Institute of Mathematical Statistics

LECTURE NOTES — MONOGRAPH SERIES

ESTIMATING FUNCTIONS AND OVER-IDENTIFIED MODELS

Tony Wirjanto
University of Waterloo

ABSTRACT

Economic theory (particularly with optimizing economic agents) usually imposes a set of moment restrictions on economic data. These restrictions are known as orthogonality conditions, which correspond to a set of unbiased estimating functions with the dimension of the estimating functions often larger than the dimension of the parameters of interest. This paper provides a selected review on the efficient methods of estimating such over-identified models, using the approach of estimating functions (see Godambe, 1960, 1976; Godambe and Heyde, 1987, and Godambe and Thompson, 1974, 1989), as an organizing principle. The discussion in this paper takes place in a random sampling framework and draws heavily from Qin and Lawless (1994), who use the estimating-functions approach to combine the estimating functions in an over-identified model optimally.

1 Introduction.

Let z be a p-dimensional vector with $z \in \mathcal{Z}$ where $\mathcal{Z}$ is a compact subset of $\mathcal{R}^p$. Let θ be a k-dimensional vector with $\theta \in \mathcal{Q}$ and define the vector-valued estimating function g as $g : \mathcal{Z} \times \mathcal{Q} \to \mathcal{R}^r$, such that it is unbiased

$$E[g(z,\theta)] = 0 \qquad (1)$$

for a unique element θ_0 of $\mathcal{Q}$. The function $g(z,\theta)$ is assumed to be twice continuously differentiable with respect to θ.

Equation (1) with $\dim[g] > \dim[\theta]$ often arises from economic theory with optimizing behavior on the part of economic agents. The parameter vector θ_0 is assumed to satisfy $E[g(z,\theta_0)] = 0$, where $g(z,\theta_0)$ is a given vector-valued function of moment conditions implied by economic theory - economic examples of this function in time-series context can be found in Hansen and Singleton (1982), Wirjanto (1995; 1996a, 1997), Amano and Wirjanto (1996a,b; 1997a,b,c) etc. Consequently this paper focuses on the over-identified case, where $\dim[g] > \dim[\theta]$. Given an identically and independently distributed (i.i.d.) sequence $\{z_t\}_{t=1}^n$, this paper is interested in estimating the parameter vector θ.

2 Two-step Estimator

One popular approach to combining the estimating functions in econometrics is briefly mentioned in Qin and Lawless (1994, page 315). This approach considers the optimal (in the sense of minimum asymptotic covariance matrix) linear combination of the r estimating functions. This leads one to estimate θ as the solution to the estimating equations

$$\sum_{t=1}^{n} M^T V g(z_t, \theta) = 0 \tag{2}$$

where $M = E[\partial g(z, \theta_0)/\partial \theta^T]$ has full rank. More generally, an estimate of θ solves the minimization program

$$\text{MIN}_\theta Q(\theta) = \text{MIN}_\theta \left[\sum_{t=1}^{n} g(z_t, \theta)\right]^T V \left[\sum_{t=1}^{n} g(z_t, \theta)\right] \tag{3}$$

over Q, for some positive semi-definite (rxr) symmetric weighting matrix V. Under standard regularity conditions, the minimand of $Q(\theta)$ is a consistent estimator of θ_0. However it is not efficient in the over-identified case. An efficient estimator can be obtained in this case by minimizing $Q(\theta)$ for $V = J^{-1}$, where $J = E[g(z, \theta_0) g(z, \theta_0)^T]$ has full rank (See Hansen, 1982; McCullagh and Nelder, 1989; White, 1984).

In practice the inverse of the weighting matrix, J, is unknown and needs to be estimated from the data. A two-step estimation strategy can be used to implement this procedure. In the first step, an initial consistent estimate θ^0 is obtained by minimizing $Q(\theta)$ for an arbitrary choice of J such as the r-dimensional identity matrix I_r. The optimal weighting matrix is then estimated as $\tilde{J}^{-1} = [\sum_{t=1}^n g(\theta^0) \sum_{t=1}^n g(\theta^0)^T/n]^{-1}$, where $g(\theta^0) \equiv g(z_t, \theta^0)$. In the second step, an efficient estimator θ is obtained from the estimating equations

$$\sum_{t=1}^{n} M^T \tilde{J}^{-1} g(z_t, \theta) = 0 \tag{4}$$

More generally, it is obtained by solving the second-stage minimization program

$$\text{MIN}_\theta \tilde{Q}(\theta) = \text{MIN}_\theta \left[\sum_{t=1}^{n} g(z_t, \theta)\right]^T \tilde{J}^{-1} \left[\sum_{t=1}^{n} g(z_t, \theta)\right] \tag{5}$$

For an obvious reason, the resultant estimator is referred to as a two-step estimator (TSE). This two-step estimator is discussed in the time-series context in Hansen (1982), White (1984), etc.

If the model is correctly specified, and there exists a unique value θ_0 such that the estimating function is unbiased, $E[g(z, \theta_0)] = 0$, then

$$\sqrt{n}(\tilde{\theta}_{\mathrm{TSE}} - \theta_0) \xrightarrow{d} N(0, \Lambda) \tag{6}$$

where $\Lambda = (M^T J^{-1} M)^{-1}$. The normalized objective function, evaluated at the estimated parameters, converges in distribution to a chi-squared random variable with $r - k$ degrees of freedom.

One potential drawback of the two-step estimator discussed above is that in the over-identified case, i.e. when $\dim[g] > \dim[\theta]$, the choice of the weighting matrix will affect efficiency considerations in an important way. More specifically in finite samples or in the case of misspecified model, the way in which the optimal weighting matrix is estimated will affect the efficiency of the estimator $\tilde{\theta}_{\mathrm{TSE}}$.

