\mathrm{d}x.$$

In the class of estimators for $\theta = 1/\mu$, $\hat{\theta} = \hat{\theta}(\bar{y}, \sigma/n^{1/2})$, with the value of $\sigma/n^{1/2}$ known, such that AR is finite relative to the loss function $L = (\theta-\hat{\theta})^2/\theta^2 = (1-\mu\hat{\theta})^2$ and improper prior pdf $p(\mu)\,d\mu \propto d\mu$, $-\infty < \mu < \infty$, $\hat{\theta}^* = \bar{y}/(\bar{y}^2 + \sigma_0^2/n)$ is the estimator in this class that minimizes AR and is thus admissible. That is, there cannot be another estimator in this class that has risk below that for $\hat{\theta}^*$ over a portion (or all) of the parameter space and risk not above that for $\hat{\theta}^*$ over the remaining portion of the parameter space. If there were such an estimator, it would have smaller AR than that for $\hat{\theta}^*$ which is impossible since $\hat{\theta}^*$ is the estimator in the class that minimizes AR.

The MELO estimator for $\theta = 1/\mu$ when σ^2 has an unknown value is from (2.6), $\hat{\theta}^* = \bar{y}/(\bar{y}^2 + \bar{s}^2/n)$, where $\bar{s}^2 = vs^2/(v-2)$, with $v = n-1 > 2$ and $vs^2 = \sum_{i=1}^n (y_i - \bar{y})^2$. By a direct calculation, the average risk (AR) of $\hat{\theta}^*$ relative to the loss function in (2.2) is found to be[27]

$$AR = (\pi/2n^{1/2})\left[\left(\frac{2}{v-2}\right)^{1/2}\Gamma\left(\frac{v+1}{2}\right) + \left(\frac{v-2}{2}\right)^{1/2}\Gamma\left(\frac{v-2}{2}\right)\right]\Big/\Gamma(v/2), \qquad v = n-1 > 2, \tag{4.5}$$

which is finite. Thus the MELO estimator $\hat{\theta}^* = \bar{y}/(\bar{y}^2 + \bar{s}^2/n)$ is admissible.

In this last expression, note that $1+x^2 = (x-i)(x+i)$, where $i = (-1)^{1/2}$, and thus there is a pole of order one in the upper half plane at $x = i$. Using a contour integration around a half circle in the upper half plane of the complex domain with center at the origin, the value of the integral

$$\int_{-\infty}^{\infty} [1/(1+x^2)]\,dx$$

is equal to $2\pi i$ times the residue at i, namely $\frac{1}{2}i$, that is $2\pi i/2i = \pi$ by the Cauchy Residue Theorem. Or more simply,

$$\int_{-\infty}^{\infty} (1+x^2)^{-1}\,dx = \pi,$$

using properties of the univariate Cauchy pdf. Thus $AR = \pi\sigma_0/n^{1/2}$ as stated in (4.4).

[27]To derive (4.5), write

$$\hat{\theta}^* = \bar{y}(\bar{y}^2 + \bar{s}^2)/n) = (n^{1/2}/\sigma)(w+\bar{\mu})^2 + z/(v-2)],$$

where $\bar{\mu} = n^{1/2}\mu/\sigma$, $w = (\bar{y}-\bar{\mu})/(\sigma/n^{1/2})$, and $z = (v-2)\bar{s}^2/\sigma^2 = vs^2/\sigma^2$. Note that w and z are independently distributed with w having a $N(0,1)$ pdf and z a χ^2 pdf with $v = n-1$ degrees of freedom. The loss function can be expressed as follows:

$$L = \left(\frac{\hat{\theta}^* - \theta}{\theta}\right)^2 = (\hat{\theta}^*\mu - 1)^2 = \left[\frac{z/(v-2) + w(w+\bar{\mu})}{z/(v-2) + (2+\bar{\mu})^2}\right]^2.$$

Some finite sample properties of MELO estimators for ratios of parameters are presented in Zellner–Park (1975. In the same paper attention is given to the properties of the MELO structural coefficient estimator $\underline{\hat{\delta}}_1^*$, shown in (3.13). Using Kadane's (1971) asymptotic small-σ approximations to the moments of $\mathscr{R}$-class estimators, it is found that the approximate risk of the MELO estimator $\underline{\hat{\delta}}_1^*$ relative to the loss function in (3.4) is smaller than that of the 2SLS estimator when the latter's moments exist under a range of conditions frequently encountered in practice.

