Institute of Mathematical Statistics

LECTURE NOTES — MONOGRAPH SERIES

OPTIMAL INSTRUMENTAL VARIABLE ESTIMATION FOR LINEAR MODELS WITH STOCHASTIC REGRESSORS USING ESTIMATING FUNCTIONS

A. C. Singh
Statistics Canada and Carleton University

R. P. Rao
Sri Sathya Sai Institute of Higher Learning
Prasanthi Nilayam, India

ABSTRACT

In the usual Gauss-Markov (GM) framework of structural linear models, the GM estimators of the regression parameters become inconsistent if at least one of the regressors is correlated with the model error. The reason for this is that the transformation matrix in the GM estimating equation, which transforms the data to the parameter space (this happens to coincide with the design matrix $\boldsymbol{X}$), cannot be regarded as conditionally fixed. Using a generalization of the method of estimating function of Godambe and Thompson (1989) to structural models, it is shown that an asymptotically consistent and optimal (in a restricted sense) estimator can be obtained by replacing the transfromation matrix $\boldsymbol{X}$ by $\tilde{E}_c(\boldsymbol{X})$, the linear regression of $\boldsymbol{X}$ on a given set of conditioning variables; the optimality is restricted in that it depends on the conditioning set. The matrix $\tilde{E}_c(\boldsymbol{X})$ can be viewed as a working (because of restricted optimality) transformation matrix with the desirable property of being uncorrelated with the model error but correlated with $\boldsymbol{X}$. Although finding an unrestricted optimal transformation matrix is not generally feasible in practice, it is shown using the estimating function framework that a lower bound to the asymptotic covariance can be found. This bound is then used to propose a measure of asymptotic efficiency of the estimator. It is observed that the concept of a working transformation matrix is equivalent to that obtained from the method of instrumental variables. Through examples from different areas of modelling such as simultaneous equations, latent variables, and measurement errors, it is illustrated that the structural model estimating function provides a unifying principle which recovers existing results as well as leads to new results.

KEY WORDS: Conditioning variables; consistency; matrix Cauchy-Schwarz inequality.

1 INTRODUCTION

We consider a linear semiparametric model $\boldsymbol{y} = \boldsymbol{X\theta} + \boldsymbol{\epsilon}$ where $\boldsymbol{\epsilon} \sim (\boldsymbol{0}, \boldsymbol{\Gamma})$, $\boldsymbol{\Gamma} = \sigma_\epsilon^2 \boldsymbol{I}$, $\boldsymbol{X}$ is a $n \times p$ matrix with rank p, and $\boldsymbol{\theta}$ is a p-vector of fixed parameters. In general, $\boldsymbol{\Gamma}$ may be $\sigma^2 \boldsymbol{\Lambda}$ where $\boldsymbol{\Lambda}$ is assumed to be nonsingular and known except for its possible dependence on $\boldsymbol{\theta}$. If some of the p-regressors, $x_1, \ldots, x_p$, are stochastic, then such a model is termed 'structural', while if all the x's are fixed, it is termed 'functional' as defined in Fuller (1987, p. 2). For the case of functional linear models, it is known that when the covariance $\boldsymbol{\Gamma}$ depends on $\boldsymbol{\theta}$, the Gauss-Markov (GM) approach, or more generally, the least squares (LS) approach may fail in that the resulting estimate may be inconsistent. However, the method of estimating function of Godambe (1960) and Godambe and Thompson (1989, henceforth GT) does give an optimal and consistent estimate; for a good review see Godambe and Kale (1991). For the case of structural linear models also, the GM or the LS approach may fail when at least one of the regressors is correlated with the model error. Some examples are the cases of latent variable, simultaneous equation, and measurement error models which are often used in econometrics. In these situations, the method of generalized instrumental variable estimation (GIVE) is commonly used, see e.g., Harvey (1981, Ch. 2, p. 80).

In this paper, for structural models we first make a connection between the concept of instrumental variables and that of conditioning variables used in GT methodology, and then propose a generalization of the GT estimator termed as the structural model estimating function (SMEF) estimator. The SMEF estimator can be used when conditional expectations in GT are not specified but can be approximated by a linear regression function. Different specifications of the conditioning or instrumental variables give rise to different SMEF estimators. For a given set of conditioning variables, it is shown that SMEF and GIVE methodologies yield identical estimates and thus optimality of GIVE estimators can be justified from the optimality of estimating functions. The optimality of each GIVE estimator is restricted in that it depends on the conditioning variables. It is also shown that a GIVE method can be improved by including a variable that is identically one to the set of instrumental variables. This improved version of GIVE arises naturally within the SMEF framework. Apparently, the distinction between instrumental variables with and without the inclusion of constant (i.e., 1) has not been emphasized in the literature. Also, since the GIVE estimators are optimal in a restricted sense, it will be useful to have a measure of the asymptotic efficiency of the GIVE estimator. Such a measure is proposed by finding a lower bound to the asymptotic covariance. The lower bound, however, is generally not attainable because the corresponding optimal instruments are not obtainable in practice.

The organization of the paper is as follows. Section 2 provides motivation of the proposed method of SMEF for optimal instrumental variable estimation. For this purpose both GM and GT estimators are first reviewed from functional and structural model perspectives. The SMEF method is presented in Section 3. Several illustrative examples are given in Section 4. Finally, Section 5 contains concluding remarks.

2 MOTIVATION OF THE PROPOSED METHOD WITH REVIEW

It will be helpful to review methods for functional linear models, i.e., models with fixed regressors. We will first consider the GM theorem (which gives BLUE-best linear unbiased estimator) and later the GT theorem which generalizes GM and gives the optimal method of estimating function. In parallel, problems arising from structural models will be discussed for motivating the proposed method.

2.1 Gauss-Markov Theorem

For the GM set-up, it is assumed that x-variables are fixed and $\boldsymbol{\Gamma}$ does not depend on $\boldsymbol{\theta}$. According to the GM theorem, the optimal (BLUE) estimator $\hat{\boldsymbol{\theta}}_{\text{BLUE}}$ is obtained as a solution of the estimating equation

$$\boldsymbol{X}'\boldsymbol{\Gamma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}) = \mathbf{0}, \tag{2.1}$$

and is given by

$$\hat{\boldsymbol{\theta}}_{\text{BLUE}} = (\boldsymbol{X}'\boldsymbol{\Gamma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}\boldsymbol{\Gamma}^{-1}\boldsymbol{y}. \tag{2.2}$$

The estimator $\hat{\boldsymbol{\theta}}_{\text{BLUE}}$ has the small sample optimality of BLUE. Also, under standard regularity conditions, we have as $n \to \infty$,

$$\hat{\boldsymbol{\theta}}_{\text{BLUE}} \to_d N_p[\boldsymbol{\theta}, (\boldsymbol{X}'\boldsymbol{\Gamma}^{-1}\boldsymbol{X})^{-1}]. \tag{2.3}$$

Note that we have used somewhat loosely the above notation for the asymptotic distribution because the covariance matrix depends on n. Now observe that the estimating equation (2.1) has four components:

(i) the n-vector of zero functions $\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}$; a zero function is a function of $\boldsymbol{y}$ *or* $\boldsymbol{\theta}$ (or both) such that it is zero in expectation,

(ii) the $n \times n$ inverse covariance matrix $\boldsymbol{\Gamma}^{-1}$, it gives differential weights to zero functions depending on their precision,

(iii) the $p \times n$ transformation matrix $\boldsymbol{X}'$ which transforms the zero function vector from the data space (of dim n) to the parameter space (of dim p), and

(iv) the p-vector of zeros on the right hand side of the equation; the transformed zero functions from the left hand side are set equal to zero which is closest under the mean squared error norm.

