Statistics Canada / Statistique Canada

Methodology Branch

Household Survey
Methods Division

Direction de la méthodologie

Division des méthodes
d'enquêtes des ménages

Canada

WORKING PAPER
METHODOLOGY BRANCH

GMM ENHANCEMENTS IN THE PRESENCE OF EXTRA INFORMATION ABOUT THE COVARIANCE OF THE MOMENT CONDITION GENERATING ZERO FUNCTIONS

HSMD-97-007E

A.C. Singh and H.J. Mantel

Household Survey Methods Division
Statistics Canada

December 1997

The work presented in this paper is the responsibility of the author and does not necessarily represent the views or policies of Statistics Canada

GMM ENHANCEMENTS IN THE PRESENCE OF EXTRA INFORMATION ABOUT THE COVARIANCE OF THE MOMENT CONDITION GENERATING ZERO FUNCTIONS

A.C. Singh and H.J. Mantel[1]

ABSTRACT

For the problem of estimating (first order) parameters of a semiparametric model, it is shown that the optimality of a commonly used method known as the generalized method of moments (GMM) developed in the econometrics literature by Hansen (1982) can be justified by the optimality of the method of estimating functions (MEF) developed in the statistics literature by Godambe (1960) and Godambe and Thompson (1989). Although the GMM framework has several appealing features, it does not always lead to the optimal estimator in a general sense as obtained by MEF when extra information about covariance of moment condition generating zero functions becomes available. In this paper, with the aid of MEF, we consider enhancements to the GMM framework so that it does give rise to an optimal estimator (asymptotically equivalent to MEF) when extra covariance information is present, while preserving the essential features of GMM. Moreover, for testing model fit a new test is proposed which recovers degrees of freedom lost (due to estimation of model parameters) with the usual GMM test. The new test is expected to be sensitive to model departures in directions in which the difference between the old and new estimating functions is not zero in expectation. A simple example along with simulation results is presented to compare both old and new methods in the context of estimation and testing.

Key Words: Estimating Functions; Minimum chi-square estimation; Moment conditions; Test of model fit.

---

[1] A.C. Singh and Harold Mantel, Household Survey Methods Division, Statistics Canada, 16th floor, R.H. Coats Building, Ottawa, Ontario, K1A 0T6.

AMÉLIORATION DE LA MÉTHODE GÉNÉRALISÉE DES MOMENTS EN PRÉSENCE DE DONNÉES SUPPLÉMENTAIRES SUR LA COVARIANCE DES FONCTIONS ZÉRO GÉNÉRATRICES DE MOMENTS

A.C. Singh et H.J. Mantel[2]

RÉSUMÉ

Concernant le problème de l'estimation des paramètres (de premier ordre) d'un modèle semi-paramétrique, on montre que l'optimalité d'une méthode courante connue sous le nom de méthode généralisée des moments (MGM) élaborée dans les travaux d'économétrie par Hansen (1982) peut être justifiée par l'optimalité de la méthode des fonctions d'estimation (MFE) élaborée dans la littérature statistique par Godambe (1960) et par Godambe et Thompson (1989). Le cadre général de la MGM a plusieurs aspects intéressants, mais il ne mène pas toujours à l'estimateur optimal dans un sens général - comme c'est le cas avec la MFE - en présence de données supplémentaires sur la covariance des conditions des moments générant des fonctions nulles. Dans le présent article, à l'aide de la MFE, nous considérons les améliorations apportées au cadre général de la MGM de sorte qu'il permette d'arriver à un estimateur optimal (asymptotiquement équivalant à la MFE) quand on dispose de données supplémentaires sur la covariance, tout en préservant les aspects essentiels de la MGM. En outre, pour tester le modèle d'ajustement, on propose un nouveau test qui récupère les degrés de liberté perdus (en raison de l'estimation des paramètres de modèle) avec le test de la MGM habituel. Le nouveau test devrait être sensible aux départs de modèle dans des directions où la différence entre les fonctions d'estimation anciennes et nouvelles n'est pas nulle en espérance. Un exemple simple est présenté avec les résultats de la simulation pour comparer les anciennes et les nouvelles méthodes dans le contexte des tests et de l'estimation.

Mots clés : fonctions d'estimation, estimation du chi carré minimum, conditions des moments, test d'ajustement de modèle.

[2] A.C. Singh et Harold Mantel, Division des méthodes d'enquêtes des ménages, Statistique Canada, 16ième étage, Immeuble R.H. Coats, Ottawa, Ontario, K1A 0T6.

1. INTRODUCTION

In econometrics, a popular and intriguing method of estimation for semiparametric model parameters is the generalized method of moments (GMM) introduced by Hansen (1982). GMM uses a minimum χ^2-type criterion as an optimality criterion which is a scalar function whose gradient gives an estimating function that can be solved to obtain the GMM estimator. In GMM we start with a given set of (linearly independent) moment conditions, chosen from substantive considerations, which are generated as linear combinations of elementary zero functions and have approximate normal distributions. Note that zero functions are simply functions of data and parameters with zero mean, and thus moment conditions are themselves zero functions, but not elementary. Next a χ^2-type statistic is constructed and then minimized to obtain the estimator. GMM is optimal in the linear class of estimating functions generated by the given set of moment conditions. However, since the choice of moment conditions is not governed by optimality considerations, the GMM estimator is optimal only in a restricted sense.

In statistics, on the other hand, an important and general method of estimation, known as the method of estimating functions (MEF), was introduced by Godambe (1960) and Godambe and Thompson (1989). The optimality criterion for MEF is an estimating-function-driven criterion, unlike the criterion for GMM which is estimator-driven. In MEF we choose an estimating function that minimises the MEF criterion; however, the GMM estimator directly minimises the GMM criterion. Thus the MEF estimator of a parameter is defined indirectly as the solution of the optimal estimating function.

