Institute of Mathematical Statistics

LECTURE NOTES — MONOGRAPH SERIES

USING GODAMBE-DURBIN ESTIMATING FUNCTIONS IN ECONOMETRICS

H. D. Vinod
Fordham University

ABSTRACT

This paper explains why Godambe-Durbin "estimating functions" (EFs) from 1960 are worthy of attention in econometrics. Godambe and Kale (1991) show the failures of Gauss-Markov and least squares and prove the small-sample superiority of EFs. There are many areas of Econometrics including unit root estimation, generalized method of moments (GMM), panel data models, etc., which can use some simplification, a little greater emphasis on finite sample properties and greater flexibility. We show why statistical inference using the EFs in conjunction with the bootstrap can be superior. For example, compared to the GMM, our EF estimates of the 'risk aversion parameter' are economically more meaningful and have shorter bootstrap confidence intervals.

Key Words: Generalized method of moments, bootstrap, confidence intervals, small sample Gauss consistent, optimal estimation

1 Introduction

The aim of this paper is to continue a dialogue between statisticians working with Godambe-Durbin estimating functions (EFs) and econometricians, apparently started by Crowder's (1986) lead article in an econometrics journal. Though Crowder proves consistency of estimates as roots of EFs, he neglects to mention (i) the "main lesson" of EF theory, and (ii) that Durbin's (1960) two-regression (TR) estimator for autoregressive distributed lag (ADL) models is an optimal EF (OptEF)

Ordinary least squares (OLS) estimators are roots of "normal equations" and maximum likelihood (ML) estimators are roots of "score equations." The "main lesson" from Godambe's EF theory is to deemphasize the estimates (roots) and focus on the underlying equations called the EFs. One considers the bias and variance of EFs themselves. Minimizing the variance of (standardized) EFs is Godambe's (1960) G-criterion. It provides optimal EF, $g^* = 0$, whose roots are the OptEF estimators. The large and recently growing EF literature, surveyed by Godambe and Kale (1991), Dunlop

(1994) and Liang and Zeger (1995), gives theory and examples of successful applications in biostatistics, survey sampling and elsewhere. They show that EFs can offer *distinct and improved* estimates, both with and without normality. It is remarkable that whenever the OLS or ML estimators do not coincide with the OptEF, it is the OptEF that have superior properties in both large and small samples.

Rather than another survey, this paper indicates avenues for further research and applications in econometrics by focusing on the regression problem. Section 2 discusses EFs for multiple regression. Section 3 reviews EFs for the GMM and ML, and includes our "results." Section 4 has further results on Instrumental variables and specific suggestions for achieving small-sample optimality in GMM. Section 5 refers to the EFs in further settings, mostly those used in biostatistics. Section 6 discusses statistical testing and inference with new bootstraps based on EF pivots. Section 7 contains some final remarks.

2 Estimating Functions And Multiple Regression

Consider the usual regression model with T observations and p regressors:

$$y = X\beta + \epsilon, \quad E(\epsilon) = 0, \quad E(\epsilon\epsilon') = \sigma^2\Omega \tag{1}$$

If the Ω matrix is known, one uses the generalized least squares (GLS) estimator, whose "normal equations" are viewed here as EFs and denoted by:

$$g_{gls} = X'\Omega^{-1}X\beta - X'\Omega^{-1}y = X'\Omega^{-1}\epsilon = 0. \tag{2}$$

The log likelihood function L_T when $\epsilon \sim N(0, \sigma^2 I)$ (normal errors and $\Omega = I$) is:

$$L_T = (-T/2)log2\pi - (T/2)log\sigma^2 - (1/2\sigma^2)(y - X\beta)'(y - X\beta). \tag{3}$$

Let S_T denote the p×1 score vector of partial derivatives of L_T with respect to (wrt) β.

$$\partial L_T/\partial\beta = S_T = (1/\sigma^2)X'(y - X\beta) = (1/\sigma^2)X'\epsilon = 0. \tag{4}$$

Under the above assumptions the score equation $S_T = 0$ is OptEF$g^* = 0$, since it minimizes Godambe's G-criterion defined for our vector case as

$$D_{g*}^{-1}E(g^*g^{*\prime})D_{g*}^{-1}, \tag{5}$$

where $D_{g*} = E\partial g_i^*/\partial\beta_j$ for i, j=1, $\cdots$, p.

The OLS normal equations from (4) are, $X'\epsilon = 0$, which are optimal when $\Omega = I$. Let us rewrite them as a sum over t:

$$g_{ols}(y, X, \beta) = \Sigma_{t=1}^{T} X_t'(y_t - X_t\beta) = 0, \tag{6}$$

where $X_t=(x_{t1}, \cdots, x_{tp})$ is a row vector and $X_t'X_t$ (for $t=1,\cdots,T$) are $p\times p$ matrices having p equations for each t.

Remark 1: The EFs (as vector random variables) are sums of T terms, but the corresponding (estimators) roots need not be sums. Hence the "central limit theorem" arguments apply directly to EFs, supporting Small and Mcleish's (1988) claim that EFs are "more normal." This justifies what we called the "main lesson" of Godambe's EF theory, asking us to focus on the EFs while deemphasizing the roots.

Using the Cauchy-Schwartz inequality, Kendall and Stuart (1979, sec. 17.17) prove that the Cramer-Rao lower bound on the variance of an unbiased estimator $\hat{y}$ of $\mu = X\beta$ is attained, if and only if, β is chosen to satisfy $S_T = A(X, \beta, \sigma^2)(y - X\beta)$ where A is the arbitrary constant (matrix) of proportionality. We refer to this important small sample property as "attaining Cramer-Rao," often readily satisfied by the EFs. Durbin (1960) proved that his simple TR estimator is OptEF for the ADL model, as it "attains Cramer-Rao," even if the autoregressive (AR) parameter has a *unit root*, or y_t is *explosive* with the root > 1. Hendry (1995, p.232) lists eleven applied econometric models, including partial adjustment, equilibrium correction, leading indicator, etc., which are special cases of ADL models, but ignores Durbin's TR estimator for them. Similarly, most econometricians discuss asymptotics of complicated estimators, which often exclude unit roots or explosive cases. Since EF theorists have recently developed a better understanding of conditioning which makes Durbin's TR estimator OptEF, it deserves a fresh look in econometrics.

Gauss's intuitive notion of consistency from 1820's is explained by Sprott (1983). An estimator is said to be small-sample Gauss-consistent (SSGC) if it equals the true value, when all errors are zero. For example, solve the OptEF $X'\epsilon = 0$ when $\epsilon_t=0$ for all t in (6), and note that the EFs are obviously SSGC. Small-sample Gauss-consistency (SSGCy) is merely a desirable property, using Gauss's tool for studying the properties of estimators. Note that SSGCy is different from unbiasedness, since it does not involve averaging. Just as the conventional asymptotics is useful, even though T may never be actually infinite, SSGCy can be useful even though all errors may never be actually zero. Kadane (1970) implemented SSGCy as small-sigma asymptotics, which is conveniently discussed later in Remark 7.