3 Maximum Empirical Likelihood Estimator

An alternative to the two-step estimator is the one-step estimator based on solving a set of estimating equations for θ and γ, the r-dimensional normalized Lagrange multiplier. Let $\Psi = (\theta, \gamma)^T$. Then the solution $\tilde{\Psi}_{\mathrm{MELE}}$ is obtained by solving a set of estimating equations

$$\sum_{t=1}^{n} l(z_t, \Psi) = 0 \tag{7}$$

where $l(z_t,)\Psi)^T, l_2(z_t, \Psi)^T]^T$, with the following estimating functions

$$\begin{aligned} l_1(z_t, \Psi) &= \frac{1}{1 + \gamma^T g(z, \theta)} \gamma^T \partial g(z, \theta)/\partial \theta^T \\ l_2(z_t, \Psi) &= \frac{1}{1 + \gamma^T g(z, \theta)} g(z, \theta) \end{aligned}$$

This is shown by Qin and Lawless (1994) to be equivalent to solving the following constrained maximization program:

$$\mathrm{MAX}_{p \cdot \theta} \sum_{t=1}^{n} n^{-1} [ln(p_t) - ln(n^{-1})] \tag{8}$$

subject to the restrictions

[a] $\sum_{t=1}^{n} \tilde{p}_t g(z_t, \theta) = 0$;

[b] $\sum_{t=1}^{n} \tilde{p}_t = 1$;

which yields the maximum empirical likelihood estimator (MELE). Under regularity conditions, $\tilde{\theta}_{\text{MELE}}$ is asymptotically efficient for θ_0, i.e.

$$\sqrt{n}(\tilde{\theta}_{\text{MELE}} - \tilde{\theta}_{\text{TSE}}) = o_p(1)$$

The MELE procedure requires that one solves a system of estimating equations in $k+r$ unknown parameters in one step compared to the procedure for the TSE which requires two steps. However some of the estimating equations in this method may potentially be unstable since the matrix of expected derivatives will not have full rank at the limiting values of the parameters, i.e. at $(\theta, \gamma) = (\theta_0, 0)$, the $(k+r) \times (kxr)$ dimensional matrix of derivatives $E[\partial l(z, \theta_0.0)/\partial \Psi^T]$ will have rank r. In practice this may not pose a serious computational problem in specific applications since the empirical likelihood in terms of (θ, γ) gives $(\tilde{\theta}, \tilde{\gamma})$ as a saddle point.

4 Alternative Characterization of the MELE

An alternative characterization of the MELE of Qin and Lawless (1994) is suggested below. It is based on solving a set of generalized estimating equations that takes account of the over-identifying restrictions on the distribution explicitly. In particular for a discrete parameterization of z with known support, this estimator (if it exists) is shown to belong to the MELE.

In the theory of estimating functions (see e.g. Godambe and Heyde (1987) and Godambe and Thompson (1989)) an estimating function $g^*(z, \theta) \in \mathcal{G}$, $\mathcal{G} \equiv \mathcal{Z} \times \mathcal{Q}$, is optimal in $\mathcal{G}$ if the estimator $\tilde{\theta}$ from $g^*(z, \theta) = 0$ has minimum asymptotic covariance matrix. To begin the analysis I partition the vector-valued estimating function $g(z(\theta)$ as

$$g(z, \theta) = \begin{bmatrix} g_1(z, \theta) \\ g_2(z, \theta) \end{bmatrix} \tag{9}$$

where g_1 is a k dimensional vector and g_2 is a $(r-k)$ dimensional vector. Similarly I partition the M and J matrices conformably to the estimating function g as

$$M = \begin{bmatrix} M_1 \\ M_2 \end{bmatrix} = E \begin{bmatrix} \partial g_1(z, \theta_0)/\partial \theta^T \\ \partial g_2(z, \theta_0)/\partial \theta^T \end{bmatrix}$$

and

$$J = \begin{bmatrix} J_{11} & J_{12} \\ J_{12} & J_{22} \end{bmatrix} = E \begin{bmatrix} g_1(z, \theta_0) g_1(z, \theta_0)^T & g_1(z, \theta_0) g_2(z, \theta_0)^T \\ g_2(z, \theta_0) g_1(z, \theta_0)^T & g_2(z, \theta_0) g_2(z, \theta_0)^T \end{bmatrix}$$

It is assumed that g_1 can be used to estimate θ consistently and t_2 represents the over-identifying restrictions on the model. The latter is equivalent to assuming that the submatrix M_1^{-1} exists. It should be pointed out that the partition of the g vector-valued function into $(g_1^T, g_2^T)^T$ will affect the interpretation given to $\tilde{\gamma}$ only. However it will have no effect on the estimator of θ.

Let γ be a $(r-k)$-dimensional vector. Let $\eta = (\eta_1^T, \eta_2^T)^T$, $\eta_1 = (\theta, \gamma)^T$ and $\eta_2 = (\mu_1, \mu_2)^T$. Define $\mu_i = vec(M_i)$ for $i = 1, 2$. Then an estimator $\tilde{\eta}$ can be obtained as a solution to the following system of $r(k+1)$ generalized estimating equaitons

$$\sum_{t=1}^{n} h(z_t, \eta) = 0 \tag{10}$$

where $h(z_t, \eta) = [h_1(z_t, \eta)^T, ..., h_4(z_t, \eta)^T]^T$, with the following estimating functions

$$\begin{aligned}
h_1(z_t, \eta) &= \frac{1}{1 + \gamma^T g_2(z, \theta) - \gamma^T M_2 M_1^{-1} g_1(z, \theta)} g_1(z, \theta) \\
h_2(z_t, \eta) &= \frac{1}{1 + \gamma^T g_2(z, \theta) - \gamma^T M_2 M_1^{-1} g_1(z, \theta)} g_2(z, \theta) \\
h_3(z_t, \eta) &= \frac{1}{1 + \gamma^T g_2(z, \theta) - \gamma^T M_2 M_1^{-1} g_1(z, \theta)} [\mu_1 - vec(\partial g_1 / \partial \theta^T)] \\
h_4(z_t, \eta) &= \frac{1}{1 + \gamma^T g_2(z, \theta) - \gamma^T M_2 M_1^{-1} g_1(z, \theta)} [\mu_2 - vec(\partial g_2 / \partial \theta^T)]
\end{aligned}$$

The resultant estimator $\tilde{\eta}$ defined above is referred to (for the sake of clarity) as a generalized estimating equations estimator (GEEE). Under regularity conditions, the GEEE $\tilde{\eta}_1$ has the following limiting properties:
(i) $\tilde{\eta}_1 \xrightarrow{p} \eta_{10}$, where $\eta_{10} = (\theta_0, 0)^T$; and
(ii) $\sqrt{n}(\tilde{\eta}_1 - \eta_{10}) \xrightarrow{d} (0, \Lambda_1)$, where $\Lambda_1 = \begin{bmatrix} \Lambda_{11} & 0 \\ 0 & \Lambda_{22} \end{bmatrix}$;
$\Lambda_{11} = (M^T J^{-1} M)^{-1}$ and $\Lambda_{22} = [M_2 (M_1^T)^{-1} J_{11} (M_1^T)^{-1} M_2^T - 2 J_{12}^T (M_1^T)^{-1} M_2^T + J_{22}]^{-1}$. The proof is given in Appendix A.