It may be of interest to note that the solution of the estimating equation (2.1) enjoys robustness (in the sense that the estimator remains unbiased and consistent) when $\boldsymbol{\Gamma}$ represents a working covariance matrix. This occurs in situations where the true covariance is difficult to specify or approximate. This robustness property holds more generally as was observed by Liang and Zeger (1986) in defining generalized estimating equations. If $\boldsymbol{\Sigma}$ denotes the working covariance, we have a sandwich-type variance for the suboptimal estimator, $\hat{\boldsymbol{\theta}}$, as $n \to \infty$, and is given by

$$\hat{\boldsymbol{\theta}} \to_d N_p[\boldsymbol{\theta}, \boldsymbol{Q}^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{Q}^{-1}], \quad \boldsymbol{Q} = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}). \tag{2.4}$$

In the above, there is quite a bit of flexibility in choosing the working covariance matrix (it can be stochastic, for example), the main requirement being that the covariance term in the normal approximation (2.4) must be $O_p(n^{-1})$. Similarly, we will define the concept of a working transformation matrix which will be useful in dealing with structural linear models in Section 3. If a (working) transformation matrix $\boldsymbol{F}'$ other than the optimal one $\boldsymbol{X}'$ is used, then the resulting suboptimal estimator $\hat{\boldsymbol{\theta}}(\boldsymbol{F})$ is also robust with a different sandwich-type variance expression. We have as $n \to \infty$

$$\hat{\boldsymbol{\theta}}(\boldsymbol{F}) \to_d N_p[\boldsymbol{\theta}, (\boldsymbol{F}'\boldsymbol{\Gamma}^{-1}\boldsymbol{X})^{-1}(\boldsymbol{F}^{-1}\boldsymbol{\Gamma}^{-1}\boldsymbol{F})(\boldsymbol{X}'\boldsymbol{\Gamma}^{-1}\boldsymbol{F})^{-1}]. \tag{2.5}$$

So far the covariates $\boldsymbol{X}$ were considered fixed, i.e. as a matrix of constants. However, if $\boldsymbol{X}$ is random as in the case of structural models, then provided that $\boldsymbol{X}$ is independent of $\boldsymbol{\epsilon}$, the GM theorem remains valid conditional on $\boldsymbol{X}$. Alternatively, if $\boldsymbol{X}$ is only uncorrelated with but not necessarily independent of $\boldsymbol{\epsilon}$, then the optimality of the GM-estimator becomes asymptotic, see Section 3.3.

Now for structural models, $\boldsymbol{X}$ is often correlated with $\boldsymbol{\epsilon}$, i.e., $\boldsymbol{y} - \boldsymbol{X\theta}$. For example, in the case of distributed lag model, we have for $1 \leq t \leq T$,

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + \alpha y_{t-1} + \epsilon_t, \tag{2.6}$$

where given $\{\boldsymbol{x}_t, y_{t-1}\}$, $\epsilon_t \sim (0, \sigma_t^2)$ and uncorrelated over t. A typical example from econometrics may define the variable y_t as rate of consumption and $\boldsymbol{x}_t$ as disposable income at time t. Here GM fails because the set of variables $\{y_{t-1} : 1 \leq t \leq T\}$ which are part of covariates are not independent of model errors $\{\epsilon_t : 1 \leq t \leq T\}$. Using the concept of estimating equation, Durbin (1960) showed that the least squares method for the model (2.6) does provide consistent estimates using the fact that for each time t, the elementary estimating equation is unbiased because ϵ_t is uncorrelated with the covariates $\boldsymbol{x}_t$ and y_{t-1}. However, Durbin did not establish optimality of the least squares estimate, although he did define an optimality criterion of the estimating equation and showed that the score equation for parametric models does satisfy the optimality criterion. Interestingly enough, around the same time, Godambe (1960) established a stronger optimality property of the score function. Further developments on Godambe's optimal estimating functions led to a general result of GT (1989) from which optimality of the least squares estimate for the distributed lag model easily follows.

2.2 Godambe-Thompson Theorem

For functional linear models, the GT theorem is more general than GM in that variance is allowed to depend on the mean parameters. For structural linear models, the GT theorem generalizes the GM theorem by allowing hierarchical conditioning with respect to a set of variables related to the random covariates. Consider, for example, the case of the distributed lag model (2.6). The conditioning variables are defined in a hierarchical manner for the vector of zero functions $\{g_t := y_t - \boldsymbol{x}_t'\boldsymbol{\beta} - \alpha y_{t-1}, 1 \leq t \leq T\}$ by the increasing sequence of conditioning sets $A_t = \{\boldsymbol{x}_{t'}, y_{t'-1}, 1 \leq t' \leq t\} \cup A_0$ for $1 \leq t \leq T$, where A_0 denotes the initial conditioning set, i.e., all those x's which are independent of error. Here the corresponding σ-fields define the conditional expectation operator to be denoted as $E_c(\cdot)$. For $t \geq 1$, define $E_t(\cdot)$ as $E(\cdot|A_{t-1})$. Then for the random variable involving t, we define E_c as E_t. Now, for a prespecified $E_c(\cdot)$, the zero functions g_t are required to be conditional zero functions, i.e., $E_c(g_t) = 0$; this holds for the above example. We need the conditional covariance of the T-vector $\boldsymbol{g}$ which for our example is easily seen to be the diagonal matrix $\sigma_t^2 \boldsymbol{I}$ using the hierarchical conditioning argument, and also need the conditional transformation matrix $-E_c(\partial \boldsymbol{g}/\partial \boldsymbol{\theta}')$ for $\boldsymbol{\theta} = (\boldsymbol{\beta}', \alpha)'$, which is simply $(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_T)$ where $\boldsymbol{z}_t = (\boldsymbol{x}_t', y_{t-1})'$. Now, the GT optimal estimating function has the same form as that of GM (see (2.1)) except that covariance and transformation matrices are replaced by the corresponding conditional ones. It is easily seen that for the model (2.6), this leads to the LS estimating equation.

To define the GT-optimality criterion with respect to a given $E_c(\cdot)$, consider in general K subsets $A_1 \subset \ldots \subset A_K$ of the conditioning variables corresponding to the K subsets of the conditional zero function vector $\boldsymbol{g}(\boldsymbol{y}, \boldsymbol{\theta})$ of dimension n. Now denoting by $\boldsymbol{G}_c'$, the conditional transformation matrix of gradients, $-E_c(\partial \boldsymbol{g}/\partial \boldsymbol{\theta}')'$, and $\boldsymbol{\Gamma}_c$, the conditional covariance of $\boldsymbol{g}$ which will be block diagonal with K blocks, the optimal estimating function of GT for estimating $\boldsymbol{\theta}$ is given by

$$\boldsymbol{G}_c'\boldsymbol{\Gamma}_c^{-1}\boldsymbol{g}(\boldsymbol{y},\boldsymbol{\theta}) = \mathbf{0}. \tag{2.7}$$

It is assumed that $\boldsymbol{G}_c$ has full rank and a unique solution $\hat{\boldsymbol{\theta}}_{\text{MEF}}$ exists where MEF denotes the method of estimating function. Clearly, the resulting estimator depends on E_c, i.e., the conditioning variables. Note that in the particular case of $A_k = A_1$, $1 \le k \le K$, i.e., when the conditioning variables are common for all g_i's, $1 \le i \le n$ (which implies that $E_c = E_2$), there are two special cases which give rise to GM: firstly when all covariates are constants, i.e., A_1 corresponds to the trivial conditioning variable for the sure event, and secondly when all covariates are conditioned, i.e., $A_1 = \boldsymbol{X}$.