Wedderburn (1974) defined integral of S_T as quasi-likelihood, used in econometrics by White (1982), without referring to Wedderburn. Quasi-likelihoods require only the mean and variance and admit the exponential family of distributions. If only the first two moments are specified (no normality), the likelihood is unknown and the ML estimator is undefined. By contrast, EFs remain available and "attain Cramer-Rao." Economists' common impression that "under normality, ML is unbeatable" is proved wrong in Godambe (1985) and further explained in Godambe and Kale (1991).

Godambe and Heyde(1987) prove that EFs yield asymptotically *shortest confidence intervals.* Vinod (1996b) explores this with regression examples. Small and McLeish (1988) show that EFs minimize the asymptotic mean squared error (MSE). The EF literature shows that solving an OptEF starting at some initial consistent estimator followed by Fisher's method of scoring, is similar to the Newton-Raphson iterates with readily available algorithms and "more normal" properties.

Let the parameter of interest be a smooth function $\nu(\beta)$ with nonzero derivatives, (e.g., long-run elasticities in econometrics). The chain rule implies that the new score for $\nu(\beta)$ is the original score S_T times $(\partial\beta/\partial\nu)$. Assuming that S_T can be written as a sum over t, the new score equation, which is the new EF, can also be a sum over t, and by Remark 1 it too is "more normal". If we ignore the "main lesson" above and focus on the roots, its sampling distribution can be quite nonnormal. For example, if $\nu(\beta)=\beta^2$, it is easy to verify that the sampling distribution of the root is χ^2.

Remark 2: The EF literature seems to ignore warnings of numerical mathematicians. For example, finding regression coefficients by solving the normal equations (EFs) directly is not advisable, especially for ill-conditioned problems, due to rounding and truncation errors, Sawitzski (1994). Vinod and Ullah (1981) formulate the regression problem in terms of the singular value decomposition (SVD) of $X = H\Lambda^{1/2}G'$, which leads to $\hat{\beta}_{ols} = G\Lambda^{-1/2}H'y$, as the computationally most reliable technique.

If the regressors X are correlated with the errors we have $X'\Omega^{-1}\epsilon \neq 0$. This leads to biased EFs from (2), $\mathrm{E}g_{gls} \neq 0$. The familiar method of instrumental variables (IV) assumes that we have data on a T×r matrix Z of r instruments which are (i) correlated with X, and (ii) uncorrelated with ϵ. The condition (ii) can be written as: $EZ'\Omega^{-1}\epsilon = 0$, which obviously achieves an unbiased EF of the IV estimator:

$$g_{iv} = Z'\Omega^{-1}(y - X\beta) = Z'\Omega^{-1}(y - \mu). \tag{7}$$

We will note [after eq. (18)] that replacing X by Z in IV-type estimation leads to "overidentification problems." Instead, Godambe and Thompson (1989) prove that it is "optimal" to replace X in the score equation (4) by its own expectation, (denoted later by X^c), hierarchically "conditional" on exogenous and lagged variables. Singh and Rao (1997) generalize "hierarchical conditioning" in Godambe-Thompson theorem and provide results on asymptotic efficiency. In (A.8) of our appendix we use AR(3) and AR(4) to replace a regressor $logC_t$ by its predicted value.

3 Estimating Functions, GMM And ML Estimators

It is well-known in econometrics texts that OLS, GLS, ML and IV can be viewed as special cases of GMM (e.g., Hamilton, 1994, ch. 14). Since the summation in (6) can be replaced by an expectation, it is obvious that the moment conditions of GMM can be viewed as EFs in (6). Since the GMM is a direct generalization of Pearson's method of moments, it has (conditional) moment equations. The first moment leads to the following EF:

$$E[g_{mom}(y, X, \beta)] = E[X_t'(y_t - X_t\beta)] = 0, \tag{8}$$

The GMM solution of (8) is obtained flexibly by minimizing an associated quadratic form $g_{mom}' W g_{mom}$, where W is *any* positive definite weight matrix.

Remark 3: The asymptotically optimal GMM minimizes $g_{mom}' W g_{mom}$, for $W = [Lim_{T\to\infty}\ TE(g_{mom}g_{mom}')]^{-1}$, Hamilton (1994, p. 413). Unfortunately, this W involves *ad hoc* choices and possibly impractical evaluation of the $Lim_{T\to\infty}$. By contrast, when available, the score is the OptEF without using asymptotic arguments. Since OptEF "attains Cramer-Rao," optimal GMM *must* be suboptimal in small samples, if it is different from the OptEF.

Dhrymes (1994, p. 368) calls GMM a "minor" modification of standard IV methods. However, GMM is popular in the context of "rational expectations" framework in economics. It can adapt to dynamic macroeconomics, where the Euler equations can be directly used as moment equations. For example, Christiano and Eichenbaum (1992) write the "first order conditions" from the "real business cycle" theory as the moment restrictions of the GMM. An Appendix illustrates another application deriving (A.4) as the EF for Tauchen's (1986) intertemporal utility maximization in a consumption based capital asset pricing model (C-CAPM). See Ogaki (1993) for C-CAPM details.

Result 1: Under certain assumptions of the regression model, EF methods yield the same $\hat{\beta}$ as the OLS, ML, GLS, IV and GMM estimators. However, OptEF methods may involve different choice of instruments or a different estimate of Ω yielding a different $\hat{\beta}$. Then, OptEF alone has small-sample optimality and assumes no more than existence of first two moments. For example, if due to heteroscedasticity, $Var(y_t)$ is a function of β, Godambe and Kale (1991) prove that the OptEF estimator is superior to GLS and ML.

It is well known that ML estimators (found by solving the score equation (2) as the EF) enjoy an equivariance property wrt a one-to-one differential transformation, a property not shared by the unbiased minimum variance estimators. Now assume that $\Omega(\phi)$ is unknown, and collect the parameters as $\theta = (\beta, \phi)$, where β are parameters of interest and ϕ are the nuisance parameters. The so-called Neyman and Scott (1948) problem is that the ML estimator obtained by ignoring nuisance parameters ϕ can be inefficient and inconsistent. Though this was published as a lead article in Econometrica, it is rarely, if ever, cited in econometric literature. The EF literature shows how to avoid the Neyman-Scott problem. Let q denote a complete sufficient statistic for ϕ for each fixed θ. Assume that q is independent of β and let f denote a generic density. We have

$$f(y; \beta, \phi) = f(y|q, \beta)\ f(q; \beta, \phi), \tag{9}$$

where the first part is a conditional density. Godambe (1976) shows that the conditional score

$$S_{cnd} = \partial log f(y|q, \theta)/\partial\theta = 0 \tag{10}$$

is an OptEF, since it maximizes the G-criterion. Lindsey's (1982) conditional score subtracts expected value to achieve unbiasedness. For time series data, Godambe (1985) uses conditional expectations using information till time t-1. Similarly Godambe and Thompson's (1989) hierarchical conditioning, and results in Godambe (1991) and Bhapkar (1991) can be used to avoid the Neyman-Scott problem.