The above limiting result suggests that solving the generalized estimating equations system in (10) yields an asymptotically efficient estimator, i.e. $\sqrt{n}(\tilde{\theta} - \tilde{\theta}_{\text{TSE}}) = o_p(1)$. However unlike the two-step estimator, this estimator is based on solving a set of estimating equations in one step. Therefore, it allows one to estimate the limiting covariance matrix of the estimator, even when the model is misspecified. Under misspecification where there is not a value for θ such that $E[g(z, \theta)] = 0$, the limiting distribution of the estimator can still be obtained as a solution to the estimating equations $\sum_{t=1}^{n} h(z_t, \tilde{\eta}) = 0$ as long as the observations are i.i.d. Let $\hat{\eta}$ be the unique solution to the unbiased estimating functions $E[h(\eta)] = 0$. Under misspecification, the solution to γ in general will not be $\hat{\gamma} = 0$, even though it exists. The limiting distribution of $\tilde{\eta}$ in the case of misspecification of the estimating function is given by $\sqrt{n}(\tilde{\eta} - \hat{\eta}) \xrightarrow{d} (0, \Lambda^*)$, where $\Lambda^* = (M^{*T} J^{*-1} M^*)^{-1}$; $M^*_. = E[\partial h(z, \hat{\eta})/\partial \eta^T]$, and $J^* = E[h(z, \hat{\eta}) h(z, \hat{\eta})^T]$.

The component of the limiting covariance matrix of $\tilde{\eta}$, which corresponds to the variance $\sqrt{n}(\tilde{\theta} - \hat{\theta})$, is not equal to Λ_{11} if there does not exist a value θ_0 at which all r estimating functions $(g_1, g_2, ..., g_r)$ are unbiased. In the case of the two-step estimator, the effects of misspecification on the weighting matrix and the second-stage estimation must be analyzed.

One potential drawback of the GEEE procedure is computation. Since in this procedure, one has to solve a system of $r(k+1)$ equations in $r(k+1)$ unknown parameters, it can be potentially more burdensome in terms of computation than solving a minimization program in k variables twice as in the two-step estimator. It is also likely to be more complex than the computation involved in MELE procedure which requires one to solve a set of estimating equations in $k + r$ unknown parameters. In practice it may or may not pose a serious computational burden in specific applications since this estimator can be used to calculate better approximations to its finite-sample distribution using a saddle-point approximation.

To further interpret the GEEE, I show that for a discrete parameterization of z with known support, γ can be interpreted as the vector of Lagrange multiplier (for the over-identifying restrictions) in the context of the empirical likelihood estimation studied by Qin and Lawless (1994). In this sense the GEEE may be viewed as an alternative formulation of the estimator of θ to Qin and Lawless (1994) as the solution to a system of generalized estimating equations.

Suppose that z has a discrete cumulative distribution function $F(z, p)$ for some unknown $p \in \mathcal{P}$, and F is known. Thus, I have a parametric estimation of θ, such that the estimating function $g(z, \theta)$ is unbiased, i.e. $E[g(z, \theta)] = 0$. I can estimate p efficiently by maximizing thelikelihood function over $\mathcal{P}_g$, where

$$\mathcal{P}_g = \left\{ p \in \mathcal{P} | \exists \theta \vee \int g(z, \theta) F(dz, p) = 0 \right\} \tag{11}$$

Once the maximum-likelihood estimate (MLE) $\tilde{p}$ is obtained, the MLE for θ is given by $\tilde{\theta}$ defined by $\int g(z, \tilde{\theta}) F(dz, \tilde{p}) = 0$.

In an over-identified case, it is important to consider the restrictions implied by the difference between $\mathcal{P}_g$ and $\mathcal{P}$. Hence consider a sequence of random variables with a known support $\{z_1, z_2, ..., z_m\}$ and $p_j = Pr(z = z_j)$ for $j = 1, 2, ..., m$. Then the probability density function is

$$f(z, p) = \prod_{j=1}^{m} p_j^{\phi_j(z)} \tag{12}$$

for $p \in \mathcal{P} = \left\{p \in \mathcal{R}^m | p_j \geq 0, \sum_{j=1}^m p_j = 1, \exists\theta \vee \int g(z,\theta)F(dz,p) = 0\right\}$ where $\phi_j(z) = 1$ if $z = z_j$, and $= 0$ if $z \neq z_j$. Since there exists a value of $\theta \vee \int g(z,\theta)F(dz,p) = 0$, p must be restricted to $\mathcal{P}_g$ defined in (11).

Next I partition g into a k dimensional vector g_1 and a $(r-k)$ dimensional vector g_2, such that $\mathcal{P}_1 = \{p \in \mathcal{P} | \exists\theta$ such that $\int g_1(z,\theta)F(dz,p) = 0\}$ is equal to $\mathcal{P}$ and $\forall p \in \mathcal{P}_1 = \mathcal{P}$. Then the solution for θ to the estimating equations $\int g_1(z,\theta)F(dz,p) = 0$ is unique. Given these assumptions, define θ as a function of p implicitly as

$$\int g_1(z,\theta(p))F(dz,p) = \sum_{j=1}^m p_j g_1(z_j,\theta(p)) = 0 \tag{13}$$

There are two useful properties that can be obtained by differentiating (13) with respect to p_τ

$$g_1(z_\tau,\theta(p)) + \sum_{j=1}^m p_j(\partial g_1(z_j,\theta(p))/\partial\theta)(\partial\theta(p)/\partial p_\tau) = 0. \tag{14}$$

These properties are given by

$$\partial\theta(p)/\partial p_\tau = -[\sum_{j=1}^m p_j(\partial g_1(z_j,\theta(p))/\partial\theta^T)]^{-1} g_1(z_\tau,\theta(p)) \tag{15}$$

and

$$\sum_{j=1}^m p_j \partial\theta(p)/\partial p_\tau = -[\sum_{j=1}^m p_j(\partial g_1(z_j,\theta(p))/\partial\theta^T)]^{-1} \sum_{j=1}^m p_j g_1(z_\tau,\theta(p)) = 0 \tag{16}$$

respectively.