Now, consider the class of estimating functions $\boldsymbol{F}'\boldsymbol{\Gamma}_c^{-1}\boldsymbol{g}$ defined by transformation matrices $\boldsymbol{F}$ such that $\boldsymbol{F}'\boldsymbol{\Gamma}_c^{-1}\boldsymbol{G}_c$ is nonsingular. This is a linear class of unbiased estimating functions except that coefficients of the linear combination are allowed to depend on the parameter $\boldsymbol{\theta}$ and the conditioning variables. Then, the GT-theorem states that the optimal $\boldsymbol{F}$ is given by $\boldsymbol{G}_c$ in the sense that it "minimizes" the following expression

$$\boldsymbol{V}(\boldsymbol{F}) := [E_1(\boldsymbol{F}'\boldsymbol{\Gamma}_c^{-1}\boldsymbol{G}_c]^{-1}[E_1(\boldsymbol{F}'\boldsymbol{\Gamma}_c^{-1}\boldsymbol{F})][E_1(\boldsymbol{G}_c'\boldsymbol{\Gamma}_c^{-1}\boldsymbol{F})]^{-1} \tag{2.8}$$

with respect to the partial order of nonnegative definite matrices. The above criterion is referred to as the small sample optimality criterion of estimating functions. A simple proof of the GT theorem is as follows. First observe that it is enough to show that

$$E_1[\boldsymbol{F}'\boldsymbol{\Gamma}_c^{-1}\boldsymbol{F}] - E_1(\boldsymbol{F}'\boldsymbol{\Gamma}_c^{-1}\boldsymbol{G}_c)[E_1(\boldsymbol{G}_c'\boldsymbol{\Gamma}_c^{-1}\boldsymbol{G}_c)]^{-1}E_1(\boldsymbol{G}_c'\boldsymbol{\Gamma}_c^{-1}\boldsymbol{F}) \ge 0, \tag{2.9}$$

i.e., nonnegative definite. The above requirement follows easily from the matrix version of the Cauchy-Schwarz inequality, namely, $\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \ge 0$ for a partitioned covariance matrix $\boldsymbol{\Sigma}$ with diagonal elements $\boldsymbol{\Sigma}_{11}$, $\boldsymbol{\Sigma}_{22}$ and off-diagonals as $\boldsymbol{\Sigma}_{12}$ and $\boldsymbol{\Sigma}_{21}$. In our case, $\boldsymbol{\Sigma}$ corresponds to the covariance of $[(\boldsymbol{F}'\boldsymbol{\Gamma}_c^{-1}\boldsymbol{g})', (\boldsymbol{G}_c'\boldsymbol{\Gamma}_c^{-1}\boldsymbol{g})']'$.

The expression (2.8) is the asymptotic covariance of the estimator $\hat{\boldsymbol{\theta}}(\boldsymbol{F})$, see equation (2.12) below and compare it with the sandwich-type finite sample covariance of $\hat{\boldsymbol{\theta}}(\boldsymbol{F})$ in (2.5) when a working transformation matrix is used for GM. For scalar θ, the optimality criterion $\boldsymbol{V}(F)$ reduces to $E_1[E_c(g^*)^2]/[E_1E_c(\partial g^*/\partial\theta)]^2$ where $g^* = F'\boldsymbol{\Gamma}_c^{-1}\boldsymbol{g}$. Except for the conditioning variables, this is same as the original criterion of Godambe (1960). For the multiparameter optimality criterion considered here, see also the important contributions of Durbin (1960), Kale (1962), and Bhapkar (1972).

An illuminating interpretation of the large sample optimality of the estimating function $\boldsymbol{G}_c'\boldsymbol{\Gamma}_c^{-1}\boldsymbol{g}$ comes from the projection approach of McLeish (1984), see also McLeish and Small (1988). If a complete parametric model is postulated, then the corresponding score vector $\varphi_\theta(\boldsymbol{y}, \boldsymbol{\theta})$ leads to asymptotically optimal maximum likelihood estimates. The optimal EF turns out to be closest to φ_θ in the conditionally linear class (in that the coefficients may depend on the conditioning variables used in E_c) generated from $\boldsymbol{g}$ under the covariance norm, because the orthogonal projection of φ_θ on $\boldsymbol{g}$ is

$$E_c(\varphi_\theta \boldsymbol{g}')\boldsymbol{\Gamma}_c^{-1}\boldsymbol{g} = -E_c(\partial \boldsymbol{g}'/\partial \boldsymbol{\theta})\boldsymbol{\Gamma}_c^{-1}\boldsymbol{g}, \tag{2.10}$$

using the fact that $\mathbf{0} = \partial(E_c\boldsymbol{g})/\partial\boldsymbol{\theta}' = E_c(\partial\boldsymbol{g}/\partial\boldsymbol{\theta}') + E_c(\boldsymbol{g}\varphi'_{\boldsymbol{\theta}})$. The asymptotic optimality of the estimator $\hat{\boldsymbol{\theta}}_{\text{MEF}}$ follows from the approximate representation (see Godambe and Heyde, 1987),

$$\boldsymbol{F}'\boldsymbol{\Gamma}_c^{-1}\boldsymbol{g}(\boldsymbol{y}, \boldsymbol{\theta}) = [E_1(\boldsymbol{F}'\boldsymbol{\Gamma}_c^{-1}\boldsymbol{G}_c)](\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + o_p(1), \tag{2.11}$$

which implies that

$$\hat{\boldsymbol{\theta}}(\boldsymbol{F}) - \boldsymbol{\theta} \to_d N_p[\mathbf{0}, \boldsymbol{V}(F)]. \tag{2.12}$$

Note that a consistent estimate of $\boldsymbol{V}(\boldsymbol{F})$ can be obtained from the expression (2.8) by dropping the E_1 operator and substituting consistent estimates for $\boldsymbol{\theta}$ if necessary. The small sample optimality of estimating function follows from the corresponding property of the score function $\varphi_{\boldsymbol{\theta}}$ (see Godambe, 1960) which seems natural in view of the projection argument. The only difference is that the general class of functions used in defining optimality of the score function is reduced to a linear class, within which the criterion $\boldsymbol{V}(F)$ is minimized. Now, the small sample optimality of the estimator $\hat{\boldsymbol{\theta}}_{\text{MEF}}$ is unknown in general. However, if we assume that $G_c \equiv -E_c(\partial\boldsymbol{g}/\partial\boldsymbol{\theta}')$ is equal to $-(\partial\boldsymbol{g}/\partial\boldsymbol{\theta}')$, i.e., without the expectation operator, then the representation (2.11) becomes exact for linear models, which in turn, implies that $\hat{\boldsymbol{\theta}}_{\text{MEF}}$ has optimality similar to but stronger than BLUE. The reason is that unlike BLUE, the linear class for MEF is larger because it allows for coefficients of the linear combination to depend on $\boldsymbol{\theta}$ as well as on the conditioning variables; see Godambe (1994) and Singh (1995) for similar results in the context of estimation for linear models with random effects.