In MEF we start with a given set of elementary zero functions and their covariance structure. Then the optimal estimating function is the best linear combination of the elementary zero functions. If the GMM moment conditions are substituted for the elementary zero functions in the MEF framework, then the GMM and MEF estimators can be seen to be asymptotically equivalent. In other words, MEF theory can be used to justify GMM.

We remark that the GMM framework has several appealing features, namely, (i) it does not have the problem of choosing among possibly multiple roots of the estimating equation, (ii) it does provide a model specification test without explicit specification of test parameters, provided that the number of moment conditions exceeds the number of model parameters, (iii) it does not require functional specification of the covariance of moment conditions, and (iv) it does provide an estimator which enjoys optimality in a suitably restricted class. Now suppose extra information about the covariance of the moment condition generating zero functions is available. Then it is known that MEF (with moment condition generating zero functions as input for elementary zero functions) provides an optimal estimator in a general class. Thus the GMM estimator developed in the absence of extra covariance information may become sub-optimal.

In this paper we consider enhancements to the GMM framework so that it indeed gives rise to an optimal estimator (asymptotically equivalent to MEF) when extra covariance information is present, while preserving the main features of the GMM framework. Apparently the existing GMM literature does not offer any guidance in this respect. Note that the trivial solution of using the MEF optimal estimating function as the sole set of moment conditions in the GMM framework is not adequate as it will satisfy only the optimality requirement but not the other GMM features, since the number of moment conditions becomes exactly equal to the number of parameters. It is shown that if the moment conditions originally chosen in the absence of extra covariance information are augmented (and not replaced) by the MEF optimal estimating function, then the GMM methodology provides a straight-forward solution to the problem mentioned above. The resulting solution, termed the minimum chi-square estimating function (MCEF) method, combines the strengths of both the GMM and MEF methodologies.

In MCEF, as expected, the original set of moment conditions, in the presence of the optimal estimating function, do not provide any new information for the estimation problem; however, they are useful in the estimation problem for choosing among possibly multiple roots, and for the problem of testing model specification. The MCEF estimator, like GMM, will depend on the choice of moment conditions, but it turns out that regardless of the choice of moment conditions, it remains asymptotically equivalent to MEF.

Using MCEF we take another look at the problem of testing fit of the semiparametric model, *i.e.* testing the hypothesis that the moment conditions have zero expectation. In the GMM literature this is referred to as testing for overidentifying restrictions. When the number of linearly independent moment conditions, m, is larger than p, the dimension of the model parameter θ, they define implicitly the test parameters and hence the direction of the alternative hypothesis. We introduce an alternative form of the GMM χ^2-test which remains valid when the MCEF estimator (or any other root-n consistent estimator) of θ is substituted. This test is asymptotically equivalent to the GMM test, and is in fact identical when the GMM estimator of θ is used. The degrees of freedom of the χ^2-test is just the number of (linearly independent) moment conditions minus the dimension of θ, that is m-p.

If a consistent estimator of θ, other than the GMM estimator, is substituted in the GMM χ^2-statistic then the asymptotic distribution will not be χ^2_{m-p}; instead it will be χ^2_{m-p} plus a convex linear combination of p independent χ^2_1 variates. This is similar to the result of Chernoff and Lehmann (1954).

We next develop a new test using MCEF which recovers the degrees of freedom lost due to estimation of θ. The new test essentially adds a correction term to the χ^2-test so that it is asymptotically χ^2_m. The new test is expected to be powerful against alternatives in the direction where the difference between GMM and MCEF estimating functions is not zero in expectation.

Section 2 contains a brief review of GMM and MEF as well as a justification of GMM *via* MEF. Section 3 presents the proposed enhancement of the GMM method, *i.e.* MCEF, and also the new tests. A simple example and some simulation results are given in Section 4.

2. GMM AND MEF: A REVIEW

2.1. GMM. Consider a semiparametric model specified by an m-vector of (linearly independent) moment conditions $\varphi(y,\theta)$ and its (nonsingular) covariance matrix $V_{\varphi}(\theta)$, where y is an n-vector of observations and θ is a p-vector of model parameters. The jth moment condition $\varphi_j(y,\theta)$ is a sum of n elementary moment (or orthogonality) conditions $\psi_j(y_i,\theta)$:

$$\varphi_j(y,\theta) = \textstyle\sum_{i=1}^{n} \psi_j(y_i,\theta). \tag{2.1}$$

The observations y_i may be correlated in general. In the GMM framework, the functional form of the covariance matrix of the vector of ψ-functions is not assumed to be known. The matrix $V_{\varphi}(\theta)$ is replaced by a consistent estimate (for large n) using sample moments of the ψ-functions. Thus GMM requires very few assumptions and can be applied in various practical situations. As mentioned in the introduction, it has several appealing features; see Davidson and MacKinnon (1993, ch. 17) for a good review. It should be noted that the moment conditions could also have the property of being conditional zero functions. Now the GMM estimator, $\hat{\theta}_{GMM}$, is obtained by minimizing the χ^2-criterion

$$\chi^2(\theta) = \varphi(y,\theta)^{T} V_{\varphi}(\hat{\theta}_0)^{-1} \varphi(y,\theta) \tag{2.2}$$

where $V_{\varphi}(\hat{\theta}_0)$ is a consistent estimator of $V_{\varphi}(\theta)$ obtained from ψ-functions evaluated at $\hat{\theta}_0$, an initial consistent estimate of θ. The estimator $\hat{\theta}_0$ can be obtained by initially setting $V_{\varphi} = I$ in (2.2); this can be updated iteratively. The GMM estimating function is then

$$f(y,\theta) = (\partial\varphi/\partial\theta)^{T} V_{\varphi}^{-1}(\hat{\theta}_0)\, \varphi(y,\theta) \tag{2.3}$$

ad the estimating equation is $f(y,\theta)=0$.