5. Summary and conclusions

In this paper a minimum expected loss (MELO) approach was utilized to generate point estimates for a number of related estimation problems including reciprocals and ratios of parameters and structural coefficients of linear structural econometric models. For these problems, well-defined and reasonable loss functions were formulated and estimates that minimize posterior expected loss were derived. These MELO estimates have relatively simple forms and thus are quite operational. In the case of 'single-equation' estimation of structural coefficients, it was found that with a normal reduced form system and a diffuse prior pdf for its parameters, the MELO estimate for the structural coefficient vector is an $\mathscr{R}$-class estimate with a value of $\mathscr{R}$ less than one in finite samples. Further, a 'systems' MELO estimate for structural coefficients was derived. These MELO point estimates are useful in

---

From $p(\mu,\sigma)\propto 1/\sigma$, $p(\bar{\mu},\sigma)\propto 1/n^{1/2}$ and thus average risk is given by

$$AR=(1/(2\pi n)^{1/2})\int_0^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} L\,\mathrm{e}^{-w^2/2}p(z)\,\mathrm{d}w\,\mathrm{d}\bar{\mu}\,\mathrm{d}z,$$

with $p(z)$ the χ_ν^2 pdf. In this last expression, let $x=w+\bar{\mu}$, $w=w$ and $z=z$, a change of variables with Jacobian equal to one, to obtain

$$\begin{aligned}AR&=(1/(2\pi n)^{1/2})\int_0^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{[z/(\nu-2)+wx]^2}{[x/(\nu-2)+x^2]^2}\mathrm{e}^{-w^2/2}p(z)\,\mathrm{d}w\,\mathrm{d}x\,\mathrm{d}z\\&=(1/n^{1/2})\int_0^{\infty}\left[\int_{-\infty}^{\infty}\frac{[z/(\nu-2)]^2+x^2}{[z/(\nu-2)+x^2]^2}\mathrm{d}x\right]p(z)\,\mathrm{d}z.\end{aligned}$$

The integral in square brackets can be evaluated using properties of the univariate Student-t pdf with three degrees of freedom to yield

$$AR=(\pi/2)\int_0^{\infty}(z/(\nu-2))^{1/2}+((\nu-2)/z)^{1/2}p(z)\,\mathrm{d}z.$$

This last integral can be evaluated given that $p(z)$ is a χ^2 pdf with ν d.f. to yield the expression shown in (4.5).

providing information about parameters' values and can supplement other measures describing properties of posterior distributions.

When the MELO point estimates are regarded as point estimators, it was pointed out that these estimators' moments and risk exist and are finite. These properties contrast markedly with those of some other estimators for these problems that do not possess finite moments and have infinite risk relative to quadratic and other loss functions. For example, maximum likelihood estimators for reciprocals and ratios of means and of regression coefficients do not possess finite moments under a wide range of distributional assumptions relating to data processes. In the case of estimating structural coefficients, it is well-known that many commonly used estimators fail to possess moments under frequently encountered conditions. That ML and some other estimators for the problems considered in this paper can or do have infinite risk relative to quadratic and other loss functions is of interest particularly in view of the widespread use of quadratic loss functions and risk considerations in choosing point estimators, for example in the Gauss–Markov Theorem, in Charles Stein's work, in Monte Carlo experiments, etc. In situations in which ML and other estimators have infinite risk whereas MELO estimators have finite risk, the former estimators are clearly inadmissible. MELO estimators based on proper prior distributions are admissible relative to the loss functions for which they have been derived. Some MELO estimators based on improper priors are admissible while others may be only under special conditions [Hill (1975)]. In large samples, the sampling properties of the ML, MELO and other consistent estimators for particular problems considered in this paper are identical.

Further, it is recognized that (1) not all agree that admissibility is an overriding or even important criterion in choosing point estimates; (2) ML and perhaps other estimators might perform better relative to performance criteria that are not sensitive to the existence or non-existence of sampling moments; and (3) while point estimates are valuable, it must be recognized that they alone are not adequate solutions to most estimating problems. Thus it is fortunate that posterior distributions are available for the problems considered in this paper that reflect the prior and sample information much better than do point estimates alone.