2.3 Relation Between Conditioning and Instrumental Variables

Consider the methodology of GIVE for estimating θ-parameters of the structural model, $\boldsymbol{y} - \boldsymbol{X\theta} = \boldsymbol{\epsilon} \sim (\mathbf{0}, \boldsymbol{\Gamma})$. For simplicity, assume $\boldsymbol{\Gamma}$ does not depend on $\boldsymbol{\theta}$. It is further assumed that for $q \geq p$, the $n \times q$ matrix of instrumental variables, $\boldsymbol{W}$ (say), is well correlated with the $n \times p$ matrix of model covariates $\boldsymbol{X}$ but uncorrelated with the model error $\boldsymbol{\epsilon}$. The $n \times p$ optimal instrument matrix $\boldsymbol{W}_*$ obtained from $\boldsymbol{W}$ is given by the linear regression of the p-vector $\boldsymbol{x}$ on the q-vector $\boldsymbol{w}$. Note that the number of instruments is the same as the number of x-variables, i.e., $\boldsymbol{p}$, but the number of instrumental variables is $\boldsymbol{q}$, which is at least as large as $\boldsymbol{p}$. It is further assumed that $x-$ and w-variables are such that $plim(\boldsymbol{X}'\boldsymbol{\Gamma}^{-1}\boldsymbol{W}_*/n)$ is nonsingular. Now, the conditioning variables used in MEF can serve as w-variables, and $\boldsymbol{G}_c \equiv -E_c(\partial \boldsymbol{g}/\partial \boldsymbol{\theta}') = E_c(\boldsymbol{X})$ as w-specific optimal instruments. Here only one conditioning set A_1 is involved so that $E_c = E_2$. However, under a given semi-parametric modelling, the information may not be sufficient to compute the actual conditional expectation $E_c(\boldsymbol{X})$ for finding w-specific optimal instruments. Instead, an approximation given by the linear regression function ($\tilde{E}_c(\boldsymbol{X})$, say) may be used as the instrument matrix. With this approximation, MEF and the method of instrumental variables will be equivalent if the instrumental variables coincide with the conditioning variables.

In the next section, we present a generalization of the GT theorem to address the question of optimality when $\tilde{E}_c(\boldsymbol{X})$ is used in estimation for structural model parameters; the resulting estimates are termed SMEF estimates. We also consider the problem of finding asymptotic efficiency of a w-specific SMEF estimator. To this end, a lower bound for the asymptotic covariance is derived. The corresponding matrix of optimal instruments ($\tilde{E}_{c^*}(\boldsymbol{X})$, say) is unfortunately not obtainable in practice. Thus, $\tilde{E}_{c^*}(\boldsymbol{X})$ can be interpreted as the conceptual optimal (in the unrestricted sense) transformation matrix while $\tilde{E}_c(\boldsymbol{X})$, corresponding to specified instrumental variables, as the working (because of restricted optimality) transformation matrix. In the following, it will be assumed for simplicity that there is only one conditioning set, i.e., $A_k = A_1$, $1 \le k \le K$ in the GT framework.

3 PROPOSED METHOD OF STRUCTURAL MODEL ESTIMATING FUNCTION

We propose two SMEF estimators $\hat{\boldsymbol{\theta}}^{(1)}_{\text{SMEF}}$ and $\hat{\boldsymbol{\theta}}^{(2)}_{\text{SMEF}}$ corresponding to two choices of conditioning or instrumental variables $\boldsymbol{w}$: (1) the w-variables are prespecified but do not contain the constant 1; $\boldsymbol{w}$ in this case will be denoted by $\boldsymbol{w}_{(1)}$, and (2) the w-variables are prespecified and contain the constant 1. In each case, we establish optimality of the SMEF estimator in a suitable class in a manner analogous to that of the MEF estimator of GT. Next the quesiton of the asymptotic efficiency of the SMEF estimator is considered in Section 3.3.

3.1 The Estimator $\hat{\boldsymbol{\theta}}^{(1)}_{\text{SMEF}}$

This case gives rise to commonly used instruments. Here the linear regression function $\tilde{E}_c(\boldsymbol{X})$ passes through the origin, because the constant 1 is not one of the w-variables. It turns out that the optimality of the estimator using $\tilde{E}_c(\boldsymbol{X})$ as instruments derived from estimating functions is equivalent to the well known property of w-specific optimal instruments in GIVE methodology. To see this, let us first define the operator $\tilde{E}_c$ on the random variable $f(\boldsymbol{y}, \boldsymbol{\theta})$ with respect to a set of q variables $\boldsymbol{w}_{(1)}(q \geq p)$ as

$$\tilde{E}_c(f) = \boldsymbol{w}'_{(1)}[E_1(\boldsymbol{w}_{(1)}\boldsymbol{w}'_{(1)})]^{-1}E_1(\boldsymbol{w}_{(1)}f) := \boldsymbol{w}'_{(1)}\boldsymbol{\beta}, \tag{3.1}$$

where $\boldsymbol{\beta}$ is the q-vector of regression coefficients of f on $\boldsymbol{w}_{(1)}$. Note that since the regression function $\tilde{E}_c(f)$ is through the origin, the w-variables are not centered. The operator $\widetilde{Cov}_c(f)$ is defined as $Cov_1(f - \tilde{E}_c(f))$, and is given by

$$\begin{aligned}\widetilde{Cov}_c(f) &= E_1(f^2) - E_1[f(\tilde{E}_c(f))], \\ &= E_1(f^2) - E_1(f\boldsymbol{w}'_{(1)})[(E_1(\boldsymbol{w}_{(1)}\boldsymbol{w}'_{(1)})]^{-1}E_1(\boldsymbol{w}_{(1)}f). \end{aligned} \tag{3.2}$$

With this definition of $\tilde{E}_c$ and $\widetilde{Cov}_c$, we have for the semiparametric structural linear model, $\boldsymbol{g} = \boldsymbol{y} - \boldsymbol{X\theta} \sim (\boldsymbol{0}, \boldsymbol{\Gamma})$, $\boldsymbol{\Gamma} = \sigma^2\boldsymbol{I}$,

$$\tilde{E}_c(\boldsymbol{g}) = \boldsymbol{0}, \quad \boldsymbol{\Gamma}_c = \widetilde{Cov}_c(\boldsymbol{g}) = \boldsymbol{\Gamma}, \tag{3.3}$$

because by definition, $\boldsymbol{g}$ is uncorrelated with the conditioning variables, $\boldsymbol{w}$.

We now have the following generalization of the GT theorem. Let $\tilde{\boldsymbol{G}}_c = \tilde{E}_c(\boldsymbol{X})$ be defined elementwise by (3.1). Thus $\tilde{G}_c$ is $\boldsymbol{W}_{(1)}\boldsymbol{B}$ where $\boldsymbol{W}_{(1)}$ is the $n \times q$ matrix of observations on the q-vector $\boldsymbol{w}_{(1)}$, and $\boldsymbol{B}$ is the $q \times q$ matrix of regression coefficients $[E_1(\boldsymbol{w}_{(1)}\boldsymbol{w}'_{(1)})]^{-1}E_1(\boldsymbol{w}_{(1)}\boldsymbol{x}')$ where $\boldsymbol{x}$ is the p-vector of x-variables. Since $\boldsymbol{B}$ is unknown in general, a consistent estimate can be substituted to approximate $\tilde{\boldsymbol{G}}_c$ as

$$\boldsymbol{W}_{(1)*} := \boldsymbol{W}_{(1)}\hat{\boldsymbol{B}} = \boldsymbol{W}_{(1)}(\boldsymbol{W}'_{(1)}\boldsymbol{W}_{(1)})^{-1}\boldsymbol{W}'_{(1)}\boldsymbol{X}. \tag{3.4}$$