The χ^2-criterion in (2.2) is an estimator-driven criterion, in the sense that the estimator is chosen to minimize this criterion. Under very general conditions the minimum of $\chi^2(\theta)$ exists and is unique. Furthermore, $\hat{\theta}_{GMM}$ is essentially invariant to scale transformations of the φ. To see this, note that if we replace φ by $A(\theta)\varphi$ then (2.3) becomes

$$\left(A(\theta)\frac{\partial\varphi}{\partial\theta} + \frac{\partial A(\theta)}{\partial\theta}\varphi \right)^{T} \left(A(\hat{\theta}_0) V_{\varphi}(\hat{\theta}_0) A^{T}(\hat{\theta}_0) \right)^{-1} A(\theta)\varphi \tag{2.4}$$

and this is asymptotically equivalent to (2.3) since φ/n and $\hat{\theta}_0 - \theta$ are $o_p(1)$. Exact invariance can be achieved if $\hat{\theta}_0$ in (2.2) is replaced by θ, which is referred to as the continuous updating method in the GMM literature.

Assuming that the minimum of $\chi^2(\theta)$ is unique, the consistency of $\hat{\theta}_{GMM}$ follows from (2.3) since the estimating function is $o_p(n)$ by a suitable law of large numbers. Furthermore, from the asymptotic normality of φ under a central limit theorem, we get

$$\hat{\theta}_{\mathrm{GMM}} - \theta \sim N\left(0, \left\{\mathrm{E}(\partial\varphi/\partial\theta)^{\mathrm{T}}\mathrm{V}_{\varphi}^{-1}(\theta)\mathrm{E}(\partial\varphi/\partial\theta)\right\}^{-1}\right) \tag{2.5}$$

where, in the asymptotic covariance, E can be dropped for estimation purposes.

The asymptotic optimality of $\hat{\theta}_{\mathrm{GMM}}$ is analogous to that of MLE by noting that the χ^2-criterion (2.2) is the kernel of the approximate likelihood of θ obtained from the asymptotic normality of φ under local alternatives. The optimality, however, is in a restricted class because the construction of moment conditions (φ) from the elementary moment conditions (ψ) is not based on optimality considerations but is simply obtained as sample moments of ψ-functions. Now, an asymptotically optimal test of the semiparametric model, $\mathrm{H}_0 : \mathrm{E}_\theta[\psi(y,\theta)] = 0$, can be obtained as a score test (*cf.* Cox and Hinkley, 1974, ch. 9), which rejects H_0 if

$$Q(\varphi)_{\theta=\hat{\theta}_{\mathrm{GMM}}} \geq \chi^2_{m-p,\alpha} \tag{2.6}$$

where $\chi^2_{m-p,\alpha}$ is the upper α-point of the χ^2_{m-p} distribution, and $Q(u)$ denotes a quadratic form $u^{\mathrm{T}}\Sigma^{-1}u$, Σ being the covariance of a random vector u; if Σ is singular, then a g-inverse is used. Note that since the number m of moment conditions is larger than the dimension p of the model parameter θ, the m-p test parameters (ξ, say) under H_0 are specified implicitly. This feature is convenient in practice because it may be difficult in general to specify ξ-parameters in a closed form.

Given that the model is accepted, a confidence set for θ can be constructed as

$$\{\theta : Q(\varphi) - Q(\varphi)_{\theta=\hat{\theta}_{\mathrm{GMM}}} \leq \chi^2_{p,\alpha}\} \tag{2.7}$$

2.2. MEF. In MEF the semiparametric model is specified by an n-vector h of elementary zero functions and its (nonsingular) covariance matrix $\mathrm{V}_h(\theta)$.

The optimal estimating function g is Ah where the $p \times n$ matrix $A(\theta)$ minimizes (in the partial order of non-negative definite matrices)

$$\left[\mathrm{E}(\partial Ah/\partial\theta)^{\mathrm{T}}\mathrm{V}_{Ah}^{-1}(\theta)\,\mathrm{E}(\partial Ah/\partial\theta)\right]^{-1} \tag{2.8}$$

The above criterion is called the Godambe Information criterion in analogy with the relation of the score function and the Fisher Information matrix. Unlike GMM, the above optimization criterion is estimating-function-driven since optimization is achieved by an estimating function.

The score-type optimal estimating function g is given by

$$g(y,\theta) = -\mathrm{E}(\partial h/\partial\theta)^{\mathrm{T}}\mathrm{V}_h^{-1}(\theta)\,h(y,\theta). \tag{2.9}$$

More generally h could be a conditional zero function, in which case both the variance V in (2.7) and the optimal A would be conditional, see *e.g.* Singh and Rao (1998). A solution of the optimal estimating equation $g(y,\theta) = 0$ exists in general but it need not be unique; a method to handle the problem of choosing among multiple roots is developed in Singh and Mantel (1998). Under mild regularity conditions the equation $g(y,\theta) = 0$ has a consistent and asymptotically normal solution, $\hat{\theta}_{\mathrm{MEF}}$. Its asymptotic distribution is given by

$$\hat{\theta}_{\mathrm{MEF}} - \theta \sim N\left(0, \left\{\mathrm{E}(\partial h/\partial\theta)^{\mathrm{T}} \mathrm{V}_h^{-1}(\theta)\mathrm{E}(\partial h/\partial\theta)\right\}^{-1}\right) \tag{2.10}$$

The asymptotic covariance matrix of $\hat{\theta}_{\mathrm{MEF}}$ is smallest in the class of all estimators derived from linear estimating functions of the form Ah. This is a natural analogue of the MLE optimality criterion when the class is restricted.