Our next task is to evaluate the G-criterion of (5) using $g^*=S_{cnd}=0$ of (10) and understand its deeper meaning. Clearly, the D_{g*} in (5) equals the Fisher information matrix $I_F = -E[\partial^2 L_T/(\partial\theta\partial\theta')]$. Let us denote this "second order partial" (Hessian) form of I_F by I_{2op}. The $E(g^* g^{*\prime})$ in (5) equals the "outer product of gradients" form, denoted here by I_{opg}. Thus we denote:

$$I_{opg} = E\{(\partial L_T/\partial\theta_i)(\partial L_T/\partial\theta_j)\},\ I_{2op} = -E\{\partial^2 L_T/\partial\theta_i\partial\theta_j\},\ (i, j = 1, \cdots, p). \tag{11}$$

Result 2: If g* proportional to a score vector similar to S_T,

$$G\text{-criterion} = I_{2op}^{-1}\ I_{opg} I_{2op}^{-1}. \tag{12}$$

For proof use (5) and (11).

White's (1982) "information matrix equivalence theorem" states that: when the model is correctly specified,

$$I_F = I_{2op} = I_{opg} \tag{13}$$

White further developed (13) into a specification test.
Corollary of Result 2: Only when the model is correctly specified in the sense of (13) the G-criterion minimand (12) reduces to I_F^{-1}, which is proportional to the variance.

Rather than using the normal likelihood of (3), we recall $\mu = X\beta$ notation and verify this corollary in a simpler setting of estimating the mean μ of independent and identically distributed (iid) variates:

$$y_t \sim IID(\mu, \sigma^2). \tag{14}$$

Now, the simpler quasi-log-likelihood of an observation is

$$log f(y_t, \mu, \sigma^2) = -0.5 log 2\pi - 0.5 log \sigma^2 - 0.5(y_t - \mu)^2/\sigma^2. \tag{15}$$

Defining $\theta = (\mu,\ \sigma^2)$, the Fisher information $I_F = I_{2op}$ is a 2×2 diagonal matrix having: $(\sigma^{-2},\ 2^{-1}\sigma^{-4})$ along the diagonal. Since its inverse has variances along the diagonal, we have verified that I_F^{-1} is proportional to the variance. Next, we have:

$$I_{\text{opg}} = \begin{bmatrix} \sigma^{-2} & 2^{-1}\gamma_1\sigma^{-3} \\ 2^{-1}\gamma_1\sigma^{-3} & 4^{-1}(\gamma_2 + 2)\sigma^{-4} \end{bmatrix} \tag{16}$$

where $\gamma_1 = E(y_t - \mu)^3/\sigma^3$ and $\gamma_2 = E(y_t - \mu)^4/\sigma^4 - 3$, measure skewness and kurtosis. Clearly, only if $\gamma_1 = 0$ and $\gamma_2 = 0$, $I_{2op} = I_{opg}$ are both diagonal matrices satisfying (13). Thus, by avoiding (13) the G-criterion offers (robustness) some protection against misspecification. This discussion of the corollary is intended to help the intuition.

Remark 4: The practical success of EF methods reported in biostatistics, sampling and elsewhere may be due to the corollary. The G-criterion of (12) seeks efficiency (low variance), as it adjusts for misspecification arising from $I_{opg} \neq I_{2op}$ in finite samples. Minimizing the G-criterion (12) is not directly attempted in econometrics, though minimizing variance (efficient estimation) is ubiquitous. The G-criterion needs only finite mean and variance, and quasi (not actual) likelihood functions. By contrast, the ML is sensitive to misspecification of the likelihood function. In short, EFs provide a more general and flexible estimation theory, which remains applicable in finite samples and is insensitive to possible misspecification or nonexistence of the likelihood function.

4 Further Results On Instrumental Variables And GMM

Consider the class $\tilde{G}$ of all estimating functions of the form $g = A\epsilon$, where A is a p×T matrix, and $E\epsilon = E(y - \mu) = 0$. In this section, we eliminate σ^2 for brevity and redefine $\Omega = E\epsilon\epsilon' = E(y-\mu)(y-\mu)'$. Also, let $H(\beta)$ denote any p×p nonsingular nonstochastic matrix whose components depend on β, and let the superscript + denote a generalized or pseudoinverse of the matrix.

Result 3: In the class $\tilde{G}$, g* is an optimum estimating function (OptEF) if and only if

$$g^* = H(\beta)[\partial(y-\mu)/\partial\beta]'\Omega^+(y-\mu). \tag{17}$$

The proof is analogous to Vijayan (1991), who has proved a special case when Ω is a diagonal matrix, which is of interest in survey sampling. Despite possibly nonzero off-diagonals, essentially the same arguments apply here. In particular, for the GLS we have $\mu = X\beta$, $X = \partial(y-\mu)/\partial\beta$. Hence the choice $A = X'\Omega^+$ gives the OptEF. This reduces to (2), the EF for the GLS, when $H(\beta) = I$ and Ω is nonsingular. Thus (2) is the OptEF under certain assumptions. If the EF is biased, $EA(y-\mu) \neq 0$, we replace A by $A_z = Z'\Omega^+$ choosing the instruments Z which satisfy $EA_z(y-\mu) = 0 = Eg^*$. This is when g_{iv} of (7) is the OptEF.

Result 4: If Z is a T×r matrix (of rank r) of instrumental variables, instead of (5) our new G-criterion minimand is

$$(Z'\Omega^+X)^+(Z'\Omega^+Z)(X'\Omega^+Z)^+. \tag{18}$$

To derive (18), use (7) and write $g_{iv} = Z'\Omega^+\epsilon$. Now, verify that $Eg_{iv}g_{iv}' = Z'\Omega^+Z$, and that the D_{g*} from (5) becomes $(Z'\Omega^+X)$. Since each instrumental variable is observable (inside the information set), our A_z is observable in the linear case. Ogaki's (1993, p. 459) extension to the nonlinear case can be adapted here.

The term "overidentification" in econometrics refers to simultaneous equation models when r>p, where r denotes the number of exogenous (predetermined) variables in the complete model, and p denotes the number of parameters in a single equation. The minimand (18) is designed for the overidentified GMM case, where r>p means that there are more (moment conditions) instrumental variables than regression parameters. In the EF theory, since one can combine many EFs into one, overidentifying restrictions can be superfluous. For example, we can combine $g_1=0$ and $g_2=0$ into $g_1+g_2=0$. The "hierarchical conditioning" mentioned earlier obtains OptEF by replacing the X in (2) by a matrix of conditional expectations denoted here by X^c. The i-th column of X^c may be obtained by regressing i-th column of X on a column of ones (intercept) and any or all r columns of Z for i=1, $\cdots$, p. Now we list some advantages of X^c over the Z matrix of instruments from (7): (i) The column dimension of Z as regressors does not matter, and one need not worry about overidentification, let alone have a complicated formal test for it. (ii) It is easy to guarantee that X^c columns are highly correlated with the X columns. (iii) A simpler G-criterion can be used to attain optimality. For an example, see our appendix.

For the overidentified case, the GMM moment condition (8) becomes $E(Z_t'(y_t - \mu_t)) = 0$ for any t, where Z_t denotes the t-th row of Z. Also, the GMM minimand becomes:

$$(y-\mu)'ZWZ'(y-\mu), \tag{19}$$

where W is an r×r matrix of weights. Assuming that X is of full rank p and r≥p, i.e., we have enough instruments, then GMM estimator $\hat{\beta}_{gmm}$ is a solution of the following p EFs:

$$g_{gmm} = (X'ZWZ'X)\beta - X'ZWZ'y = 0. \tag{20}$$

Recent GMM theory notes that the optimal r×r weight matrix W is: $W = S_T\Omega^{-1}S_T'$, where $S_T = Z'\Omega^{-1}(y - X\beta)$ is the r×1 score vector similar to (7). Since score equations are OptEFs this brings GMM and EF theories closer. However, substituting this W in (20) seems to be unnecessarily complicated.