Consider the following problem

$$\max_{p\in\mathcal{P}_g} \sum_{t=1}^n \sum_{j=1}^m \phi_j(z_t) ln(p_j). \tag{17}$$

The assumptions suggest that the maximization can be performed on the set given by

$$\mathcal{P}_g = \left\{p \in \mathcal{R}^m | p_j \geq 0, \sum_{j=1}^m p_j = 1, \int g_2(z,\theta)F(dz,p) = 0\right\} \tag{18}$$

i.e. the maximization program can be written as

$$\max \sum_{t=1}^n \sum_{j=1}^m \phi_j(z_t) ln(p_j) \text{ s.t. } p_j \geq 0; \sum_{j=1}^m p_j = 1; \text{ and } \sum_{j=1}^m p_j g_2(z_j,\theta(p)) \tag{19}$$

in the framework of empirical likelihood of Qin and Lawless (1994).

In the above maximization program I have $(r-k+1)$ nonlinear restrictions, which contain an implicit function that can be solved using a numerical approximation method. However the solution to (19), if exists, will be identical to the solution to the generalized estimating equations in (10), which only requires a solution to a system of $r(k+1)$ exactly-identified equations.

Let t_1 be the Lagrange multiplier associated with the restriction $\sum_{j=1}^{m} p_j$, and t_2 be the vector of Lagrange multipliers associated with the restrictions $g_2(z_j, \theta(p)) = 0$ respectively. Then the first-order conditions for the maximization problem in (19) are

$$[\text{a}] \quad \sum_{j=1}^{m}[\phi_j(z_t)/p_j] - \tilde{t}_1 - \tilde{t}_2^T g_2(z_j, \theta(\tilde{p}))$$

$$-\tilde{t}_2^T[\sum_{\tau=1}^{m} \tilde{p}_\tau(\partial g_2(z_\tau, \theta(\tilde{p}))/\partial\theta^T)](\partial\theta(\tilde{p})/\partial p_j) = 0; j = 1, 2, ..., m;$$

$$[\text{b}] \quad \sum_{j=1}^{m} \tilde{p}_j = 1;$$

$$[\text{c}] \quad \sum_{j=1}^{m} \tilde{p}_j g_2(z_j, \theta(\tilde{p})) = 0.$$

Multiplying the left-hand side of [a] above by $\tilde{p}_j$ and summing over $j = 1, 2, ..., m$ shows that the Lagrange multiplier $\tilde{t}_1$ for the restriction $\sum_{j=1}^{m} p_j = 1$ is n. Thus the solution for $(\tilde{p}, \tilde{t}_1, \tilde{t}_2)$ is characterized by
[a] $\tilde{t}_1 = n$;

$$[\text{b}] \quad \tilde{p}_j = \sum_{t=1}^{n} \phi_j(z_t)/[n + \tilde{t}_2^T g_2(z_j, \theta(\tilde{p})) - \tilde{t}_2^T(\sum_{\tau=1}^{m} \tilde{p}_\tau \partial g_2(z_\tau, \theta(\tilde{p}))/\partial\theta^T)$$

$$(\sum_{\tau=1}^{m} \tilde{p}_\tau \partial g_2(z_\tau, \theta(\tilde{p}))/\partial\theta^T)^{-1} g_1(z_j, \theta(\tilde{p}))];$$

[c] $\sum_{j=1}^{m} \tilde{p}_j g_2(z_j, \theta(\tilde{p})) = 0$.

Given $\tilde{p}$ and $\tilde{t}_2$, the MELE for $\theta = \theta(p)$, which is $\tilde{\theta} = \theta(\tilde{p})$, can be obtained as a solution to the system of equations

$$\sum_{j=1}^{m} \tilde{p}_j g_1(z_j, \theta(\tilde{p})) = 0 \tag{20}$$

Below I show that the MELE for θ given by (20) is equivalent to the GEEE obtained from (10). To this end, let $\tilde{M}_1 = \sum_{j=1}^{m} \tilde{p}_j \partial g_1(z_j, \theta(\tilde{p}))/\partial\theta^T$,

and let $\tilde{M}_2 = \sum_{j=1}^{m} \tilde{p}_j \partial g_2(z_j, \theta(\tilde{p}))/\partial\theta^T$. Let $\tilde{\gamma} = \tilde{t}_2/\tilde{t}_1$. Then the solution for η is characterized by
[a] $\sum_{j=1}^{m} \tilde{p}_j g_1(z_j, \theta(\tilde{p})) = 0$;
[b] $\sum_{j=1}^{m} \tilde{p}_j g_2(z_j, \theta(\tilde{p})) = 0$;
[c] $\tilde{p}_j = n^{-1} \sum_{t=1}^{n} \phi_j(z_t)/[1 + \tilde{\gamma}^T g_2(z_j, \theta(\tilde{p})) - \tilde{\gamma}^T \tilde{M}_2 \tilde{M}_1^{-1} g_1(z_j, \theta(\tilde{p}))]$.
Lastly, substituting the expression for $\tilde{p}_j$, the solution for the estimator $\tilde{\eta}$ can be obtained from the system of generalized estimating equations

$$\sum_{j=1}^{m} \tilde{p}_j g_1(z_j, \theta(\tilde{p})) = 0 \tag{21}$$

Below I show that the MELE or θ given by (20) is equivalent to the GEEE obtained from (10). To this end, let $\tilde{M}_1 = \sum_{j=1}^{m} \tilde{p}_j \partial g_1(z_j, \theta(\tilde{p}))/\partial\theta^T$, and let $\tilde{M}_2 = \sum_{j=1}^{m} \tilde{p}_j \partial g_2(z_j, \theta(\tilde{p}))/\partial\theta^T$. Let $\tilde{\gamma} = \tilde{t}_2/\tilde{t}_1$. Then the solution for η is characterized by
[a] $\sum_{j=1}^{m} \tilde{p}_j g_1(z_j, \theta(\tilde{p})) = 0$;
[b] $\sum_{j=1}^{m} \tilde{p}_j g_2(z_j, \theta(\tilde{p})) = 0$;
[c] $\tilde{p}_j = n^{-1} \sum_{t=1}^{n} \phi_j(z_t)/[1 + \tilde{\gamma}^T g_2(z_j, \theta(\tilde{p})) - \tilde{\gamma}^T \tilde{M}_2 \tilde{M}_1^{-1} g_1(z_j, \theta(\tilde{p}))]$.
Lastly, substituting the expression for $\tilde{p}_j$, the solution for the estimator $\tilde{\eta}$ can be obtained from the system of generalized estimating equations

$$\sum_{t=1}^{n} h(z_t, \eta) = 0 \tag{22}$$

which is equation (10).