To test model fit, the model is first extended to specify explicitly a set of r test parameters ξ ($=\xi_0$ under H_0). Then the optimal estimating functions $g_\theta(y,\theta,\xi)$ and $g_\xi(y,\theta,\xi)$ for (θ,ξ) are found. An optimal test for H_0 is given by

$$Q(g_\theta, g_\xi)_{\theta=\hat{\theta}_{\mathrm{MEF}},\, \xi=\xi_0} \geq \chi^2_{r,\alpha} \tag{2.11}$$

where $Q(u_1, u_2)$ is defined as $Q(u)$ with u denoting the stacked column vector of (u_1, u_2). This can be interpreted as a score-type test by considering the asymptotic likelihood of g_θ and g_ξ under local alternatives to H_0. If the model fits then a confidence set for θ is obtained as

$$\{\theta : Q(g_\theta)_{\xi=\xi_0} \leq \chi^2_{p,\alpha}\}. \tag{2.12}$$

2.3. Comparison of GMM and MEF. The main difference between MEF and GMM is that the elementary zero functions used in MEF could be more elementary than the elementary moment conditions used in GMM. In other words, elementary moment conditions can be generated as linear combinations of the underlying elementrary zero functions. Thus the optimality of the MEF can be more general than that of the GMM, depending on the level at which the elementary zero functions are defined, which depends on the level at which the covariance structure is available.

If the moment conditions used in GMM are treated as elementary zero functions then MEF and GMM are asymptotically equivalent. To see this, note that $(\partial\varphi/\partial\theta)/n - \mathrm{E}(\partial\psi/\partial\theta) = o_p(1)$ by a weak law of large numbers, and thus iteratively evaluating the GMM estimating function $f(y,\theta)$ in (2.3) is asymptotically equivalent to solving the MEF estimating function $g(y,\theta)$ in (2.9) when $h=\varphi$. However, if we take the moment condition generating zero functions as elementary zero functions, assuming that the extra covariance information is present, then GMM can be improved *via* MEF since the MEF optimality is in a wider class of estimating functions.

Thus MEF provides a general and flexible method of constructing optimal estimating functions, whose optimality depends on the choice of elementary zero functions and their covariance structure used as input. This is the main strength of MEF, and can be used to strengthen the GMM.

The GMM approach, on the other hand, can be used to overcome some of the problems of MEF. First of all, GMM provides a readily available test of model fit in the χ^2-statistic (2.2) evaluated at $\hat{\theta}_{\mathrm{GMM}}$ which has an asymptotic χ^2_{m-p} distribution under the null hypothesis that $\mathrm{E}(\psi)=0$. Thus in the GMM approach it is not necessary to explicitly define the ξ-parameters to construct a test of model fit. Secondly, the GMM estimation, involving the minimization of a χ^2-statistic, does not have the potential problem of choosing among multiple roots that the MEF estimation, which involves solving an estimating equation, has.

The minimum χ^2 estimating function (MCEF) method that we propose here modifies GMM so that the GMM estimator becomes asymptotically equivalent to the MEF estimator while preserving the essential GMM framework with its corresponding benefits.

3. MCEF: THE PROPOSED GMM ENHANCEMENT

Suppose extra information about the covariance matrix $V_h(\theta)$ of moment condition generating zero functions h is available. Let $g(y,\theta)$ denote the optimal estimating function which gives rise to $\hat{\theta}_{MEF}$ as the (unrestricted) optimal estimator. Now consider the semiparametric model in which φ is augmented with g, *i.e.*,

$$\varphi_\bullet := \begin{bmatrix} \varphi \\ g \end{bmatrix} \sim (0, V_{\varphi_\bullet}(\theta)). \tag{3.1}$$

The proposed GMM enhancement consists of applying the GMM methodology to the model (3.1).

3.1. Estimation. The proposed estimator, $\hat{\theta}_{MCEF}$, is defined by minimizing

$$\chi^2_\bullet(\theta) = \varphi^T_\bullet(\theta)\, V^-_{\varphi_\bullet}(\hat{\theta}_0)\, \varphi_\bullet(\theta) \tag{3.2}$$

where $\hat{\theta}_0$ is an initial consistent estimator of θ. Note that in (3.2), we have used a generalized inverse of the matrix $V_{\varphi_\bullet}(\hat{\theta}_0)$ because it need not be of full rank. In the following we assume, for simplicity, that $V_{\varphi_\bullet}(\hat{\theta}_0)$ has full rank, *i.e.* $m+p$. The estimating function is given by

$$f_\bullet(y,\theta) = (\partial\varphi^T_\bullet/\partial\theta) V^{-1}_{\varphi_\bullet} \varphi_\bullet(\theta). \tag{3.3}$$

The estimator $\hat{\theta}_{MCEF}$ is not only asymptotically equivalent to $\hat{\theta}_{MEF}$, but it shares all the important features of $\hat{\theta}_{GMM}$. To show the equivalence, observe that we can write $f_\bullet(y,\theta)$ as (here write $\varphi = M^T h$, $-E(\partial h/\partial\theta) = G$),

$$\begin{pmatrix} M^T(\partial h/\partial\theta) \\ G^T V_h^{-1}(\partial h/\partial\theta) \end{pmatrix}^T \begin{pmatrix} M^T V_h M & M^T G \\ G^T M & G^T V_h^{-1} G \end{pmatrix}^{-1} \begin{pmatrix} M^T h \\ G^T V_h^{-1} h \end{pmatrix}$$

$$\approx -\begin{pmatrix} M^T G \\ G^T V_h^{-1} G \end{pmatrix}^T \begin{pmatrix} M^T V_h M & M^T G \\ G^T M & G^T V_h^{-1} G \end{pmatrix}^{-1} \begin{pmatrix} M^T h \\ G^T V_h^{-1} h \end{pmatrix} \tag{3.4}$$