Therefore $\boldsymbol{W}_{(1)*}$ is the (orthogonal) projection of $\boldsymbol{X}$ on the column space of $\boldsymbol{W}_{(1)}$. Note that for general $\boldsymbol{\Gamma}$, i.e., for $\boldsymbol{\Gamma} = \sigma^2\boldsymbol{\Lambda}$, $\boldsymbol{W}_{(1)*}$ will be defined as $\boldsymbol{W}_{(1)}(\boldsymbol{W}'_{(1)}\boldsymbol{\Gamma}^{-1}\boldsymbol{W}_{(1)})^{-1}\boldsymbol{W}'_{(1)}\boldsymbol{\Gamma}^{-1}\boldsymbol{X}$ to ensure invariance of $\boldsymbol{W}_{(1)*}$ to transformations of $\boldsymbol{y} - \boldsymbol{X\theta}$, and to achieve optimal instruments (see below). Now in the linear class of estimating functions defined by $n \times p$ transformation matrices $\boldsymbol{F}$ (which may depend on the parameters $\boldsymbol{\theta}$ and are in the column space of $\boldsymbol{W}_{(1)}$) such that $\boldsymbol{F}'\boldsymbol{\Gamma}^{-1}\tilde{\boldsymbol{G}}_c$ is nonsingular, the optimal $\boldsymbol{F}$ is given by $\tilde{\boldsymbol{G}}_c$ in the sense that

$$\tilde{V}(\boldsymbol{F}) := [E_1(\boldsymbol{F}'\boldsymbol{\Gamma}^{-1}\tilde{\boldsymbol{G}}_c)]^{-1}[E_1(\boldsymbol{F}'\boldsymbol{\Gamma}^{-1}\boldsymbol{F})][E_1(\tilde{\boldsymbol{G}}_c\boldsymbol{\Gamma}^{-1}\boldsymbol{F})]^{-1} \tag{3.5}$$

is minimized. The term minimization is defined as in the case of the GT criterion (2.8). the estimator $\hat{\boldsymbol{\theta}}^{(1)}_{\text{SMEF}}$ is obtained by solving (when $\tilde{\boldsymbol{G}}_c$ is replaced by its estimate $\boldsymbol{W}_{(1)*}$)

$$\boldsymbol{W}'_{(1)*}\boldsymbol{\Gamma}^{-1}\boldsymbol{g}(\boldsymbol{y}, \boldsymbol{\theta}) = \boldsymbol{0}. \tag{3.6}$$

We have, for large samples,

$$\hat{\boldsymbol{\theta}}^{(1)}_{\text{SMEF}} - \boldsymbol{\theta} \to_d N_p[\mathbf{0}, (\boldsymbol{W}'_{(1)*}\boldsymbol{\Gamma}^{-1}\boldsymbol{W}_{(1)*})^{-1}]. \tag{3.7}$$

In fact the estimated asymptotic covariance matrix takes the sandwich form $(\boldsymbol{W}'_{(1)*}\boldsymbol{\Gamma}^{-1}\boldsymbol{X})^{-1}(\boldsymbol{W}'_{(1)*}\boldsymbol{\Gamma}^{-1}\boldsymbol{W}_{(1)*})(\boldsymbol{X}'\boldsymbol{\Gamma}^{-1}\boldsymbol{W}_{(1)*})^{-1}$ (compare with (2.5)), which can be simplified by noting that $\boldsymbol{W}'_{(1)*}\boldsymbol{\Gamma}^{-1}\boldsymbol{X}$ is the same as $\boldsymbol{W}'_{(1)*}\boldsymbol{\Gamma}^{-1}\boldsymbol{W}_{(1)*}$ using the properties of the projection form (3.4). The estimator $\hat{\boldsymbol{\theta}}^{(1)}_{\text{SMEF}}$ is asymptotically optimal in the sense that it minimizes the asymptotic covariance $\tilde{\boldsymbol{V}}(\boldsymbol{F})$. This optimality is restricted in that the class of transformation matrices $\boldsymbol{F}$ is allowed to depend only on $\boldsymbol{W}_{(1)}$. The (estimated) asymptotic covariance further simplifies to $[(\boldsymbol{X}'\boldsymbol{\Gamma}^{-1}\boldsymbol{W}_{(1)})(\boldsymbol{W}'_{(1)}\boldsymbol{\Gamma}^{-1}\boldsymbol{W}_{(1)})^{-1}(\boldsymbol{W}'_{(1)}\boldsymbol{\Gamma}^{-1}\boldsymbol{X})]^{-1}$ which can be seen to coincide with the known result for optimal instruments under GIVE methodology (Harvey, 1981, p. 80).

3.2 The Estimator $\hat{\boldsymbol{\theta}}^{(2)}_{\text{SMEF}}$

In this case the linear regression $\tilde{E}_c(\boldsymbol{X})$ is allowed to have the intercept term, and thus is more general than $\tilde{E}_c(\boldsymbol{X})$ of Section 3.1. In fact, $\tilde{E}_c(\boldsymbol{X})$ now corresponds to the commonly used definition of linear regression in which the regressors (w-variables in the present case) are centred. For a random variable $f(\boldsymbol{y}, \boldsymbol{\theta})$, $\tilde{E}_c$ is defined with respect to conditioning variables $\boldsymbol{w} = (1, \boldsymbol{w}'_{(1)})'$ as

$$\begin{aligned} \tilde{E}_c(f) &= E_1(f) + Cov_1(f, \boldsymbol{w}_{(1)})Cov_1(\boldsymbol{w}_{(1)})^{-1}[\boldsymbol{w}_{(1)} - E_1(\boldsymbol{w}_{(1)})] \\ &= E_1(f) + E_1(f\boldsymbol{w}')[E_1(\boldsymbol{w}\boldsymbol{w}')]^{-1}[\boldsymbol{w} - E_1(\boldsymbol{w})] \\ &= E_1(f\boldsymbol{w}')[E_1(\boldsymbol{w}\boldsymbol{w}')]^{-1}\boldsymbol{w} \end{aligned} \tag{3.8}$$

where $Cov_1(f, \boldsymbol{w}_{(1)})$, for example, is $E_1[(f - E_1(f))((\boldsymbol{w}_{(1)} - E_1(\boldsymbol{w}_{(1)}))']$. The operator $\widetilde{Cov}_c(f)$ is defined as before by $Cov_1(f - \tilde{E}_c(f))$ except that $\tilde{E}_c(f)$ is different in this case.

Analogous to $\hat{\boldsymbol{\theta}}^{(1)}_{\text{SMEF}}$, the optimal SMEF estimator $\hat{\boldsymbol{\theta}}^{(2)}_{\text{SMEF}}$ can be defined from (3.4) with $\boldsymbol{W}_*$ instead of $\boldsymbol{W}_{(1)*}$ and its asymptotic covariance will be $(\boldsymbol{W}'_*\boldsymbol{\Gamma}^{-1}\boldsymbol{W}_*)^{-1}$. Note that it will have stronger optimality property because of a larger linear class of estimating functions due to introduction of the unit vector in $\boldsymbol{W}$. Thus, this SMEF estimator will be superior to the usual GIVE estimator unless the constant 1 is already used as one of the instrumental variables. However, its optimality is still restricted because the class of transformation matrices is allowed to depend only on $\boldsymbol{W}$.