Remark 5: Only when the W matrix which minimizes (19) also minimizes the G-criterion of (18) will the GMM solution coincide with the EF solution. Otherwise the GMM is suboptimal, as we have already noted in remark 3.

Result 5: Denote the root of an EF by $\hat{\beta}_{ef}$, and let the corresponding vector containing T residuals be $\hat{\epsilon} = y - \hat{\mu}(\hat{\beta}_{ef})$. Since the GLS normal equations (2) are EFs, their "equation residuals" are (degenerate) identically zero, or $X'\Omega^{-1}\hat{\epsilon} \equiv 0$. Recall that we may avoid biased EFs by replacing X by X^c, a conditional expectation based on Z instruments. Denote $A_z\hat{\epsilon} \equiv X^{c\prime}\Omega^{+}\hat{\epsilon} \equiv 0$. To verify SSGCy we set $\hat{\epsilon} \equiv 0$ and note that this amounts to solving the normal equations. Thus, we always have SSGCy for these unbiased EFs. However, the true unknown "equation errors" $g^* = A_z\epsilon$ are a non-degenerate vector of random variables with zero mean and variance

$$V = Var(g^*) = Eg^*g^{*\prime} = EA_z\epsilon\epsilon'A_z' = A_z\Omega A_z'. \tag{21}$$

Since $A_z\hat{\epsilon} \equiv 0$, it is incorrect to replace ϵ in (21) by $\hat{\epsilon}$. Instead, our estimate of variance is:

$$\hat{V} = X^{c\prime}\hat{\Omega}^{+}X^{c}. \tag{22}$$

Remark 6: Equation (22) needs a consistent estimate of Ω^{+} (or Ω^{-1}, if it exists). If regression errors follow an AR(1) process with parameter ρ, it is well-known that Ω^{-1} is proportional to a tridiagonal matrix: It has $1+\rho^2$ along the main diagonal, except for ones in the two corners and a $-\rho$ along the sub and super diagonal. More generally, we find the parameters of appropriate (by Schwartz criterion) autoregressive integrated moving average

(ARIMA) process for residual $\hat{\epsilon}$ and indirectly determine autocovariances $\hat{\phi}_j$ for j=1, $\cdots$, q$\leq$T. Next, create a (symmetric) Toeplitz matrix $\hat{\Omega}$ from $\hat{\phi}_j$ and substitute in (22). To achieve robustness against both autocorrelation and heteroscedasticity, we need consistent estimates of heteroscedastic variances to insert into the diagonals of Ω.

Remark 7: Kadane's (1970) small-sigma asymptotics explained in Vinod and Ullah (1981, p. 162) is an implementation of SSGCy. It rewrites the error as $u = \sigma\epsilon$, $Euu' = \sigma^2\Omega$ and finds limits as the constant $\sigma \to 0$. Consider a biased EF arising when appropriate instrumental variables Z are unavailable: $g = A_x\epsilon = A_x(y - \mu)$, where $A_x = X'\Omega^+$, $A_x\hat{\epsilon} \neq 0$. Denote the bias by $B_s = EA_x\epsilon = EA_x(y - \hat{\mu} + \hat{\mu} - \mu) = EA_x\hat{u}\sigma + EA_x(\hat{\mu} - \mu)$, where $g = A_x\hat{u}\sigma + A_x(\hat{\mu} - \mu)$ is now a function of Kadane's σ. Since $EA_x\hat{u}\sigma \to 0$ as $\sigma \to 0$, small-σ approximate bias is $B_{s\sigma} = EA_x(\hat{\mu} - \mu)$, linking the bias of EFs with the bias of estimators $\hat{\mu}$. Writing the bias, variance and MSE as a power series in σ, one obtains small-σ approximations by ignoring high powers of σ. The variance of the biased EF is

$$E(gg') = (EA_x\epsilon\epsilon' A_x') = (EA_x uu' A_x')/\sigma^2 \tag{23}$$

If $\mu = X\beta$, small-σ bias is $B_{s\sigma} = EX'\Omega^+X(\hat{\beta}_{ef} - \beta)$. If $\mu = X\beta$ and X is nonstochastic, $E(gg') = (X'\Omega^+X)$, where we use $Euu' = \sigma^2\Omega$ and cancel the σ^2 in the denominator of (23). More research is needed to explore small-sigma methods for EFs.

5 Estimating Functions In Further Settings

Cox's (1975) partial likelihood score, which is a sum of appropriately conditioned scores (10), is widely used in biostatistics and elsewhere. Godambe (1985) proves that the optimal coefficient in g^* is $E_{t-1}(\partial^2 S_{cnd}/\partial\theta^2)/E_{t-1}(S_{cnd})^2 = -I_{2op}/Var(S_{cnd})$. Next, he proves that the EF theory implies optimality of the partial likelihood scores, which have been used by practitioners since the late 1970's.

Wedderburn (1974) considers a general linear model (GLM) with independent responses. The GLM is specified by a known (monotonic differentiable) link function $\nu(\mu_t) = \nu(X_t'\beta)$, and the random part is from the exponential family. The Box-Cox transformation is one kind of link function well-known in econometrics. The quasi-score function for GLM is

$$D'\Omega^{-1}(y - \mu) = 0, \tag{24}$$

where $\mu = \nu^{-1}(X\beta)$, $D = \partial\mu_i/\partial\beta_j$ is a p$\times$p matrix, and $\Omega = diag(\Omega_t)$ is T$\times$T diagonal matrix. The logistic model has $\mu_t = exp(X_t'\beta)/[1+exp(X_t'\beta)]$,

$\partial\mu/\partial\beta_j = x_{tj}\mu_t(1-\mu_t)$. Its familiar link function $\mu_t = log(y_t/(1-y_t))$ provides a useful simplification and linear estimating functions.

Godambe and Heyde (1987) show that the quasi score function (which minimizes the G-criterion) is an optimal estimating function. Godambe and Thompson (1989) extend the quasi score function to include the quadratic $w_t = (y_t - \mu_t)^2$ and show that the variance matrix for $(y_t,\ w_t)$ involves skewness and kurtosis parameters in (16). The GLS estimating functions (2) are a special case when $D = X$.