using a law of large numbers for linear combinations of $\partial h/\partial\theta^T$. Now since $\varphi_\bullet$ has expected value 0, it belongs with probability 1 to the column space of its covariance matrix. Noting also that the left most matrix of (3.4) is equal to a part of that same covariance matrix, we see that (3.4) simplifies to $G^{\mathrm{T}}\mathrm{V}_h^{-1}h$ which is the MEF estimating function $g(y,\theta)$. It follows that $\hat{\theta}_{\mathrm{MCEF}}$ is asymptotically equivalent to $\hat{\theta}_{\mathrm{MEF}}$. Moreover, one can construct several $\hat{\theta}_{\mathrm{MCEF}}$, depending on the choice of M in $\varphi = M^{\mathrm{T}}h$, which will all be asymptotically equivalent to $\hat{\theta}_{\mathrm{MEF}}$.

3.2. Testing. First we construct a χ^2-test, which is asymptotically equivalent to χ^2_{GMM} such that any root n-consistent estimate of θ can be used in evaluating the test statistic. The construction follows easily from that of the score test. We transform the moment conditions φ as

$$C\varphi = \begin{pmatrix} B\varphi \\ f \end{pmatrix} \begin{matrix} (m-p)\times 1 \\ p\times 1 \end{matrix} \tag{3.5}$$

where C is an $m\times m$ nonsingular transformation matrix, and f is the GMM estimating function for θ. It follows that if l denotes the asymptotic loglikelihood based on φ, then f can be viewed as the score function l_θ and $B\varphi$ as l_ξ (ξ denoting the test parameters) where, *e.g.*, l_θ denotes the partial derivative of l with respect to θ. Thus, the score test for $\mathrm{H}_0\!:\mathrm{E}(\psi)=0$ is given by

$$Q(B\varphi - \tilde{\mathrm{E}}(B\varphi|f))_{\theta=\hat{\theta}_{\mathrm{MCEF}}} \geq \chi^2_{m-p,\alpha}. \tag{3.6}$$

Here $\tilde{\mathrm{E}}(B\varphi|f)$, similar to the conditional expectation, is the best linear predictor of $B\varphi$ given f and is therefore given by $B\mathrm{E}(\partial\varphi/\partial\theta^{\mathrm{T}})f$. The test (3.6) can be expressed as

$$\chi^2_{\mathrm{MCEF}}(1) := (Q(\varphi) - Q(f))_{\theta=\hat{\theta}_{\mathrm{MCEF}}} \geq \chi^2_{m-p,\alpha} \tag{3.7}$$

Note that the test statistic can be evaluated at any root n-consistent estimate of θ without affecting the asymptotic optimality. When it is evaluated at $\theta = \hat{\theta}_{\mathrm{GMM}}$, we get χ^2_{GMM} because $f=0$. Thus $\chi^2_{\mathrm{MCEF}}(1)$ provides an asymptotically equivalent alternative to χ^2_{GMM}. Also, note that unlike the case of testing with MEF, the above test is easy to compute because the test parameters ξ are not required to be defined explicitly.

Now, it is possible to construct another test based on $\varphi_.$ which recovers p *d.f.* lost in the above test based on φ. This would be analogous to the χ^2 goodness-of-fit test of Rao and Robson (1974) when the raw data MLE of model parameters is used; see also Singh (1987). The reason for this is that the elementary zero functions h (like the raw data in the goodness-of-fit problem) are available for estimating θ as an alternative to the moment conditions φ (which are like the grouped data). The new test is defined in a manner similar to (3.7) as

$$\chi^2_{\mathrm{MCEF}}(2) := (Q(\varphi_.) - Q(f_.))_{\theta=\hat{\theta}_{\mathrm{MCEF}}} \geq \chi^2_{m,\alpha}. \tag{3.8}$$

Note that at $\theta = \hat{\theta}_{\mathrm{MCEF}}$ it simply reduces to $Q(\varphi_.)_{\theta=\hat{\theta}_{\mathrm{MCEF}}}$. As before $\chi^2_{\mathrm{MCEF}}(2)$ can be easily justified as a score test by using the asymptotic Guassian likelihood based on $\varphi_.$.

In view of the fact that $\chi^2_{\mathrm{MCEF}}(2)$ recovers lost *d.f.*, it seems natural to expect that $\chi^2_{\mathrm{MCEF}}(2) - \chi^2_{\mathrm{MCEF}}(1)$ should provide another χ^2 test of H_0 with p *d.f.* This is indeed the case. To see this, note that $Q(\varphi_.)$ can be written as

$$\varphi_{\cdot}^{\mathrm{T}} \mathrm{V}_{\varphi_{\cdot}}^{-1} \varphi_{\cdot} = (C_{\cdot}\varphi_{\cdot})^{\mathrm{T}} (C_{\cdot} \mathrm{V}_{\varphi_{\cdot}} C_{\cdot}^{\mathrm{T}})^{-1} (C_{\cdot}\varphi_{\cdot}) \tag{3.9}$$

for a nonsingular transformation $C_{\cdot}$ such that $C_{\cdot}\varphi_{\cdot}$ defines a stacked column vector of $B\varphi$, f, and $f_{\cdot}$. It follows that

$$Q(\varphi_{\cdot}) = Q(B\varphi - \tilde{\mathrm{E}}(B\varphi \mid f, f_{\cdot})) + Q(f, f_{\cdot}). \tag{3.10}$$

Now, since $\tilde{\mathrm{E}}(B\varphi \mid f, f_{\cdot}) \approx \tilde{\mathrm{E}}(B\varphi \mid f)$ (it is approximate because of the relation $\partial\varphi/\partial\theta \approx \mathrm{E}(\partial\varphi/\partial\theta)$), we have

$$Q(\varphi_{\cdot}) \approx Q(B\varphi - \tilde{\mathrm{E}}(B\varphi \mid f)) + Q(f, f_{\cdot}). \tag{3.11}$$