3.3 Asymptotic Efficiency of the SMEF Estimator

Using the framework of estimating functions, we consider the question of defining an optimal choice of conditioning variables such that they are correlated with the covariates $\boldsymbol{X}$ but uncorrelated with $\boldsymbol{y} - \boldsymbol{X\theta}$, and that they give rise to minimum asymptotic covariance in the class of all SMEF estimators. To this end, let $\boldsymbol{\theta}^0$ denote the true unknown value of $\boldsymbol{\theta}$ and define $\boldsymbol{X}_*$ as the residual of $\boldsymbol{X}$ after regressing on $\boldsymbol{y} - \boldsymbol{X\theta}^0$, i.e., for $j = 1$ to p, the jth column $\boldsymbol{x}_{Cj*}$ of $\boldsymbol{X}_*$ is (the n-vector $\boldsymbol{X}_{Cj}$ denotes the jth column of $\boldsymbol{X}$)

$$\boldsymbol{x}_{cj*} = \boldsymbol{x}_{Cj} - Cov_1(\boldsymbol{x}_{Cj}, \boldsymbol{y} - \boldsymbol{X\theta}^0)\boldsymbol{\Gamma}^{-1}(\boldsymbol{y} - \boldsymbol{X\theta}). \tag{3.9}$$

Note that if $\boldsymbol{x}_{Cj}$ is uncorrelated with $\boldsymbol{y} - \boldsymbol{X\theta}^0$, the corresponding $\boldsymbol{x}_{Cj*}$ will coincide with it. For given $\boldsymbol{\theta}^0$, the regression coefficients in (3.9) can be estimated consistently as follows. Since $\boldsymbol{\Gamma} = \boldsymbol{\sigma}_\epsilon^2\boldsymbol{I}$, it may be reasonable to assume that $Cov_1(\boldsymbol{x}_{Cj}, \boldsymbol{y} - \boldsymbol{X\theta}^0)/\sigma_\epsilon^2$ is also of the form $\gamma_j(\boldsymbol{\theta}^0)\boldsymbol{I}$ where $\gamma_j(\boldsymbol{\theta}) = Cov_1(x_{ij}, y_i - \boldsymbol{x}_i'\boldsymbol{\theta})/\sigma_\epsilon^2$ for all $i = 1$ to n. Now $\gamma_j(\boldsymbol{\theta}^0)$ can be estimated consistently as a function of $\boldsymbol{\theta}^0$ by using

$$\hat{\gamma}_j(\boldsymbol{\theta}^0) = \boldsymbol{x}_j'(\boldsymbol{y} - \boldsymbol{X\theta}^0)/n\sigma_\epsilon^2. \tag{3.10}$$

In an analogous manner σ_ϵ^2 (if unknown) can be consistently estimated as a function of $\boldsymbol{\theta}^0$. Note that for general $\boldsymbol{\Gamma}$, $\hat{\gamma}_j(\boldsymbol{\theta}^0)$ should be modified to $\boldsymbol{x}_j'\boldsymbol{\Gamma}^{-1}(\boldsymbol{y} - \boldsymbol{X\theta}^0)/n$. Thus for a given $\boldsymbol{\theta}^0$, $\boldsymbol{X}_*$ can be computed. However, since $\boldsymbol{\theta}^0$ is unknown, it is not computable in practice.

Now, treating $\boldsymbol{X}_*$ as conditioning variables, it easily follows that the optimal transformation matrix $\tilde{E}_{c*}(\boldsymbol{X})$ is $\boldsymbol{X}_*$. We will now show that the asymptotic covariance corresponding to $\boldsymbol{X}_*$ provides a lower bound for the asymptotic covariance of an SMEF estimator. To see this, note that for any given set of conditioning variables $\boldsymbol{w}$, $Cov_1(\boldsymbol{w}, y - \boldsymbol{x}'\boldsymbol{\theta}) = \boldsymbol{0}$ by definition, and therefore, $E_1(\boldsymbol{x}\boldsymbol{w}') = E_1(\boldsymbol{x}_*\boldsymbol{w}')$ where $\boldsymbol{x}_*$ is the p-vector of x_*-variables. This implies using the definition $E_c(\boldsymbol{x}) = \boldsymbol{B}'\boldsymbol{w}$ (see 3.4) that

$$E_1[\boldsymbol{x}_*\boldsymbol{x}_*'] = E_1(E_c(\boldsymbol{x})E_c(\boldsymbol{x}'))] + E_1(\boldsymbol{x}_* - E_c(\boldsymbol{x}))(\boldsymbol{x}_* - E_c(\boldsymbol{x}))']. \tag{3.11}$$

In terms of the corresponding consistent estimates, we have

$$\boldsymbol{X}_*'\boldsymbol{X}_* = \boldsymbol{W}_*'\boldsymbol{W}_* + (\boldsymbol{X}_* - \boldsymbol{W}_*)'(\boldsymbol{X}_* - \boldsymbol{W}_*). \tag{3.12}$$

Incidentally, the above decomposition also follows easily by noting that $\boldsymbol{W}_*$ is the (orthogonal) projection of $\boldsymbol{X}_*$ on $\boldsymbol{W}$ under the Euclidean norm. Now for general $\boldsymbol{\Gamma}$, (3.12) takes the form

$$\boldsymbol{X}_*'\boldsymbol{\Gamma}^{-1}\boldsymbol{X}_* = \boldsymbol{W}_*'\boldsymbol{\Gamma}^{-1}\boldsymbol{W}_* + (\boldsymbol{X}_* - \boldsymbol{W}_*)'\boldsymbol{\Gamma}^{-1}(\boldsymbol{X}_* - \boldsymbol{W}_*). \tag{3.13}$$

It follows from (3.13) that $(\boldsymbol{W}_*'\boldsymbol{\Gamma}^{-1}\boldsymbol{W}_*)^{-1} - (\boldsymbol{X}_*'\boldsymbol{\Gamma}^{-1}\boldsymbol{X}_*)^{-1}$ is nonnegative definite for any choice of w-variables and therefore $\boldsymbol{X}_*$ provides optimal instruments. Although $\boldsymbol{X}_*$ is not computable in practice as it depends on $\boldsymbol{\theta}^0$, a consistent estimate of the corresponding minimum asymptotic covariance can nevertheless be obtained by replacing $\boldsymbol{\theta}^0$ in $(\boldsymbol{X}_*'\boldsymbol{\Gamma}^{-1}\boldsymbol{X}_*)^{-1}$ by a consistent estimate given by an SMEF estimator. Note that the matrix $\boldsymbol{W}_*$ plays the role of a working transformation matrix. To see this observe that the covariance matrix $(\boldsymbol{W}_*'\boldsymbol{\Gamma}^{-1}\boldsymbol{W}_*)^{-1}$ is asymptotically equivalent to $(\boldsymbol{W}_*'\boldsymbol{\Gamma}^{-1}\boldsymbol{X}_*)^{-1}(\boldsymbol{W}_*'\boldsymbol{\Gamma}^{-1}\boldsymbol{W}_*)(\boldsymbol{X}_*'\boldsymbol{\Gamma}^{-1}\boldsymbol{W}_*)^{-1}$ in view of the comments given below (3.7) and the fact that $\boldsymbol{W}_*'\boldsymbol{\Gamma}^{-1}\boldsymbol{X}$ is asymptotically equivalent to $\boldsymbol{W}_*'\boldsymbol{\Gamma}^{-1}\boldsymbol{X}_*$.