For repeated k measurements (panel) of correlated data, the D matrix of (24) becomes block diagonal, and we have similar estimating functions with an additional summation over k blocks. These are called generalized estimating equations (GEE) defined by:

$$\Sigma_{j=1}^{k} D_j' \Omega_j^{-1}(y_j - \mu_j) = 0, \tag{25}$$

where the matrices have subscript j to distinguish them from those of (24) in analogous notation. This will correctly account for the correlations ϕ incorporated in $\Omega_j(\phi)$ if they are known. Otherwise, one uses the feasible version by replacing them by consistent estimators denoted by $\hat{\Omega}_j$. The asymptotic covariance matrix of GEE estimates of β is given by

$$Var(\hat{\beta}_{gee}) = \sigma^2 A^{-1} B A^{-1}, \text{ with } A = \Sigma_{j=1}^{k} D_j' \hat{\Omega}_j^{-1} D_j \text{ and}$$

$$B = \Sigma_{j=1}^{k} D_j' \hat{\Omega}_j^{-1} \Omega_j \hat{\Omega}_j^{-1} D_j. \tag{26}$$

Although repeated measurements are rare in econometrics, the above methodology can be useful. For example, Vinod (1989) uses a fuzzy range of rounding errors around each measurement to artificially create repeated measurements. In biostatistics, GEE has been applied to study various situations where y is continuous, binary or count. Hence econometric applications to similar data (e.g., panel) are worth considering.

6 Statistical Testing And Inference With New Bootstraps

Statistical testing is important in econometrics and the bootstrap has been proved to be useful for difficult inference problems, Hall and Horowitz (1996), Vinod (1993). Denote Fisher's pivotal for i-th regression coefficient as: $\epsilon_F = (\hat{\beta}_i - \beta_i)/SE_i$, where SE_i denotes the standard error. This pivot is asymptotically standard normal, N(0,1), and since it is a function of (y, X, β), it is an EF itself. Inverting Fisher-type pivotals (having $\hat{\beta}_i - \beta_i$ in the numerator) always yields symmetric confidence intervals. For the 95% case, inversion yields $(\hat{\beta}_i \mp 1.96 SE_i)$, a symmetric confidence interval for β_i.

Godambe and Heyde(1987) prove that EFs yield asymptotically shortest confidence intervals. Lele (1991) suggests a symmetric EF bootstrap for dependent time-series data. Vinod (1995) suggests using the double bootstrap to deal with the problem of a lack-of-a-pivot, whereas Vinod (1996b) considers asymmetric intervals and uses the double bootstrap. Thavaneswaran (1991) notes that the sequence of EFs based on Godambe (1985) is a "martingale difference sequence" and provides Wald, Rao's score and nonparametric test statistics involving martingales and semimartingales. His applications include testing for structural change, common in econometrics. Heyde and Morton (1993) show that the restricted OLS estimator (RLS in econometrics texts) can be derived from projections in EF theory without using Lagrangians. For testing linear restrictions, one can use their EF estimators and corresponding Fisher information as covariance matrices for construction of test statistics. However, we propose a different approach where the pivotal is constructed from g and Egg' rather than the usual $[\hat{\beta}_{ef} - \beta] \times [Var(\hat{\beta}_{ef})]^{-1/2}$.

Given $Eg = 0$ (unbiased g) we seek the square root matrix of the p×p variance matrix $V = Egg'$ of EF if V is known. Otherwise, we use a consistent, robust and nonparametric estimators $\hat{V}$ of V from (22). For related literature, see Ogaki (1993) and Andrews and Monahan (1992), among others. Now consider a square root decomposition:

$$V = \sigma_L \sigma_R, \quad \text{if available, or} \quad \hat{V} = \sigma_L \sigma_R, \tag{27}$$

where σ_L denotes the p×p left-square-root matrix decomposition of the matrix, and σ_R a similar right-square-root matrix, assuming they exist. The computations may be based on either the Cholesky or symmetric versions. If $Eg \neq 0$ due to nuisance parameters being estimated, Godambe (1991, p.144) suggests using $\tilde{g} = g - Eg$ as the revised EF, where the expectation is conditional on minimal sufficient statistic for the nuisance parameter. For biased estimators, Vinod (1984) argues that it is useful to replace the variance in the denominator of the usual t ratio by the square root of an unbiased estimate of the mean squared error ($\sqrt{}$UMSE). However, he shows that the resulting distribution is similar to a noncentral Chi-square, depending on unknown noncentrality; hence the revised t-ratio is only an approximate pivotal. Thus, if g is biased, we may replace the $\hat{V}$ in (27) by a matrix containing the UMSE. Further work is needed to explore the use of UMSE here.

Result 6: An approximate pivotal for the EF estimator $\hat{\beta}_{ef}$ obtained by solving $g = 0$ can be

$$U = \sigma_L^{-1} g, \tag{28}$$

where we assume $E(g) = 0$ and use the square root decomposition in (27).

For example, substituting $\hat{V}$ from (22) and $\hat{\Omega}$ from Remark 6 in (27) one can incorporate autocorrelations and heteroscedasticity. As a pivotal, we know that asymptotically, U is well behaved and $Var(U) \to I$. By contrast, a pivotal for $\hat{\beta}_{ef}$ needs variance of the sampling distribution of the root $\hat{\beta}_{ef}$ of $g = 0$, which can be more complicated. When nuisance parameters ϕ are present, the distribution of g may depend on ϕ. To solve this problem Morton (1981) considers a pivotal g, where the distribution depends only on the parameters of interest β. Parzen et al (1994) provide two mild regularity conditions on general EFs for asymptotically valid bootstrap confidence intervals.

Now we outline typical steps of a bootstrap for time-series regressions, designed to overcome the time dependence (among regression errors). These should be modified to suit specific examples, and all computations should use numerically reliable methods (e.g., use SVD of Remark 2 when applicable).

Steps for a Parametric Bootstrap using EFs:

(i) Denote i-th columns of X and X^c by x_i and x_i^c, respectively. Let $x_i^c = x_i$ for (deterministic or exogenous) variables in X known to be uncorrelated with errors, $(y - \mu)$. Regress each of the remaining (correlated) x_i on a column of ones (for intercept) and all relevant instruments Z. Use the predicted value from this regression to complete the X^c matrix. Construct p optimal EFs $g^* = X^{c\prime}\hat{\Omega}^+(y - \mu(\beta))$, where we permit some nonlinear functions $\mu(\beta))$ instead of the usual $X\beta$. Unless Ω is known, use $\Omega = I$, to obtain a preliminary p×1 vector $\hat{\beta}_{pre}$ by solving $g^* = 0$.

(ii) Use the T residuals $\hat{\epsilon}_t = y_t - \mu(\hat{\beta}_{pre})_t$ and Remark 6 to construct q autocovariances $\hat{\phi}_j$ and hence a T×T Toeplitz matrix $\hat{\Omega}$ to approximately represent time dependence. Make further adjustments to the diagonal terms to allow for heteroscedasticity, if desired.

(iii) Solve the revised $g^* = X^{c\prime}\hat{\Omega}^+(y - \mu(\beta)) = 0$ to obtain $\hat{\beta}_{ef}$, new residuals $\tilde{e}_t = y_t - \mu_t(\hat{\beta}_{ef})$, and revised $\tilde{\Omega}$. In some examples it may be useful to repeat this step until $\hat{\beta}_{ef}$ values converge within a given tolerance.

(iv) Obtain a (Cholesky) factorization of the p×p matrix $\tilde{V} = X^{c\prime}\tilde{\Omega}^+X^c = \sigma_L\sigma_R$, from (22) defining σ_L and σ_R.

(v) Generate J(=999, say) simulated p×1 vectors $e_j^* = \sigma_L v$, of "equation errors," where v is a p×1 vector of N(0,1) unit normal deviates and j=1, $\cdots$ J.