Thus, from (3.6) and (3.7),

$$\chi^2_{\mathrm{MCEF}}(1) \approx Q(\varphi_{\cdot}) - Q(f, f_{\cdot}). \tag{3.12}$$

Moreover,

$$\chi^2_{\mathrm{MCEF}}(2) - \chi^2_{\mathrm{MCEF}}(1) \approx Q(f, f_{\cdot}) - Q(f_{\cdot}) \tag{3.13}$$

which is a nonnegative difference of two χ^2 variables and hence provides yet another χ^2 test of H_0 with p *d.f.* to be denoted by $\chi^2_{\mathrm{MCEF}}(3)$. Now, using the fact that $f_{\cdot} - f$ is uncorrelated with f, we can write

$$\chi^2_{\mathrm{MCEF}}(3) = Q(f_{\cdot} - f) + Q(f) - Q(f_{\cdot}). \tag{3.14}$$

Note that the last term in (3.14) vanishes at $\theta = \hat{\theta}_{\mathrm{MCEF}}$. There is an interesting practical interpretation of the relation between χ^2_{GMM} and χ^2_{MCEF}. Observe that from (3.7) we have

$$\chi^2_{\mathrm{GMM}} \approx [Q(\varphi) - Q(f)]_{\theta = \hat{\theta}_{\mathrm{MCEF}}} = \chi^2_{\mathrm{MCEF}}(1)) \tag{3.15}$$

and from (3.8) and (3.11),

$$\begin{aligned} \chi^2_{\mathrm{MCEF}}(2) &\approx [Q(\varphi) - Q(f) + Q(f, f_{\cdot}) - Q(f_{\cdot})]_{\theta = \hat{\theta}_{\mathrm{MCEF}}} \\ &= [Q(\varphi) + Q(f_{\cdot} - f)]_{\theta = \hat{\theta}_{\mathrm{MCEF}}}. \end{aligned} \tag{3.16}$$

Thus, χ^2_{GMM} (or $\chi^2_{\mathrm{MCEF}}(1)$) subtracts a correction term $Q(f)$ from $Q(\varphi)$ in order to have a χ^2_{m-p} distribution; $Q(\varphi)_{\theta = \hat{\theta}_{\mathrm{MCEF}}}$ can be shown to have a $\chi^2_{m-p} + \Sigma_{i=1}^{p} \alpha_i \chi^2_{1i}$ $(0 < \alpha_i < 1)$ distribution, where χ^2_{1i} are independent χ^2_1-variables. (see *e.g.* Chernoff & Lehmann 1954). Similarly $\chi^2_{\mathrm{MCEF}}(2)$ adds a correction term $Q(f_{\cdot} - f)_{\theta = \hat{\theta}_{\mathrm{MCEF}}}$ to have a χ^2_m distribution. The sum of the two corrections *i.e.* $(Q(f) + Q(f_{\cdot} - f))_{\theta = \hat{\theta}_{\mathrm{MCEF}}}$ has a χ^2_p $(= \Sigma_{i=1}^{p}[\alpha_i \chi^2_{1i} + (1 - \alpha_i)\chi^2_{1i}])$ distribution, which gives rise to the test $\chi^2_{\mathrm{MCEF}}(3)$.

4. A SIMPLE EXAMPLE

We will illustrate GMM and its enhancement (*i.e.*, MCEF) by means of a simple example based on linear models. For each unit in the sample of size n, the response variable is y_i and the covariates are x_i and z_i. Now, consider two elementary moment (or orthogonality) conditions:

$$\psi_1(y_i, \theta) = x_i(y_i - x_i\theta) \tag{4.1}$$

$$\psi_2(y_i, \theta) = z_i(y_i - x_i\theta) \tag{4.2}$$

which, in turn, define the moment conditions

$$\varphi_1(y,\theta) = \Sigma_{i=1}^{n} x_i(y_i - x_i\theta) = X^{\mathrm{T}}(y - X\theta) \tag{4.3}$$

$$\varphi_2(y,\theta) = \Sigma_{i=1}^{n} z_i(y_i - x_i\theta) = Z^{\mathrm{T}}(y - X\theta) \tag{4.4}$$

The elementary zero functions which generate the moment conditions are

$$h_i(y_i, \theta) = y_i - x_i\theta \tag{4.5}$$

Suppose the extra information about the covariance of h is available, and has the functional form $\sigma^2\mathrm{diag}(x_i)$. This, in turn, specifies the covariance of moment conditions φ. The above moment conditions can be motivated by postulating the underlying model (in reality, this is not known),

$$y_i = x_i\theta + z_i\lambda + \varepsilon_i \tag{4.6}$$

where λ=0, and ε_i are uncorrelated with mean 0 and variance $\sigma^2 x_i$.