Appendix

A.1. Bimodality of the posterior pdf for $\theta = 1/\mu$[28]

To illustrate a case in which a posterior pdf is bimodal, consider $y_i = \mu + \varepsilon_i$ where the ε_i's are NID$(0, \sigma^2)$ with the value of σ^2 assumed known. With a

[28] Analysis of the bimodality of the sampling distribution of the maximum likelihood estimator for θ, $\hat{\theta} = 1/\bar{y}$, with $\bar{y} \sim N(\mu, \sigma^2/n)$ is similar to that presented below.

diffuse prior pdf for μ, $p(\mu) \propto$ const., $-\infty < \mu < \infty$, the posterior pdf for μ is $p(\mu|\sigma, \bar{y}) \propto \exp\{-n(\mu-\bar{y})^2/2\sigma^2\}$, a normal pdf. Making a change of variable $\theta = 1/\mu$, the posterior pdf for θ is

$$p(\theta|\sigma, \bar{y}) \propto \theta^{-2} \exp\left\{-\frac{z^2}{2}\left(\frac{\theta-\hat{\theta}}{\theta}\right)^2\right\}, \tag{A.1}$$

where $\hat{\theta} = 1/\bar{y}$ and $z = n^{1/2}\bar{y}/\sigma$. Note that z is the ratio of the posterior mean for μ, $\bar{y}$, to its posterior standard deviation, $\sigma/n^{1/2}$. Also z is a sampling theory test statistic for testing the hypothesis $\mu = 0$.

For convenience in analyzing the modes of (A.1), let $\eta = (\hat{\theta}-\theta)/\theta$. Then

$$\theta^{-2} \exp\left\{-\frac{z^2}{2}\left(\frac{\theta-\hat{\theta}}{\theta}\right)^2\right\} \propto (1+\eta)^2 \exp\{-z^2\eta^2/2\}. \tag{A.2}$$

From the form of (A.2), when z^2 has a small value, it is clear from a plot of the two factors on the r.h.s. of (A.2) that there will be pronounced bimodality. On taking the logarithm of the expression on the r.h.s. of (A.2) and differentiating it, the two modal values of η, η_1 and η_2, can be determined. They are $\eta_1, \eta_2 = (1/2)(-1 \pm (1+8/z^2)^{1/2})$. From $\eta = (\hat{\theta}-\theta)/\theta$, $\theta = \hat{\theta}/(1+\eta)$ and thus the modal values for θ are

$$\theta_1 = \hat{\theta}\left(\frac{2}{1+(1+8/z^2)^{1/2}}\right) \quad \text{and} \quad \theta_2 = \hat{\theta}\left(\frac{2}{1-(1+8/z^2)^{1/2}}\right). \tag{A.3}$$

It is clear that $\theta_1 > 0$ and $\theta_2 < 0$. Also, the modal values of the posterior pdf can be quite different from the ML estimate $\hat{\theta}$, particularly when $z = n^{1/2}\bar{y}/\sigma$ is small in absolute value. For example, when $z = 1$, $\theta_1 = \hat{\theta}/2$ and $\theta_2 = -\hat{\theta}$. The MELO estimate for θ is $\hat{\theta}^* = (1/\bar{y})(1+\sigma^2/n\bar{y}^2)^{-1} = \hat{\theta}z^2/(1+z^2)$. Thus when $z^2 = 1$, $\hat{\theta}^* = \hat{\theta}/2 = \theta_1$, a modal value. As z^2 gets large, θ_1 and $\hat{\theta}^*$ both approach $\hat{\theta}$ while θ_2 approaches $-\infty$.

On inserting the modal values for η in the expression on the r.h.s. of (A.2), denoted by $f(\eta)$, we have

$$f(\eta_1)/f(\eta_2) = \left(\frac{1+(1+8/z^2)^{1/2}}{1-(1+8/z^2)^{1/2}}\right)^2 \exp\left\{\frac{z^2}{2}(1+8/z^2)^{1/2}\right\}. \tag{A.4}$$

When $z^2 = 1$, $f(\eta_1)/f(\eta_2) = 4e^{3/2} = 17.93$, the ratio of the ordinates of the posterior pdf at the modal values, η_1 and η_2 or θ_1 and θ_2. Thus even for $z^2 = 1$, one mode is much higher than the other. When z^2 grows in value, the relative heights of the modes grows. As $z^2 \to 0$, $f(\eta_1)/f(\eta_2) \to 1$.