(vi) Solve J different EFs: $e_j^* = X^{c\prime}\tilde{\Omega}^+(y - \mu(\beta))$ to yield J estimates of the p×1 vector β denoted by $\hat{\beta}^*_{efj}$. If some function f(β) is of interest, (e.g., long-run elasticity), compute J estimates of that function denoted by $\hat{f}_j(\hat{\beta}^*_{ef})$.

(vii) Order the p components of β separately in an increasing order and denote the order statistics by $\hat{\beta}^*_{ef(j)}$. For a 95% two-sided confidence interval use $\hat{\beta}^*_{ef(25)}$ as the lower limit and $\hat{\beta}^*_{ef(975)}$ as the upper limit. Similarly, from the function estimates $\hat{f}_j(\hat{\beta}^*_{ef}) = \hat{f}_j$ the confidence interval is $[\hat{f}_{(25)}, \hat{f}_{(975)}]$.

Since the EFs are "more normal" than estimators of β (due to averaging), and since we are using EFs as pivots, we have reliable statistical inference according to the bootstrap and EF theories. A null hypothesis regarding functions f(β), similar to long-run elasticities of Remark 1, can be tested by using appropriate confidence intervals from f($\hat{\beta}^*_{ef}$) in step (vii). The following nonparametric bootstrap can avoid estimation of Ω.

A Nonparametric Bootstrap using EFs:

Vinod (1996a) suggests using scaled recursive residuals, recently surveyed by Kianifard and Swallow (1996). For regressions with dependent data they are attractive, because they are iid with zero mean and unit variance. Denote by X_τ a $\tau \times$p matrix consisting of the first τ rows of X, provided $\tau \geq$p+1. Let y$_\tau$ denote the $\tau \times 1$ vector of initial τ observations of y, and let $b_\tau = (X'_\tau X_\tau)^{-1} X'_\tau y_\tau$ denote the corresponding OLS estimator. Denote x_τ such that $\hat{y}_\tau = x'_\tau b_{\tau-1}$ is the "conditional forecast" of y_τ for $\tau = p+1, \cdots$,T. Instead of using $\hat{y}_\tau$, we define recursive residuals in an equivalent simpler form:

$$w_\tau = (y_\tau - x'_\tau b_\tau)/\tilde{\sigma}_\tau, \quad \text{where} \quad \tilde{\sigma}_\tau = [1 - x'_\tau (X'_\tau X_\tau)^{-1} x_\tau]^{1/2}, \qquad (29)$$

where the residual in the numerator, $y_\tau - \hat{\mu}_\tau$, is scaled by $\tilde{\sigma}_\tau$ to achieve $Var(w_\tau) = 1$. Kianifard and Swallow (1996, p. 393) provide a computational strategy and write the (T − p)×1 vector $w = w_\tau = Cy$, where C is a (T− p)×T complicated matrix. They show that $CX = 0, Ew = 0$, $Eww' = \sigma^2 I_{T-p}$, and $w'w = \epsilon'\epsilon$. Denote the residual sum of squares by RSS$_\tau$ = $(y_\tau - x'_\tau b_\tau)'(y_\tau$-$-x'_\tau b_\tau)$. Since $RSS_\tau = RSS_{\tau-1} + w^2_\tau$, this gives a numerically reliable method of estimating $\tilde{\sigma}_\tau$.

From the OLS normal equations, the EFs are (T−p) separate sets for $\tau = p+1, \cdots$,T: $g^*_\tau = X^{c\prime}(y_\tau - \mu_\tau)/\tilde{\sigma}_\tau$, where $\mu_\tau = x'_\tau \beta_\tau$. Each τ leads to p equations $g^*_\tau = 0$ yielding T−p sets of $\hat{\beta}_{ef\tau}$. However, we are mainly interested in the last solution $\hat{\beta}_{ef} = \hat{\beta}_{efT}$ using all available data. Bootstrap shuffling of the T−p iid w_τ residuals avoids the problems with bootstrap for time series (dependent) data. Denote shuffled recursive residuals as $w_{\tau j}$ for j=1,$\cdots$, J(=999). Substitute $w_{\tau j}$ in (29) to get $y_{\tau j} = \tilde{y}_\tau + w_{\tau j}\tilde{\sigma}_\tau$. Next, we assemble $y^*_j = (y_1, \cdots y_p, y_{1j}, \cdots y_{\tau j}, \cdots y_{Tj})'$, into a T×1 vector of y values for the j-th bootstrap resample. Now, solve $X^{c\prime}(y^*_j - \mu(\beta)) = 0$, to yield J estimates of p×1 vectors denoted by $\hat{\beta}^*_{efj}$. Finally follow the step (viii) mentioned earlier for statistical inference and confidence intervals. An Appendix implements this for the C-CAPM model using the GAUSS and RATS computer languages.

7 Conclusion And Final Remarks

This paper discusses EFs, which are functions of parameters and data, $g(y, X, \beta) = 0$, in the regression context of econometrics. The main lesson of the EF theory is to focus attention on EFs while deemphasizing their roots (or estimators) $\hat{\beta}$. The EFs are shown to be "more normal" than the roots, leading to better behaved pivotals for inference and easier conditioning arguments. Applied researchers sometimes wish to avoid the ML estimator, since the likelihood exists only under the (often unverifiable) normality assumption. If they are confident about the existence of the mean and variance, Wedderburn's (1974) quasi likelihood score can be the optimal EF. Using the EFs econometricians can avoid some fancy asymptotic theory and 'overidentifying restrictions'. Also, some recent unit root and other estimators of autoregressive distributed lag (ADL) models are shown to be unnecessarily complicated and suboptimal in small samples compared to Durbin's (1960, p.151) two regression (TR) estimator, further justified by the recent EF theory.

We show that the optimum EF estimate $\hat{\beta}$ based on the quasi score "attains Cramer-Rao," and is small-sample Gauss-consistent (SSGC). Further advantages of EFs proved in the literature are: minimization of the asymptotic mean square error, ability to yield shortest confidence intervals, solve the Neyman-Scott problem and provide efficient Newton-Raphson iterates. Our references also give examples of practical uses of the EFs. We include an appendix where a familiar econometric example of dynamic rational expectation model (C-CAPM) shows that EF estimates are similar to GMM estimates, while avoiding some of the *ad hoc* features of the GMM. For inference on this example, we implement the nonparametric bootstrap using the iid recursive-residuals from (29). Our EFs yield shorter confidence intervals and our estimates of both risk aversion and time preference parameters are economically more meaningful than the GMM estimates. This application using commonly available data demonstrates that EFs are practical and potentially valuable tools for econometrics.

This paper also includes several "results" and "remarks" which are potentially useful, and provides detailed steps for two *new* time series bootstraps for possibly dependent data. Since Euler equations or first-order conditions are readily regarded as EFs, there are many potential applications in economic dynamics. We can test complicated dynamic economic theories directly by using the EF methods, and recommend using the bootstrap when inference becomes analytically complicated. We have demonstrated that EFs can improve upon some esoteric econometrics and simplify applications.