The discussion so far has suggested a link between the GEEE and an estimator of the distribution function. To see this, consider z having a discrete distribution, so that the distribution function can be estimated by

$$\tilde{F}^*(z) = \sum_{j=1}^{m} \tilde{p}_j I(z_j \le z) \tag{23}$$

which, after substitution of the expression for $\tilde{p}_j$, yields

$$\tilde{F}^*(z) = n^{-1} \sum_{t=1}^{n} I(z_t \le z)[1 + \gamma^T\{g_2(z_t, \tilde{\theta}) - \tilde{M}_2 \tilde{M}_1^{-1} g_1(z_t, \tilde{\theta})\}]^{-1} \tag{24}$$

can be stated as follows. The probability $q_n = pr(z \in \mathcal{B})$ is estimated by

$$\tilde{q} = \int I(z \in \mathcal{B}) \tilde{F}^*(dz); \quad \text{as } n \to \infty \tag{25}$$

The limiting properties of this estimator can be obtained by noticing that to the unbiased elementary estimating functions $E[g(z, \theta)]$, I can augment the estimating functions $g_0(z, q) = 1q - I(z \in \mathcal{B})$. Let $\varphi = (q, \theta^T)^T$, and let

$$\jmath(z, \varphi) = (g_0(z, q), g(z, \theta)^T)^T$$

Then I partition the estimating functions $1(z, \varphi)$ as

$$\jmath(z, \Psi) = (\jmath_1(z, \varphi), \jmath_2(z, \varphi)^T)^T$$

where $\jmath_1(z, \varphi) = (g_0(z, q), g_1(z, \theta)^T)^T$ and $\jmath_2(z, \varphi) = g_2(z, \theta)$. The new estimating functions are therefore given by

$$\begin{aligned}
\hat{h}_1(z, \varphi) &= (\hat{h}_{10}(z, \varphi), \hat{h}_{11}(z, \varphi)^T)^T; \text{ with} \\
\hat{h}_{10}(z_{,t}, \varphi) &= \frac{1}{1 + \gamma^T g_2(z, \theta) - \gamma^T M_2 M_1^{-1} g_1(z, q)}[q - I(z \in \mathcal{B})]; \text{ and} \\
\hat{h}_{11}(z_t, \varphi) &= h_1(z_t, \eta); \\
\hat{h}_2(z_t, \varphi) &= h_2(z_t, \eta); \\
\hat{h}_3(z_t, \varphi) &= h_3(z_t, \eta); \\
\hat{h}_4(z_t, \varphi) &= h_4(z_t, \eta).
\end{aligned}$$

It folows that the η-component of the solution to the generalized estimating equations

$$\sum_{t=1}^{n} \hat{h}(z_t, \varphi) = 0 \tag{26}$$

where $\hat{h}(z_t, \varphi) = [\hat{h}_{10}(z_t, \varphi)^T, \hat{h}_{11}(z_t, \varphi)^T, ..., \hat{h}_4(z_t, \varphi)^T]^T$, is identical to the estimator $\tilde{\eta}$ obtained from solving the generalized estimating equations in (10). This implies that the q-component of the solution to the above generalized estimating equations is given by (24), rewritten as

$$\tilde{q} = \frac{1}{1 + \tilde{\gamma}^T g_2(z, \tilde{\theta}) - \tilde{\gamma}^T \tilde{M}_2 \tilde{M}_1^{-1} g_1(z, \tilde{q})} n^{-1} \sum_{t=1}^{n} I(z \in \mathcal{B})] \tag{27}$$

The limiting properties of this estimator for the distribution function are:
[i] $\tilde{q} \xrightarrow{p} q_0$; and
[ii] $\sqrt{n}(\tilde{q} - q_0) \xrightarrow{d} N(0, \Omega_q)$,
where $\Omega_q = q(1-q) - g_{\mathcal{B}}^T[J^{-1} - J^{-1}M(M^T J^{-1} M)^{-1} M^T J^{-1}] g_{\mathcal{B}}$, and $g_{\mathcal{B}} = E[g(z, \theta_n) I(z \in \mathcal{B})] = q_0 E[g(z, \theta_0) | z \in \mathcal{B}]$. The estimate $\tilde{q}$ is fully efficient in the sense that its variance attains the semi-parametric efficiency bound. The proof is given in Appendix B.

5 The Minimum Discriminant Information Adjusted Estimator

The GEEE method in Section 4 solves a system of estimating equations of dimension $rx(k+1)$ much larger than $(k+r) \times (k+r)$ as in the MELE procedure, even though the GEEE procedure has the advantage that $E[\partial h(z; \theta_0, 0, \mu_{10},$

$\mu_{20})/\partial\eta^T]$ has full rank, whereas the MELE procedure has the disadvantage that $E[\partial l(z;\theta_0,0)/\partial\Psi^T]$ does not have full rank. In practice this dimensional issue may or may not be a serious handicap to applied workers, depending on specific applications.

However given the potentially computational burden involved in the GEEE procedure it is worthwhile looking at alternative procedures. One such procedure is Haberman's (1984) procedure mentioned in Qin and Lawless (1994, example 3, page 314). Instead of maximizing the empirical likelihood as in (7), this estimator is obtained by minimizing the Kullback-Leibler divergence from the estimated distribution to the empirical distribution. Thus the resultant estimator is referred to as the minimum discriminant information adjusted estimator (MDIAE).