It follows that the semiparametric model for GMM is $\varphi \sim (0, \mathrm{V}_\varphi(\theta))$ where $\mathrm{V}_\varphi(\theta)$ has $(\sigma^2\Sigma_i x_i, \sigma^2\Sigma_i z_i)$ on the diagonal and $\sigma^2\Sigma_i x_i z_i$ on the offdiagonal. The semiparametric model for MEF with h as input for elementary zero functions is $h \sim (0, \mathrm{V}_h(\theta))$ with $\mathrm{V}_h(\theta) = \sigma^2\mathrm{diag}(x_i)$. Note that $\mathrm{V}_h(\theta)$ is n×n while $\mathrm{V}_\varphi(\theta)$ is only m×m (m=2 in this case). Now, the GMM optimality criterion $\chi^2(\theta)$ is

$$(X_\bullet^{\mathrm{T}}(y - X\theta))^{\mathrm{T}}(X_\bullet^{\mathrm{T}}\mathrm{V}_h(\theta)X_\bullet)^{-1}X_\bullet^{\mathrm{T}}(y - X\theta) \tag{4.7}$$

where $X_\bullet = (X, Z)$ and the GMM estimating function is

$$f(y,\theta) = -(\partial\chi^2/\partial\theta)/2 = X_\bullet^{\mathrm{T}}\Gamma(y - X\theta) \tag{4.8}$$

where $\Gamma = X_\bullet(X_\bullet^{\mathrm{T}}\mathrm{V}_h(\theta)X_\bullet)^{-1}X_\bullet^{\mathrm{T}}$. Thus

$$\hat{\theta}_{\mathrm{GMM}} = (X_\bullet^{\mathrm{T}}\Gamma X_\bullet)^{-1}X_\bullet^{\mathrm{T}}\Gamma y. \tag{4.9}$$

For MEF, choosing h=φ leads to $\hat{\theta}_{\mathrm{MEF}} = \hat{\theta}_{\mathrm{GMM}}$, although one would only get asymptotic equivalence in general. However, with h as (4.5), it is easily seen that

$$g(y,\theta) = X^{\mathrm{T}}\mathrm{V}_h^{-1}(y - X\theta),$$
$$\hat{\theta}_{\mathrm{MEF}} = (X^{\mathrm{T}}\mathrm{V}_h^{-1}X)^{-1}X^{\mathrm{T}}\mathrm{V}_h^{-1}y, \tag{4.10}$$

which is identical to the generalized least squares estimator.

Now for MCEF, we define the augmented moment conditions

$$\varphi_{\bullet}(y,\theta) = \begin{pmatrix} X_{\bullet}^{\mathrm{T}}(y - X\theta) \\ X^{\mathrm{T}}\mathrm{V}_h^{-1}(y - X\theta) \end{pmatrix} \tag{4.11}$$

with the corresponding covariance matrix $\mathrm{V}_{\varphi_{\bullet}}$. The estimating function $f_{\bullet}(y,\theta)$ turns out to be

$$\begin{aligned} &(X^{\mathrm{T}}X_{\bullet} \quad X^{\mathrm{T}}\mathrm{V}_h^{-1}X)\begin{pmatrix} X_{\bullet}^{\mathrm{T}}\mathrm{V}_hX_{\bullet} & X_{\bullet}^{\mathrm{T}}X \\ X^{\mathrm{T}}X_{\bullet} & X^{\mathrm{T}}\mathrm{V}_h^{-1}X \end{pmatrix}^{-1}\varphi_{\bullet} \\ &= (0 \;\; I)\varphi_{\bullet} = X^{\mathrm{T}}\mathrm{V}_h^{-1}(y - X\theta) \end{aligned} \tag{4.12}$$

which is the same as $g(y,\theta)$. Thus, for this simple example, $\hat{\theta}_{\mathrm{MCEF}} = \hat{\theta}_{\mathrm{MEF}}$ although in general the two would only be asymptotically equivalent. Now, with φ, f, $\varphi_{\bullet}$, and $f_{\bullet}$ defined for the above example, various χ^2-tests of Section 3 can be constructed for testing $\mathrm{H}_0: \mathrm{E}(\psi_j) = 0$, $j = 1, \ldots, m$.

For an empirical illustration of the behaviour of estimates and test procedures, we generated 50,000 Monte Carlo samples of size n (=10) using the model

$$y_i = x_i\theta + z_i\lambda + w_i\eta + \varepsilon_i\,,\;\; \varepsilon_i \sim (0, \sigma^2 x_i)\,, \tag{4.13}$$

where x_i, z_i were drawn independently from a Uniform (3,10). Two distributions for ε_i were chosen, one $N(0,\sigma^2 x_i)$ and other a location and scale transformed χ_1^2 with mean and variance $(0,\sigma^2 x_i)$. The variables w_i were generated in a somewhat different manner to demonstrate the difference between the two tests χ^2_{GMM} and $\chi^2_{\mathrm{MCEF}}(2)$. If we choose W such that

$$X^{\mathrm{T}}W = 0,\; Z^{\mathrm{T}}W = 0,\; \text{but } X^{\mathrm{T}}\mathrm{V}_h^{-1}W \neq 0, \tag{4.14}$$

then the values of test parameters $\xi = (\lambda, \eta)$ with λ=0, $\eta \neq 0$ will define alternatives $\mathrm{E}(\psi_j) \neq 0$ to H_0, such that $\mathrm{E}(\varphi_1) = \mathrm{E}(\varphi_2) = 0$. This implies that χ^2_{GMM} or $\chi^2_{\mathrm{MCEF}}(1)$ will have asymptotically no power in detecting these alternatives. However, $\chi^2_{\mathrm{MCEF}}(2)$ and $\chi^2_{\mathrm{MCEF}}(3)$ will have some power because $\mathrm{E}(f_{\bullet}) \neq 0$, $\mathrm{E}(f) = 0$ and therefore $\mathrm{E}(f_{\bullet} - f) \neq 0$. Therefore, W can be defined as

$$\left(1 - (X \;\; Z)\begin{pmatrix} X^{\mathrm{T}}X & X^{\mathrm{T}}Z \\ Z^{\mathrm{T}}X & Z^{\mathrm{T}}Z \end{pmatrix}^{-1}\begin{pmatrix} X^{\mathrm{T}}1 \\ Z^{\mathrm{T}}1 \end{pmatrix}\right) \times \mathrm{const} \tag{4.15}$$

The constant was chosen as $\sqrt{84.5}$ to make variance of w similar to that of x and z. The value of the model parameter θ was set at 1 and the variance σ^2 was chosen as 0.25. In the empirical results, σ^2 was treated as known.