APPENDIX A

GMM and EFs for C-CAPM

Generalized Method of Moments and Estimating Functions for Estimation of Consumption-Based Capital Asset Pricing Model

Consider a representative agent model leading to C-CAPM, see Tauchen (1986) and Ogaki (1993). Denote c_t= consumption, p_{it}=price ex-dividend of i-th asset, and d_{it}=real dividend. Let an intertemporal budget constraint be: $c_t + \Sigma_{i=1}^{M} q_{it}p_{it} = \Sigma_{i=1}^{M} q_{i,t-1}(p_{it} + d_{it})$, where q_{it}=amount of the i-th asset the agent carries into the next period t+1. Now the intertemporal utility is assumed to be additively separable. Lifetime utility of an agent with infinite time horizon (due to bequests) is

$$E_t \Sigma_k^{\infty} \beta^k u(c_{t+k}), \tag{A.1}$$

where E_t is conditional expectation using all information till time t, and $0< \beta < 1$ is subjective rate of time preference. If the constant subjective interest rate is i_s, the parameter β is defined as $1/(1+i_s)$. The market interest may determine i_s and an "instrumental variable" may be based market interest rate data. In (A.1) the first and second derivatives (denoted by primes) of the utility function satisfy: $u' > 0$ and $u'' < 0$. A popular utility function is: $u(c) = c^{1-\gamma}/(1-\gamma)$, with $\gamma > 0$, has $u' = c^{-\gamma}$ and $u'' = -\gamma c^{-\gamma-1}$ where $u'' < 0$. Since $cu''/u' = -\gamma$, is a constant, this $u(c)$ belongs to the constant relative risk aversion (CRRA) family of utility functions. The estimation of risk aversion parameter γ is more important than that of the discount parameter β. We do not permit the agent to continuously roll over the debt. The already mentioned first-order conditions for maximizing the lifetime utility, subject to the budget constraint, are the Euler equations:

$$p_{it}u'(c_t) = \beta E_t[u'(c_{t+1})(p_{i,t+1} + d_{i,t+1})], \tag{A.2}$$

where (i=1,$\cdots$,M). Using the CRRA utility function for u, we rewrite (A.2) as

$$p_{i,t}c_t^{-\gamma} = \beta E_t p_{i,t+1} c_{t+1}^{-\gamma} + \beta E_t d_{i,t+1} c_{t+1}^{-\gamma}, \tag{A.3}$$

which can be written in terms of observable growth variables defined as ratios $\tilde{d}_{i,t} = d_{i,t}/d_{i,t-1}$, and $C_t = c_t/c_{t-1}$ and the price dividend ratio $\pi_{it} = p_{i,t}/d_{i,t}$ for the i-th asset as the estimating function (EF)

$$\Sigma_{t=1}^{T} g_{it}(\beta,\gamma) = \Sigma_{t=1}^{T}[\pi_{it} - \beta E_t(1+\pi_{i,t+1})C_{t+1}^{-\gamma}\,\tilde{d}_{i,t+1}] = 0, \tag{A.4}$$

where all terms are observable, except the parameters denoted by $\theta = (\beta, \gamma)$. Denote the gross return per dollar invested on the *entire market* from t to t+1, $[(1+\pi_{t+1})/\pi_t]\tilde{d}_{t+1}$, by R_{t+1} and rewrite (A.4) as:

$$\Sigma_{t=1}^{T} g_{mom,t}(\theta) = \Sigma_{t=1}^{T}[\beta R_{t+1} C_{t+1}^{-\gamma} - 1] = 0, \tag{A.5}$$

Typical GMM estimation involves following steps: (i) Choose r instrumental variables defined from lagged C and lagged asset returns to form an r×1 vector z_t for each t. (ii) Define $g_t(\theta) = g_{mom,t} z_t$. (iii) Compute time averages $g_T(\theta) = \Sigma_{t=1}^T g_t(\theta)/\mathrm{T}$. (iv) Compute $A = (1/T)\Sigma_{t=1}^T g_t g_t'$ with diagonals a_t. (v) Use the starting W_j for j=1 as a diagonal matrix having $(1/a_t)$. (vi) Compute the $\hat{\theta}_j$ for j=1 by minimizing a quadratic form $g_t(\theta)' W_j g_t(\theta)$. (vii) For j=2 let $W_j = [(1/T)\Sigma_{t=1}^T g_t(\hat{\theta}_{j-1}) g_t(\hat{\theta}_{j-1})']^{-1}$. (viii) Compute $\hat{\theta}_j$ as minimizers of $g_t(\hat{\theta}_{j-1})' W_j g_t(\hat{\theta}_{j-1})$. (ix) Repeat steps (vii) and (viii) till convergence and obtain the GMM estimator $\hat{\theta}_{gmm}$. It is assumed that W_j converge to $W_0 = [E g_t g_t']^{-1}$ evaluated at the true value θ_0, and that $E g_t g_t'$ is nonsingular and symmetric. (x) The final step is to find the covariance matrix $Var(\hat{\theta}_{gmm}) = D_t^{*\prime} W_0 D_t^*$, where $D_t^* = E_t D_t$, where D_t matrix contains partials of $g_t(\theta)$ with respect to θ. In practice, one substitutes time averages of D_t for D_t^* and replaces W_0 by the time average $[(1/T)\Sigma_{t=1}^T g_t g_t']^{-1}$, where the g_t are evaluated at the $\hat{\theta}_{gmm}$, Tauchen (1986).

Now we turn to our EF estimation. We rewrite (A.5) as $E(\beta R_{t+1} C_{t+1}^{-\gamma}) = 1$, which may be written as a nonlinear regression model with errors (not utilities) denoted by subscripted u:

$$\beta R_{t+1} C_{t+1}^{-\gamma} = 1 + u_{t+1}, \quad E u_{t+1} = 0, \quad u_{t+1} = \sigma \epsilon_{t+1}, \tag{A.6}$$

where σ refers to small-sigma asymptotics (Remark 7). Now consider limits as $\sigma \to 0$. The SSGCy requires that the method of estimation should yield the correct estimates when the model equation errors $u_t \equiv 0$ for all t. In models containing forward-looking economic time series, it is reasonable to extend SSGCy to require that the estimation method performs correctly when "expectational errors" are zero. Log of the right hand side of the regression in (A.6) is $log(1 + u_{t+1}) = u_{t+1} - u_{t+1}^2/2 + \cdots = \sigma\epsilon_{t+1} - \sigma^2 \epsilon_{t+1}^2/2$. Since $\sigma \to 0$, we can omit terms with σ^j for j$\geq$ 2. Next, we replace t+1 by t, and $\sigma\epsilon_{t+1}$ by u_t to write a new C-CAPM linearized regression:

$$logR_t = -log\beta + \gamma logC_t + u_t. \tag{A.7}$$

This new method suggested by the EF approach is potentially useful elsewhere in macroeconomics. Of course, the EFs are explicitly nonlinear functions, and small-σ asymptotics leading to (A.7) is not at all essential for selling the EFs. Since the correlation, $corr(u_t, logC_t)$, between the error and the (stochastic) regressor of (A.7) may be significant, we use k lagged values as instruments. We construct X^c from predicted value $log\hat{C}_t$ from an AR(k) regression: $logC_t = b_0 + \Sigma_{j=1}^k logC_{t-j}$. The k is chosen by the information criteria (AIC-type). It is well-known from the econometric literature dealing with two stage least squares that $corr(u_t, log\hat{C}_t) \to 0$. This supports similar "hierarchical conditioning" arguments from the EF

=====================================

TABLE 1a GMM estimation by RATS package and Nonlinear Instrumental Variables. (Quarterly data from 1960:01 to 1987:04) GMM estimates of β and γ with t-statistics in parentheses:
Case 1) T(Usable Observations)=110, df(Degrees of Freedom)=108, iterations=3, NLAG (Lags used for instruments)=1, Residual Sum of Squares (RSS)= 0.0092, Durbin-Watson statistic (DW)=1.6104. $\hat{\beta}$=0.99097 (54.86), $\hat{\gamma}$=1.24629 (t-value=0.4823, SE=2.58). Upper limit of confidence interval for γ (UpL) is much larger than 2 (UpL>> 2). The $\chi^2(1)$= 2.3841 < 3.8415, the tabulated 95 % value for df=1.