The MDIAE of Ψ can be obtained by solving a set of estimating equations for θ and γ, the r-dimensional normalized tilting parameter in

$$\mathrm{MAX}_{p\cdot\theta}\sum_{t=1}^{n}p_t[ln(n^{-1})-ln(p_t)] \tag{28}$$

subject to the restrictions
[a] $\sum_{t=1}^{n}\tilde{p}_t g_1(z_j,\theta(\tilde{p}))=0$;
[b] $\sum_{t=1}^{n}\tilde{p}_t=1$.
Thus the solution $\tilde{\Psi}_H$ is obtained by solving a set of estimating equations

$$\sum_{t=1}^{n}\omega(z_t,\Psi)=0 \tag{29}$$

where $\omega(z_t,\Psi)=[\omega_1(z_t,\Psi)^T,\omega_2(z_t,\Psi)^T]^T$, with following estimating functions $\omega_1(z_t,\Psi)=\gamma^T(\partial g(z,\theta)/\partial\theta^T)\exp[\gamma^T g(z,\theta)]$; $\omega_2(z_t,\Psi)=g(z,\theta)\exp[\gamma^T g(z,\theta)]$. Under regularity conditions, $\tilde{\theta}_{\mathrm{MDIAE}}$ is asymptotically efficient for θ_0, i.e. $\sqrt{n}(\tilde{\theta}_{\mathrm{MDIAE}}-\tilde{\theta}_{\mathrm{TSE}})=o_p(1)$.

There is a number of attractive features of the MDIAE. First, in the MDIAE procedure, the discrepancy between the estimated probabilities p_t and the empirical frequency n^{-1} is weighted using the efficient estimate of these probabilities $\tilde{p}_t$; in contrast in the MELE procedure, the discrepancy is weighted using an inefficient estimate of these probabilities n^{-1}. Second, built on Huber (1980), the influence functions of the estimators defined by estimating equations for the MELE and MDIAE are given by $E[\partial l(z,\Psi)/\partial\Psi^T]^{-1}l(z,\Psi)$ and $E[\partial\omega(z,\Psi)/\partial\Psi^T]^{-1}\omega(z,\Psi)$ respectively. At the limiting values $(\theta,\gamma)=(\theta_0,0)$ the influence functions for the MELE and MDIAE are identical. However the influence function for MELE, in contrast to the MDIAE, can be unbounded at $\gamma=\epsilon$ where $\epsilon\ni 0$, even if the estimating functions themselves $g(z,\theta)$ are bounded. Third, the MDIAE procedure potentially is more attractive in terms of computation than the MELE and

GEEE procedures. This is because the estimated probabilities in the MDIAE procedure is given by $\tilde{p}_t = \exp[\gamma^T g(z,\theta)]/\sum_{t=1}^n \exp[\gamma^T g(z,\theta)]$. Replacing p_t in (27) with this and re-arranging terms result in the constrained maximization program

$$\text{MAX}_{\gamma\cdot\theta}\{ln(\sum_{t=1}^{n}\exp[\gamma^T g(z_t,\theta)]) - ln(n)\}$$

$$\text{subject to} \tag{30}$$

$$\partial\{ln(\sum_{t=1}^{n}\exp[\gamma^T g(z_t,\theta)]) - ln(n)\}/\partial\gamma = 0$$

which is computationally easier to solve. It is clear that the estimating equations in (28) amounts to choosing (θ,γ) with the first derivatives of the object function in (29) with respect to θ and with respect to γ set equal to zero.

Lastly it also is possible to obtain an estimator that has identical first-order limiting properties as the GEEE, using the principle of minimizing the Kullback-Leibler divergence measure in Haberman (1984). In terms of computation however, it is potentially inferior to the MDIAE procedure applied to the formulation proposed by Qin and Lawless (1994). For this reason, it is not discussed in this paper.

6 Concluding Remarks

This paper has provided a selected survey on the efficient estimation of over-identified models common in Economic models with optimizing agents, using the theory of estimating functions as an organizing principle. It was argued that the two-step estimator is sensitive to the way in which the optimal weighting matrix is estimated from the data. The MELE of Qin and Lawless (1994) solves this problem by estimating a set of estimating equations in one step. However some of the estimating equation in the MELE procedure may potentially be unstable because the matrix of their expected derivatives does not have full rank at the limiting values. Although in practice this may not pose a computation burden in specific applications, an alternative characterization of the MELE was suggested in this paper. It is based on a set of generalized estimating equations, which does not share the shortcoming of the MELE mentioned above. This set of generalized estimating equations incorporates information provided by the over-identifying restrictions on the distribution function explicitly, resulting in a just-identified set of estimating equations with an exact solution. Unfortunately the computation of the result GEEE appears to be much less attractive than that of the MELE since the estimating equations in GEEE method has the dimension much larger than the dimension of the estimating equation in the MELE method. In practice this may or may not pose a computational problem in particular applications since it is possible to use this estimator to compute a saddle-point approximation to the finite-sample distribution. Next an alternative estimator to the MELE based on Haberman's (1984) minimization of Kullback-Leibler divergence measure is discussed. It has a number of appealing theoretical features in terms of finite-sample efficiency and robustness and a priori it has a more tractable computational requirement.

It is worthwhile to make two final remarks: [i] the estimating functions approach is a natural framework to use in over-identified models. One of its major advantage over other known approaches, which has not been emphasized much in this literature, is that in this approach it is possible to treat over-identified models as just-identified models essentially be re-defining the estimating functions appropriately; [ii] another asymptotically equivalent estimator to the ones discussed in this paper can be obtained by combining the set of elementary estimating functions $g_1(z, \theta)$, which is unbiased, i.e. $E[g_1(z, \theta)] = 0$ with the set of auxiliary estimating functions $t_2(z, \theta)$ which represents the over-identifying restrictions. This set of estimating functions is biased with the biased term given by a r-dimensional auxiliary parameter vector λ, i.e. $E[g_2(z, \theta)] = \lambda$. However λ will be a zero vector identically, so that the set of elementary estimating functions will be unbiased, when the over-identifying restrictions are satisfied by the data. Under regularity conditions the resultant estimators will be consistent (i.e. $\tilde{\lambda}$ will be consistent for a zero vector), asymptotically normal and asymptotically efficient. However the efficiency is obtained from a joint estimation of θ and λ, without using the prior information that $\lambda = 0$. This prior information can be used simply by projecting the estimator of θ orthogonally on to the tangent space in which $\lambda = 0$. This projection, which will be done approximately at the appropriate tangent space, corresponds to the one-step estimators discussed in this paper. Recently Wirjanto (1996b) has applied this approach to study a collection of generalized linear models in McCullagh and Nelder (1989) with contemporaneous correlations across the regression error terms. The resulting one-step estimator is shown to be not only efficient but also robust to nonconstant variances in each model's error terms as well as nonconstant correlations across the equation error terms.