We thus propose the following measure of asymptotic efficiency of an SMEF estimator

$$\text{asy.eff}(\hat{\boldsymbol{\theta}}_{\text{SMEF}}) = |(\boldsymbol{X}_*'\boldsymbol{\Gamma}^{-1}\boldsymbol{X}_*)^{-1}| \times |(\boldsymbol{W}_*'\boldsymbol{\Gamma}^{-1}\boldsymbol{W}_*)^{-1}|^{-1}. \tag{3.14}$$

Alternatively, in the above definition of efficiency, trace can be used instead of the determinant operator.

4 EXAMPLES OF SMEF ESTIMATORS

4.1 Latent Variable Models

Consider a linear model $\boldsymbol{y} = \boldsymbol{X\theta} + \boldsymbol{v}\alpha + \boldsymbol{\delta}$, $\boldsymbol{\delta} \sim (\boldsymbol{0}, \boldsymbol{\Gamma}_{\boldsymbol{\delta}})$, with $\boldsymbol{\Gamma} = \sigma_\delta^2\boldsymbol{I}$, $\boldsymbol{v} \sim (\boldsymbol{0}, \sigma_v^2\boldsymbol{I})$ and $\boldsymbol{\delta}$ uncorrelated with $\boldsymbol{v}$. Here $\boldsymbol{v}$ corresponds to a latent covariate and is therefore unobserved. A typical example may define y as production of a company, x as size and v as management motivation. If we treat $\boldsymbol{v}\alpha + \boldsymbol{\delta}$ as the model error $\boldsymbol{\epsilon}$ with mean $\boldsymbol{0}$ and covariance $\boldsymbol{\Gamma}$, where $\boldsymbol{\Gamma} = \sigma_\epsilon^2\boldsymbol{I}$, $\sigma_\epsilon^2 = \alpha^2\sigma_v^2 + \sigma_\delta^2$, then $\boldsymbol{X}$ and $\boldsymbol{\epsilon}$ will be correlated in general due to the presence of the covariate $\boldsymbol{v}$ in $\boldsymbol{\epsilon}$. Such problems are often addressed in longitudinal surveys where data can be obtained for a second occasion. Assuming that the parameters $\boldsymbol{\theta}$, α, and the latent variables $\boldsymbol{v}$ do not change over the two occasions, then from the model

$$\boldsymbol{y}_1 - \boldsymbol{y}_2 = (\boldsymbol{X}_1 - \boldsymbol{X}_2)\boldsymbol{\theta} + (\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2), \tag{4.1}$$

where the subscript 1, for example, denotes the first occasion, one can estimate $\boldsymbol{\theta}$ by GM provided $\boldsymbol{X}_1 - \boldsymbol{X}_2$ has full rank. (Note that if the model has an intercept term, then $\boldsymbol{X}_1 - \boldsymbol{X}_2$ will have a column of zeros which should be dropped. In other words, the intercept can't be estimated from the model (4.1)). Alternatively, we can treat $\boldsymbol{X}_1 - \boldsymbol{X}_2$ as instrumental variables because they are uncorrelated with $\boldsymbol{\epsilon}_1 = \boldsymbol{y}_1 - \boldsymbol{X}_1\boldsymbol{\theta}$ and $\boldsymbol{\epsilon}_2 = \boldsymbol{y}_2 - \boldsymbol{X}_2\boldsymbol{\theta}$, while they are correlated with $\boldsymbol{X} = (\boldsymbol{X}_1', \boldsymbol{X}_2')'$. Thus, from Section 3.1, we can find $\hat{\boldsymbol{\theta}}^{(1)}_{\text{SMEF}}$ as a solution of

$$(\boldsymbol{X}_1 - \boldsymbol{X}_2)'_* \boldsymbol{\Gamma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}) = \mathbf{0}, \tag{4.2}$$

where $\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta} = [(\boldsymbol{y}_1 - \boldsymbol{X}_1\boldsymbol{\theta})', (\boldsymbol{y}_2 - \boldsymbol{X}_2\boldsymbol{\theta})']'$, $\boldsymbol{\Gamma}$ has $\sigma_\epsilon^2 \boldsymbol{I}$ on the diagonal and $\alpha^2\sigma_v^2\boldsymbol{I}$ on the offdiagonal, and $(\boldsymbol{X}_1 - \boldsymbol{X}_2)_*$ is similar to (3.3) except that $\boldsymbol{W}_{(1)}$ is replaced by $\boldsymbol{X}_1 - \boldsymbol{X}_2$. Again it is assumed that $\boldsymbol{X}_1 - \boldsymbol{X}_2$ has full rank. If the model has an intercept term, then all the linear functions of $\boldsymbol{\theta}$ are not estimable. Note that in practice it may be difficult to estimate $\boldsymbol{\Gamma}$ and therefore a working covariance may be used to obtain suboptimal estimates. Alternatively, using the suboptimal estimates as initial consistent estimates, we can estimate $\boldsymbol{\Gamma}$ as in Zellner's feasible GLS approach and then obtain optimal estimates in a second step.

The above estimator $\hat{\boldsymbol{\theta}}^{(1)}_{\text{SMEF}}$ can be easily improved by adding a column of ones in $\boldsymbol{X}_1 - \boldsymbol{X}_2$ to obtain $\hat{\boldsymbol{\theta}}^{(2)}_{\text{SMEF}}$. Moreover, the second estimator has an additional advantage in that it allows for estimation of the intercept term if it is present in the model. Also the asymptotic efficiencies of SMEF estimators can be computed as given by (3.14).

4.2 Simultaneous Equation Models

Consider a system of two equations

$$\begin{aligned} \boldsymbol{y}_{(1)} &= \boldsymbol{X}_{(1)}\boldsymbol{\beta}_{(1)} + \boldsymbol{y}_{(2)}\alpha_{(1)} + \boldsymbol{\epsilon}_{(1)} := \boldsymbol{X}^+_{(1)}\boldsymbol{\theta}_{(1)} + \boldsymbol{\epsilon}_{(1)} \\ \boldsymbol{y}_{(2)} &= \boldsymbol{X}_{(2)}\boldsymbol{\beta}_{(2)} + \boldsymbol{y}_{(1)}\alpha_{(2)} + \boldsymbol{\epsilon}_{(2)} := \boldsymbol{X}^+_{(2)}\boldsymbol{\theta}_{(2)} + \boldsymbol{\epsilon}_{(2)} \end{aligned} \tag{4.3}$$

where $\boldsymbol{\epsilon}_{(1)} \sim (\mathbf{0}, \boldsymbol{\Gamma}_{(1)}), \boldsymbol{\epsilon}_{(2)} \sim (\mathbf{0}, \boldsymbol{\Gamma}_{(2)}), \boldsymbol{X}^+_{(1)} = (\boldsymbol{X}_{(1)}, \boldsymbol{y}_{(2)}), \boldsymbol{\theta}_{(1)} = (\boldsymbol{\beta}'_{(1)}, \alpha_{(1)})'$, and so on.