A.2. Point estimation of θ with bimodal distribution

As shown above, (A.1) is bimodal with one modal value positive and the other negative. Further, as $|\theta| \to 0$, $p(\theta|\sigma, \bar{y}) \to 0$ and the posterior probability that $\theta = 1/\mu > 0$ is

$$\int_0^{\infty} p(\mu|\sigma, \bar{y})\,\mathrm{d}\mu.$$

Conditional upon $\mu > 0$ or $\theta > 0$, the MELO estimate of θ relative to $(\theta - \hat{\theta})^2/\theta^2$ is given by $\hat{\theta}^*_+ = E(\mu|\sigma, \mu > 0)/E(\mu^2|\sigma, \mu > 0)$ where E is the posterior expectation operator. An explicit expression for $\hat{\theta}^*_+$ is[29]

$$\hat{\theta}^*_+ = \frac{\bar{y}[1 - F(-z)] + (\sigma/n^{1/2})P(z)}{\bar{y}^2[1 - F(-z)] + 2\bar{y}(\sigma/n^{1/2})P(z) + \sigma^2/n[1 - F(-z) - zP(z)}$$

$$= \hat{\theta}^* \frac{(1+z^2)(1+Rz)}{(1+z^2)(1+Rz) - 1} = \hat{\theta}\frac{z(1+Rz)}{z(1+Rz)+R}, \tag{A.5}$$

where $z = n^{1/2}\bar{y}/\sigma$, $P(z) = (1/(2\pi)^{1/2})\exp\{-z^2/2\}$, $F(-z) = \int_{-\infty}^{-z} P(z)\,\mathrm{d}z$, $\hat{\theta}^* = (1/\bar{y})/(1 + \sigma^2/n\bar{y}^2)$, $\hat{\theta} = 1/\bar{y}$, and $R = [1 - F(-z)]/P(z)$ Mills' ratio. From (A.5), as $z \to \infty$, $\hat{\theta}^*_+/\hat{\theta}^* \to \hat{\theta}^*_+/\hat{\theta} \to 1$. This is reasonable since as $z \to \infty$, the mode situated over positive values of θ becomes very sharp and centered at the ML estimate $\hat{\theta} = 1/\bar{y}$. On the other hand, for $z = n^{1/2}\bar{y}/\sigma = 1$, $\hat{\theta}^*_+ = 1.4326\hat{\theta}^* = 0.7163\hat{\theta}$ from (A.5) and $\hat{\theta}^* = 0.5\hat{\theta}$. Thus with $z = 1$, introducing the conditional information that $\mu > 0$ or $\theta > 0$, results in the MELO estimate $\hat{\theta}^*_+$ being expanded relative to $\hat{\theta}^*$, the MELO estimate not incorporating the information $\mu > 0$. However, with $z = 1$, $\hat{\theta}^*_+$ is still smaller in value than the ML estimate $\hat{\theta}$. Similar calculations can be performed in the case that the conditioning information is $\mu < 0$. In this way MELO point estimates, $\hat{\theta}^*_+$ and $\hat{\theta}^*_-$ can be obtained. With a 2×2 loss structure and posterior probabilities for the hypotheses, $\mu > 0$ and $\mu < 0$, one can choose between $\hat{\theta}^*_+$ and $\hat{\theta}^*_-$ on the basis of which act has lower expected loss. In case

[29]The expression for $\hat{\theta}^*_+$ in (A.5) was obtained by straightforward evaluation of the following integrals:

$$\hat{\theta}^*_+ = \int_0^{\infty} \mu p(\mu|\sigma, \bar{y})\,\mathrm{d}\mu \bigg/ \int_0^{\infty} \mu^2 p(\mu|\sigma, \bar{y})\,\mathrm{d}\mu,$$

with

$$p(\mu|\sigma, \bar{y}) = \frac{1}{(2\pi)^{1/2}(\sigma/n^{1/2})}\exp\left\{-\frac{n}{2\sigma^2}(\mu - \bar{y})^2\right\}.$$

no loss structure is available, one has posterior probabilities that $\theta > 0$ and $\theta < 0$ as well as the point estimates $\hat{\theta}^*_+$ and $\hat{\theta}^*_-$.

References

Anderson, T.W. and T. Sawa, 1973, Distributions of estimates of coefficients of a single equation in a simultaneous system and their asymptotic expansions, Econometrica 41, 683–714.