APPENDIX A

Since $E[h(\eta_0)] = E[h(\theta_0, 0, \mu_{10}, \mu_{20})^T] = 0$, there exists a consistent root of the generalized estimating equation

$$\sum_{t=1}^{n} h(z_t, \tilde{\eta}) = 0.$$

Regularity conditions ensure that

$$\sqrt{n}(\tilde{\eta}-\eta_0) \xrightarrow{d} (0,\Lambda)$$

where $\Lambda(D^T\Sigma^{-1}D)^{-1}$, $D = E[\partial h(z,\eta_0)/\partial\eta^T]$, and $\Sigma = E[h(z,\eta_0)h(z,\eta_0)^T]$. Next let me partition the matrices D and Σ conformably to $\eta = [\eta_1;\eta_2]^T$, where $\eta_1 = (\theta,\gamma)^T$, and $\eta_2 = (\mu_1,\mu_2)^T$, as

$$D = \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}; \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

where

$$D_{11} = \begin{bmatrix} M_1 & -J_{12} + J_{11}(M_1^T)^{-1}M_2^T \\ M_2 & -J_{22} + J_{12}^T(M_1^T)^{-1}M_2^T \end{bmatrix}; D_{12} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

and

$$D_{22} = \begin{bmatrix} I_{2211} & 0 \\ 0 & I_{2222} \end{bmatrix}; \Sigma_{11} = \begin{bmatrix} J_{11} & J_{12} \\ J_{12}^T & J_{22} \end{bmatrix}$$

Here, I_{2211} and I_{2222} are identity matrices of dimensions $(k(k-1))/2$ and $((r-k)(r-k-1))/2$ respectively.

The limiting covariance matrix of $\tilde{\eta}_1$ is thus given by

$$\begin{aligned} \Lambda_1 &= [D_{11}^{-1}\Sigma_{11}(D_{11}^T)^{-1}]^{-1} \\ &= D_{11}^T\Sigma_{11}^{-1}D_{11} \end{aligned}$$

Since I can write

$$\begin{bmatrix} M_1 & -J_{12} + J_{11}(M_1^T)^{-1}M_2^T \\ M_2 & -J_{22} + J_{12}^T(M_1^T)^{-1}M_2^T \end{bmatrix} = J \begin{bmatrix} (M_1^T)^{-1}M_2^T \\ -I_{r-k} \end{bmatrix}$$

the above expression for the matrix Λ_1 simplifies to

$$\begin{bmatrix} (J^TM^{-1}J)^{-1} & 0 \\ 0 & [M_2(M_1^T)^{-1}J_{11}(M_1^T)^{-1}M_2^T - 2J_{12}^T(M_1^T)^{-1}M_2^T + J_{22}]^{-1} \end{bmatrix}$$

which yields the intended result.

APPENDIX B

The minimum bound for the parameter vector Ψ is given by

$$\Omega(\Psi) = (A^TB^{-1}A)^{-1}$$

where $A = E[\partial \jmath(z,\Psi_0)/\partial\Psi^T]$ and $B = E[\jmath(z,\Psi_0)\jmath(z,\Psi_0)^T]$.

The minimum bound for the parameter q can be obtained by noting that

$$A = \begin{bmatrix} 1 & 0 \\ 0 & M \end{bmatrix}; B^{-1} = \begin{bmatrix} q(1-q) & -g_{\mathcal{B}} \\ -g_{\mathcal{B}} & J \end{bmatrix} = \begin{bmatrix} B^{11} & B^{12} \\ B^{21} & B^{22} \end{bmatrix}$$

where
$g_{\mathcal{B}} = E[g(z,\theta_n)I(z \in \mathcal{B})] = q_n E[g(z,\theta_n)|z \in \mathcal{B}]$;
$B^{11} = [q(1q) - g_{\mathcal{B}}^T J^{-1} g_{\mathcal{B}}]^{-1}$;
$B^{12} = [q(1q) - g_{\mathcal{B}}^T J^{-1} g_{\mathcal{B}}]^{-1} g_{\mathcal{B}}^T J^{-1}$;
$B^{22} = J^{-1} + J^{-1} g_{\mathcal{B}}[q(1q) - g_{\mathcal{B}}^T J^{-1} g_{\mathcal{B}}]^{-1} g_{\mathcal{B}}^T J^{-1}$.
Therefore I have

$$\Omega(\Psi)^{-1} = \begin{bmatrix} \Omega^{11} & \Omega^{12} \\ \Omega^{21} & \Omega^{22} \end{bmatrix}$$

where
$\Omega^{11} = [q(1q) - g_{\mathcal{B}}^T J^{-1} g_{\mathcal{B}}]^{-1}$;
$\Omega^{12} = -[q(1q) - g_{\mathcal{B}}^T J^{-1} g_{\mathcal{B}}]^{-1} g_{\mathcal{B}}^T J^{-1} M$;
$\Omega^{22} = M^T J^{-1} M + M J^{-1} g_{\mathcal{B}}[q(1q) - g_{\mathcal{B}}^T J^{-1} g_{\mathcal{B}}]^{-1} g_{\mathcal{B}}^T J^{-1} M$.