A typical example may define $y_{(1)}$ as consumption expenditure, $y_{(2)}$ as income and x's as other explanatory variables. Since the regressors and model error are obviously correlated, we can't use GM. A commonly used solution is two stage least squares in which $\boldsymbol{W} = (\boldsymbol{X}_{(1)}, \boldsymbol{X}_{(2)})$ is used as a common set of instrumental variables for each equation. Thus for (4.3a), in stage I, $\hat{\boldsymbol{y}}_{(2)}$ is obtained by regressing the $y_{(2)}$-variable on w-variables, and in stage II, parameter estimates are obtained as

$$\hat{\boldsymbol{\theta}}_{(1)} = (\boldsymbol{W}'_{(1)+}\boldsymbol{\Gamma}^{-1}_{(1)}\boldsymbol{X}^+_{(1)})^{-1}\boldsymbol{W}'_{(1)+}\boldsymbol{\Gamma}^{-1}_{(1)}\boldsymbol{y}_{(1)} \tag{4.4}$$

where $\boldsymbol{W}_{(1)+}$ denotes the optimal instruments obtained by regressing $\boldsymbol{X}^+_{(1)}$ on $\boldsymbol{W}$, i.e., $\boldsymbol{W}_{(1)+}$ is $(\boldsymbol{X}_{(1)}, \hat{\boldsymbol{y}}_{(2)})$. It is easily seen that the SMEF estimator $\hat{\boldsymbol{\theta}}^{(1)}_{\text{SMEF}}$ with $\boldsymbol{W}$ as conditioning variables will be identical to the above estimate. Moreover, it can be improved by adding the unit vector in $\boldsymbol{W}$ and its asymptotic efficiency can be calculated as discussed earlier.

We remark that all parameters of the two equations can be estimated simultaneously but it requires estimation of covariance between $\epsilon_{(1)}$ and $\epsilon_{(2)}$. For this reason, the method of three stage least squares is commonly used. Using SMEF, alternative estimators can be developed as in the case of two stage least squares.

4.3 Models with Measurement Error

Consider the model $\boldsymbol{y} = \boldsymbol{Z\theta} + \boldsymbol{\delta}$, $\boldsymbol{\delta} \sim (\mathbf{0}, \boldsymbol{\Gamma}_{\boldsymbol{\delta}})$, $\boldsymbol{\Gamma}_{\boldsymbol{\delta}} = \sigma_\delta^2 \boldsymbol{I}$, where $\boldsymbol{Z}$ is subject to measurement error. Thus the observed $\boldsymbol{Z}$ is $\boldsymbol{X}$, where $\boldsymbol{X} = \boldsymbol{Z} + \boldsymbol{U}$, $\boldsymbol{U}$ is the measurement error which is assumed to be uncorrelated with the model error $\boldsymbol{\delta}$. For the jth column $\boldsymbol{u}_{Cj}$ of $\boldsymbol{U}$, it is assumed that $\boldsymbol{u}_{Cj} \sim (\mathbf{0}, \sigma_{uj}^2 \boldsymbol{I})$ and uncorrelated over j, $1 \le j \le p$. A typical example may define $\boldsymbol{y}$ as corn yield and $\boldsymbol{Z}$ as nitrogen content in the soil as considered by Fuller (1987, p. 2). Now rewriting the model as $\boldsymbol{y} = \boldsymbol{X\theta} + \boldsymbol{\epsilon}$ where $\boldsymbol{\epsilon} = (\boldsymbol{\delta} - \boldsymbol{U\theta})$, we have a structural linear model with $\boldsymbol{X}$ correlated with $\boldsymbol{\epsilon}$. The covariance $\boldsymbol{\Gamma}$ of $\boldsymbol{\epsilon}$ will depend on unknown parameters σ_δ^2 and $\boldsymbol{\theta}$ and has the form $\boldsymbol{\Gamma} = \sigma_\epsilon^2 \boldsymbol{I}$ where $\sigma_\epsilon^2 = \sigma_\delta^2 + \Sigma_j \sigma_{uj}^2 \theta_j^2$. Often an instrumental variable is obtained by taking a second independent observation on $\boldsymbol{Z}$, see e.g. Fuller (1987, p. 52). Denoting the first and second observations on $\boldsymbol{Z}$ as $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, $\boldsymbol{\theta}$ can be estimated from

$$\boldsymbol{X}_2' \boldsymbol{\Gamma}^{-1} (\boldsymbol{y} - \boldsymbol{X}_1 \boldsymbol{\theta}) = \mathbf{0}. \tag{4.5}$$

Similarly, using $\boldsymbol{X}_1$ as the instrumental variable, another estimate can be obtained, and then a final estimate can be obtained by combining the two. Above estimates can be alternatively obtained as $\hat{\boldsymbol{\theta}}^{(1)}_{\text{SMEF}}$, and can be further improved by using $\hat{\boldsymbol{\theta}}^{(2)}_{\text{SMEF}}$.

5 CONCLUDING REMARKS

For structural linear models, two alternative SMEF estimators using instrumental variables were considered based on a generalization of the GT theory of estimating functions. The first SMEF estimator uses the available instrumental variables but does not include constant as an instrumental variable. The second SMEF estimator includes constant as an instrumental variable and was shown to be more efficient than the first SMEF estimator. These estimators correspond to the commonly used estimators based on GIVE methodology. However, the advantage of using constant as an instrumental variable does not seem to have been emphasized in the literature on GIVE methodology. Using the theory of estimating functions, a measure of asymptotic efficiency was proposed by finding a lower bound to asymptotic covariance of all GIVE estimators. Such a measure is expected to be useful in practice as optimality of each GIVE is restricted to its own class. Through several examples drawn mostly from econometrics, it was shown that the SMEF methodology provides a useful generalization of the GM methodology as well as an important unified statistical technique for dealing with linear models with stochastic regressors.

ACKNOWLEDGEMENT

We are grateful to Professors V. P. Godambe, M. E. Thompson, and Dr. H. J. Mantel for comments and helpful discussions. The first author's research was supported in part by a grant from Natural Sciences and Engineering Research Council of Canada. He is especially grateful to Sai Baba, Chancellor, Sri Sathya Sai Institute of Higher Learning, for providing the opportunity to visit the institute and conduct this research.

References

Bhapkar, V. P. (1972). On a measure of efficiency of an estimating equation. *Sankhya A*, 34, 467-72.

Durbin, J. (1960). Estimation of parameters in time series regression models. *J. Roy. Stat. Soc. Ser. B*, 22, 139-53.

Fuller, W. A. (1987). *Measurement Error Models.* John Wiley and Sons, New York.

Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *Ann. Math. Stat.* 31, 1208-12.

Godambe, V. P. (1994). Linear Bayes and optimal estimation. *STAT-94-11*, Department of Statistics and Actuarial Sciences, University of Waterloo, Waterloo, 16 pages.

Godambe, V. P. and Heyde, C. C. (1987). Quasi-likelihood and optimal estimation. *Int. Stat. Rev.*, 55, 231-44.

Godambe, V. P. and Kale, B. K. (1991). Estimating functions: an overview. In *Estimating Functions* (ed. Godambe, V. P.), Oxford University Press, Oxford, 3-20.

Godambe, V. P. and Thompson, M. E. (1989). An extension of quasi-likelihood estimation (with discussion). *J. Stat. Plan. Inf.*, 22, 137-72.

Harvey, A. C. (1981). *The Econometric Analysis of Time Series.* Phillip Allan Publishers Ltd., Oxford.

Kale, B. K. (1962). An extension of Cramér-Rao inequality for statistical estimation functions. *Skand. Aktur.*, 45, 60-89.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13-22.

McLeish, D. L. (1984). Estimation for aggregate models: the aggregate Markov chain. *Can. J. Stat.*, 12, 265-282.

McLeish, D. L. and Small, C. G. (1988). The Theory and Applications of Statistical Inference Functions. *Lecture Notes in Statistics* 44, Springer-Verlag, New York.

Singh, A. C. (1995). Predicting functions for generalization of BLUP to mixed nonlinear models. *ASA Proc. Biom. Sec.*, 300-5.