Table 1 shows that confidence intervals based on $\hat{\theta}_{MCEF}$ are shorter in average length (AL) as expected. The corresponding coverage probabilities (CP) are very accurate in the case of Normal errors (in fact, the intervals are based on the Normal distribution and therefore exact in this case), and slightly conservative when the errors are χ_1^2. More interesting results appear in Table 2 on testing. The usual χ^2_{GMM} (or $\chi^2_{MCEF}(1)$) test has no power in detecting alternatives where $\lambda=0$ but $\eta\neq0$. The third test $\chi^2_{MCEF}(3)$ is powerful in detecting these alternatives, but has hardly any power against alternatives where $\eta=0$. On the other hand, the new test $\chi^2_{MCEF}(2)$ performs reasonably well for various alternatives. In practice, it may be advisable to perform both tests χ^2_{GMM} (or $\chi^2_{MCEF}(1)$) and $\chi^2_{MCEF}(3)$ (with a suitable control on the overall size) before drawing any conclusions.

Table 1. Interval Estimation, *n*=10
(Average Length and Coverage Probability)

		Normal Error		χ_1^2 Error	
1-α	Estimator	AL	CP	AL	CP
0.95	MCEF	0.244	0.951	0.244	0.957
	GMM	0.248	0.951	0.248	0.955
0.9	MCEF	0.205	0.898	0.205	0.923
	GMM	0.208	0.899	0.208	0.926

Table 2. Testing of model-fit (Empirical Level and Power), *n*=10

Test Parameter Values	Nominal Level	Normal error			χ_1^2 error		
		$\chi^2_{MCEF}(1)$	$\chi^2_{MCEF}(2)$	$\chi^2_{MCEF}(3)$	$\chi^2_{MCEF}(1)$	$\chi^2_{MCEF}(2)$	$\chi^2_{MCEF}(3)$
$\lambda=0, \eta=0$	α=0.05	.04994	.05002	.04976	.06068	.07444	.05914
	α=0.10	.10072	.09968	.09966	.09724	.11212	.09398
$\lambda=.5, \eta=0$	α=0.05	.86954	.81806	.07384	.88440	.83142	.07410
	α=0.10	.91824	.87852	.12968	.92420	.88674	.11920
$\lambda=0, \eta=.5$	α=0.05	.04990	.54328	.63232	.05978	.52724	.64680
	α=0.10	.10008	.64878	.72538	.09388	.64772	.75600
$\lambda=.5, \eta=.5$	α=0.05	.86984	.95674	.71128	.88470	.96244	.73140
	α=0.10	.91688	.97606	.79548	.92542	.97854	.81196

5. CONCLUDING REMARKS

It is well known that the optimality of the GMM estimator depends on the set of moment conditions which are chosen from substantive considerations. Unfortunately, the GMM estimator may not be optimal in a general sense although the GMM framework has several other appealing features. This may happen when extra information about moment condition generating zero functions becomes available because it leads to optimal estimating functions via MEF. To overcome this limitation, a simple recipe is suggested which consists of augmenting (and not replacing) the original set of moment conditions by the optimal estimating function. The GMM estimator with the augmented set of moment conditions (termed MCEF) provides the necessary enhancements to the usual GMM in that not only the MCEF estimator becomes asymptotically equivalent to the MEF estimator, it also preserves the essential features of the GMM framework, namely, it provides a way to choose from possibly multiple roots of the estimating equation, and also a readily available test of model-fit without having to specify test parameters, i.e. directions of alternatives. Moreover, with MCEF, one can construct a new test of model fit which recovers degrees of freedom lost due to estimation of model parameters. Limited empirical results suggest that the new test in comparison to the old one could have substantially more power in detecting certain alternatives.

ACKNOWLEDGEMENT

The authors are grateful to V.P. Godambe and M.E. Thompson of the Department of Statistics and Actuarial Science, University of Waterloo for useful comments and to R. Davidson and J. MacKinnon of the Department of Economics, Queen's University for helpful discussions. The first author's research was supported in part by an NSERC grant held at Carleton University under an adjunct research professorship.

REFERENCES

Chernoff, H. and Lehmann, E.L. (1954). The use of maximum likelihood estimates in χ^2 test for goodness of fit. *Annals of Mathematical Statistics*, 25, 579-586.

Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics*, Chapman and Hall Ltd., London.

Davidson, R. and MacKinnon, J. (1993). *Estimation and Inference in Econometrics*, Oxford University Press.

Godambe, V.P. (1960). An optimum property of regular maximum likelihood estimation. *Annals of Mathematical Statistics*, 31, 1208-1212.

Godambe, V.P. and Thompson, M.E. (1989). An extension of quasilikelihood estimation (with discussion). *J. Stat. Plan. Inf.*, 22, 137-122.

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

Rao, K.C. and Robson, D.S. (1974). A chi-square statistic for goodness-of-fit tests within the exponential family. *Communications in Statistics - Theory and Methods*, 3, 1139-1153.

Singh, A.C. (1987). On the optimality and a generalization of Rao-Robson's statistic. *Comm. Stat. - Theory and Methods*, 16, 3255-3273.

Singh A.C. and Mantel, H.J. (1998) Minimum chi-square estimating function and the problem of choosing among multiple roots. Technical Report, HSMD, Statistics Canada, Ottawa, (under preparation)

Singh, A.C. and Rao, R.P. (1998). Optimal instrumental variable estimation for linear models with stochastic regressors using estimating functions. In *Selected Proceedings of the Symposium on Estimating Functions* (ed. I.V. Basawa, V.P. Godambe, and R.L. Taylor), IMS Lecture Notes - Monograph Series, Volume 32 (to appear).