Bergstrom, A.R., 1962, The exact sampling distributions of least squares and maximum likelihood estimators of the marginal propensity to consume, Econometrica 30, 480–490.

Box, G.E.P. and G.C. Tiao, 1973, Bayesian inference in statistical analysis (Addison-Wesley, Reading, MA).

Cramer, H., 1946, Mathematical methods of statistics (Princeton University Press, Princeton, NJ).

DeGroot, M.H., 1970, Optimal statistical decisions (McGraw-Hill, New York).

Dhrymes, P.J., 1974, Econometrics: Statistical foundations and applications (Springer-Verlag, New York).

Dickey, J.M., 1975, Bayesian alternatives to the F test and least-squares estimate in the normal linear model, in: S.E. Fienberg and A. Zellner, eds., Studies in Bayesian econometrics and statistics in honor of Leonard J. Savage (North-Holland, Amsterdam), 515–554.

Ferguson, T.S., 1967, Mathematical statistics: A decision theoretical approach (Academic Press, New York).

Goldberger, A.S., 1964, Econometric theory (Wiley, New York).

Hatanaka, M., 1973, On the existence and the approximation formulae for the moments of the $\mathscr{K}$-class estimators, Economic Studies Quarterly 24, 1–15.

Hill, B.M., 1975, On coherence, inadmissibility and inference about many parameters in the theory of least squares, in: S.E. Fienberg and A. Zellner, eds., Studies in Bayesian econometrics and statistics in honor of Leonard J. Savage (North-Holland, Amsterdam), 555–584.

Jeffreys, H., 1967, Theory of probability, 3rd rev. ed. (Clarendon Press, Oxford).

Kadane, J.B., 1971, Comparison of $\mathscr{K}$-class estimators when the disturbances are small, Econometrica 39, 723–738.

Kappenman, R.F. and S. Geisser, 1970, Bayesian and fiducial solutions for the Fieller–Creasy problem, Sankhya 32, 331–340.

Leamer, E.E., 1973, Multicollinearity: A Bayesian interpretation, Review of Economics and Statistics 55, 371–380.

Le Cam, L., 1958, Les propriétés asymptotiques des solutions de Bayes, Publ. Inst. Statist. (University of Paris) 7, 17–35.

Lindley, D.V., 1961, The use of prior probability distributions in statistical inference and decisions, in: J. Neyman, ed., Proc. Fourth Berkeley Symp. Math. Statist. and Probab., Vol. I, 453–468.

Madansky, A., 1964, On the efficiency of three-stage least-squares estimation, Econometrica 32, 51–56.

Mariano, R.S., 1973, Approximations to the distribution functions of Theil's $\mathscr{K}$-class estimators, Econometrica 41, 715–722.

Nagar, A.L., 1959, The bias and moment matrix of the general $\mathscr{K}$-class estimators of the parameters in simultaneous equations, Econometrica 29, 575–595.

Press, S.J., 1969, The t-ratio distribution, Journal of the American Statistical Association 64, 242–252.

Sargan, J.D., 1964, Three-stage least-squares and full maximum likelihood estimates, Econometrica 32, 77–81.

Sawa, T., 1972, Finite-sample properties of the $\mathscr{K}$-class estimators, Econometrica 40, 653–680.

Theil, H., 1971, Principles of econometrics (Wiley, New York).

Zellner, A., 1971a, An introduction to Bayesian inference in econometrics (Wiley, New York).

Zellner, A., 1971b, Bayesian and non-Bayesian analysis of log-normal distribution and log-normal regression, Journal of the American Statistical Association 66, 327–330.

Zellner, A., 1976a, A note on the relationship of minimum expected loss (MELO) and other structural coefficient estimates, manuscript (H.G.B. Alexander Research Foundation, University of Chicago, Chicago, IL).

Zellner, A., 1976b, Reply, Econometrica 44, 619–624.

Zellner, A. and S.B. Park, 1975, Further properties of minimum expected loss (MELO) estimators for nonlinear functions of parameters including structural coefficients, Manuscript (H.G.B. Alexander Research Foundation, University of Chicago, Chicago, IL).

Zellner, A. and W. Vandaele, 1975, Bayes–Stein estimators for k-means, regression and simultaneous equation models, in: S.E. Fienberg and A. Zellner, eds., Studies in Bayesian econometrics and statistics in honor of Leonard J. Savage (North-Holland, Amsterdam) 627–654.