Since the new estimating function does not affect the variance of the estimator $\tilde{\theta}$, I have

$$\Omega_{22} = (M^T J^{-1} M)^{-1}$$

regardless of what the set $\mathcal{B}$ is. The element of $\Omega(\Psi)$ corresponding to the parameter q is simply

$$\Omega_{11}[q(1q) - g_{\mathcal{B}}^T J^{-1} g_{\mathcal{B}}] + g_{\mathcal{B}}^T J^{-1}(M^T J^{-1} M)^{-1} M^T J^{-1} g_{\mathcal{B}}.$$

Therefore, I obtain the result that

$$var(\tilde{q}) = q(1-q)$$

If I have extra information in the form of unbiased estimating function with θ known, then

$$var(\tilde{q}) = q(1-q) - g_{\mathcal{B}}^T J^{-1} g_{\mathcal{B}}$$

It remains to show that the variance of the estimator $\tilde{q}$ in (34) attains this minimum bound. Following the steps in section 2 for the new estimating function $\jmath(z,\Psi)$ results in a one-step estimator. Using the result in Appendix A, it follows that its variance will be identical to the variance of the two-step estimator. Therefore, it attains the minimum bound and the resultant one-step estimator is $\tilde{q}$. The key to this result is that

$$\begin{aligned} L_1 &= E[\partial \jmath_1(z,\Psi_0)/\partial \Psi^T] \\ &= \begin{bmatrix} 1 & 0 \\ 0 & M_1 \end{bmatrix} \\ L_2 &= E[\partial \jmath_2(z,\Psi_0)/\partial \Psi^T] = (0, M_2) \end{aligned}$$

so that

$$L_2 L_1^{-1} \jmath_1 = M_2 M_1^{-1} g_1$$

Therefore I have the results that

$$\hat{h}_{10}(z_t, \eta, q) = \frac{g_0(z,\theta)}{1 + \gamma^T \jmath_2(z,\theta) - \gamma^T L_1^{-1} \jmath_1(z,q)}$$

which can be rewritten as

$$\hat{h}_{10}(z_t, \eta, q) = \frac{q - I(z \in \mathcal{B}]}{1 + \gamma^T g_2(z,\theta) - \gamma^T M_2 M_1^{-1} g_1(z,q)}]$$

Similarly, it can be shown that

$$\begin{aligned}
\hat{h}_{11}(z_t, \eta, q) &= \frac{g_1(z,\theta)}{1 + \gamma^T g_2(z,\theta) - \gamma^T M_2 M_1^{-1}(z,\theta)} \\
\hat{h}_{2}(z_t, \eta, q) &= \frac{g_2(z,\theta)}{1 + \gamma^T g_2(z,\theta) - \gamma^T M_2 M_1^{-1} g_1(z,\theta)} \\
\hat{h}_{3}(z_t, \eta, q) &= \frac{g_1(z,\theta) - vec(\partial g_1/\partial\theta^T)}{1 + \gamma^T g_2(z,\theta) - \gamma^T M_2 M_1^{-1} g_1(z,\theta)} \\
\hat{h}_{4}(z_t, \eta, q) &= \frac{g_2(z,\theta) - vec(\partial g_2/\partial\theta^T)}{1 + \gamma^T g_2(z,\theta) - \gamma^T M_2 M_1^{-1} g_1(z,\theta)}
\end{aligned}$$

Since $\hat{h}_{11}(z_t, \eta, q) = h_1(z_t, \eta)$, $\hat{h}_2(z_t, \eta, q) = h_2(z_t, \eta)$, $\hat{h}_3(z_t, \eta, q) = h_3(z_t, \eta)$, and $\hat{h}_4(z_t, \eta, q) = h_4(z_t, \eta)$, the intended results follow.

Acknowledgements. Support from the Social Science and Humanities Council of Canada is acknowledged. The first version of this paper was prepared for the Institute of Mathematical Statistic (IMS) Special Topics Meeting Symposium on Estimating Functions, held at the University of Georgia, Athens Georgia, U.S.A. March'96. I am indebted to two anonymous referees and the editor, whose probing questions have substantially improved both the presentation and the content of the paper. The usual disclaimer applies.

REFERENCES

AMANO, R. A. and WIRJANTO, T. S. (1996a). Money stock targetting and money supply: a closer examination of the data. *Journal of Applied Econometrics*, **11** 93-104.

AMANO, R. A. and WIRJANTO, T. S. (1996b). Intertemporal substitution, imports and the permanent income model. *Journal of International Economics*, **40** 439-457.

AMANO, R. A. and WIRJANTO, T. S. (1997a). intratemporal substitution and government spending. *The Review of Economics and Statistics* (in press).

AMANO, R. A. and WIRJANTO, T. S. (1997b). Adjustment costs and import demand behaviour: evidence from Canada and the United States. *Journal of International Money and Finance* (in press).

AMANO, R. A. and WIRJANTO, T. S. (1997c). An empirical study of dynamic labour demand with integraded forcing processes. *Journal of Macroeconomics* (in press).

GODAMBE, V. P. (1960). An optimum property of regular maximum lieklihood estimation. *Ann. Math. Statist.* **31** 1208-1212.

GODAMBE, V. P. (1976). Conditional likelihood and unconditional optimum estimating equations. *Biometrika*, **63** 277-284.

GODAMBE, V. P. and HEYDE, C. C. (1987). Quasi-likelihood and optimal estimation. *Int. tatist. Rev.* **55** 231-244.

GODAMBE, V. P. and THOMPSON, M. E. (1974). Estimating equations in the presence of nuisance parameter. *Ann. Statist.* **2** 568-571.

GODAMBE, V. P. and THOMPSON, M. E. (1989). An extension of quasi-likelihood (with discussion). *J. Statist. Plan. Infer.* **22** 137-172.

HABERMAN, S. J. (1984). Adjustment by minimum discriminant information. *Ann. Statist.* **12** 971-988.

HANSEN, L. P. (1982). Large sample properties of generalized method of moment estimators. *Econometrica* **50** 1029-1054.

HANSEN, L. P. and SINGLETON, K. J. (1982). Generalized instrumental variables estimation of non-linear rational expectations model. *Econometrica* **50** 1269-1286.

HUBER, P. J. (1980). *Robust Statistics.* Wiley, New York.

McCULLAGH, P. and NELDER, J. (1989). *Generalized Linear Models*, Second ed. Chapman and Hall, London.

QIN, J. and LAWLESS, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22** 300-325.

WHITE, H. (1984). *Asymptotic Theory for Econometricians.* Academic Press, Orlando.

WIRJANTO, T. S. (1995). Aggregate consumption behaviour and liquidity constraints: the Canadian evidence. *Canadian Journal of Economics*, **28** 1135-1152.

WIRJANTO, T. S. (1996a). An empirical investigation into the permanent income hypothesis: further evidence from the Canadian data. *Applied Economics* **28** 1451-1461.

WIRJANTO, T. S. (1996b). On the efficiency of the seemingly unrelated regressions, unpublished manuscript, University of Waterloo, Waterloo, Ontario, Canada.

WIRJANTO, T. S. (1997). Aggregate consumption behaviour with time nonseparable preferences and liquidity constraints. *Applied Financial Economics* **7** 107-104.