Case 2) T=109, df=107, iterations=3, NLAG=2, RSS= 0.0092, DW=1.5633. $\hat{\beta}$=0.99058 (72.83), $\hat{\gamma}$=1.18231 (0.6077, SE=1.94). Again UpL for $\gamma >> 2$. The $\chi^2(1)$= 4.0682.

=====================================

literature mentioned after (7). Here, (7) becomes:

$$g_{iv} = Z'\Omega^{-1}(y - X\tilde{\theta}), \tag{A.8}$$

where $Z = [\iota, log\hat{C}_t]$, where ι is a column of ones, $y = logR_t, X = [\iota, logC_t]$, and $\tilde{\theta} = (-log\beta, \gamma)'$.

Table 1a reports estimates for the GMM (using RATS software, rather than Tauchen's). Tables 1b to 1d have our EF estimates of $(A.7)$. Both methods use the same quarterly US data from 1960:1 to 1987:4, used by many others and which comes bundled with the RATS software. We find that the EF results are quite comparable in terms of the Durbin-Watson statistics and residual sum of squares as measures of fit. Both methods correctly estimate $\beta < 1$, $(\hat{\beta}_{ef} \approx 0.99)$.

We use subscripts to compare $\hat{\gamma}$ across tables. Using the subscript 0 for the no instrument case in Table 1b we have $\hat{\gamma}_0$=2.26782. Similarly, when AR(3) is the instrument to obtain $log\hat{C}_t$ in Table 1c, $\hat{\gamma}_{ar3}$= 1.62983. When AR(4) is used in Table 1d, we have $\hat{\gamma}_{ar4} = 1.24253$. The GMM estimate (Case 1, Table 1a) when NLAG (the number of lags used to form instruments) equals unity is $\hat{\gamma}_{gmm} = 1.24628$. This is close to $\hat{\gamma}_{ar4} = 1.24253$.

Recall from the comments before (7) that good instruments should be (i) highly correlated with the replaced regressors, and (ii) uncorrelated with the errors. In the GMM literature, lagged values of variables are directly used as instruments. The users of GMM software often throw-in many lagged values as instrumental variables, with little attempt to check both requirements of good instruments. Tauchen (1986) notes that Hansen's two-step GMM estimator with lagged endogenous variables as instruments may not "possess moments." He recommends reporting medians and interquartile ranges.

==

TABLE 1b EF estimation (No instrument used)
T=111, df=109, RSS=0.009, DW=1.635, R^2=0.206.
$\hat{\alpha}_0$=intercept=0.00225 (0.7208), $\hat{\beta}=exp(-\hat{\alpha}_0)$=0.99775. Since the transformation $exp(-\hat{\alpha}_0)$ is needed, only Taylor series approximate t-values for $\hat{\beta}$ are possible, $\hat{\gamma}$=2.26782 (5.32).

==

TABLE 1c (Instrument created using AR(3))
T=108, df=106, RSS=0.011, DW=1.624, R^2=0.006
$\hat{\alpha}_0$=intercept=0.00644 (0.4605), $\hat{\beta} = exp(-\hat{\alpha}_0)$=0.99358, $\hat{\gamma}$=1.62983 (0.8241).

==

TABLE 1d (Instrument created using AR(4) model)
T=107, df=105, RSS=0.011, DW=1.619, R^2=0.005
$\hat{\alpha}_0$=intercept=0.00921 (0.729), $\hat{\beta} = exp(-\hat{\alpha}_0)$=0.99084, $\hat{\gamma}$=1.24253 (0.6966).

==

They are not needed here, since our C-CAPM does not suffer from the infinite moment problem. In any case, the GMM strategy for choosing instruments is somewhat *ad hoc*.

Since the GMM routinely involves overidentified models (r>p), it needs a χ^2 test for overidentifying restrictions. Since the observed χ^2=2.3841 for Case 1 in Table 1a is smaller than the tabulated 95 % value of 3.8415 for df=1, we do not reject the overidentifying restrictions of the GMM. However, Tauchen (1986, p. 412) cautions that these χ^2 values are slightly biased toward accepting model specification. The Case 2 of Table 1a does reject the GMM model restrictions. To avoid problems with nonstationarity of variables most researchers follow the current literature. They use consumption growth $C_t = c_t/c_{t-1}$ and dividend growth in the definition of R_t instead of c_t and d_t levels. However, then we have to pay a price: the statistical fits are often poor when level variables are absent. The R^2 values are artificially small and do not imply that the model specification is flawed. Unfortunately, the low R^2 values are rarely reported in the GMM papers. Perhaps, some alternatives to the usual R^2-type criteria are needed.

Tauchen and others use the "reliability" of confidence intervals, as well as, $Var(\hat{\beta}_{gmm})$ for evaluating the reliability of GMM estimation. Plots of simulated sampling distributions in Tauchen seem to be new. They are generally skewed to the right and suggest nonnormality. For EF estimation it is easy to use the standard t-ratios based on $Var(\hat{\beta}_{ef})=\sigma^2(X'\Omega^{-1}X)$, although the nonnormality (skewness) implies that the bootstrap may be better. However, when the residuals are shuffled for a bootstrap, they lose the time subscript and fail to retain original time dependence. Hence recent literature suggests that bootstrap of time series residuals is problematic. Following Vinod (1996a) we use the recursive residuals defined in (29) to develop a proper (iid) bootstrap.

Table 2 reports the bootstrap 95 % confidence intervals for β and γ. Note that our upper limit (UpL) of the interval for β does not exceed unity. By contrast, for NLAG> 2, the GMM estimate $\hat{\beta} > 1$ is economically meaningless, since it implies perverse discounting (negative interest rate). The bias corrected (BC) upper and lower limits move the confidence intervals slightly to the right, consistent with the positive skewness (the mean over 999 resamples exceeds the median). Our UpL for γ is 1.9744, while both GMM UpL's (Table 1a) are very large. Since Mehra and Prescott (1985) concluded that "sensible" values of γ should be less than 2, our intervals are more "sensible" than the GMM's.

Table 2: Bootstrap results based on J=999 resamples

Mean	Median	Low	Up	Low-BC	Up-BC	Coefficient
0.99069	0.99066	0.98630	0.99550	0.98674	0.99633	β
1.2204	1.2196	0.59527	1.8924	0.65285	1.9744	γ

where Low and Up are the lower and upper limits of percentile confidence intervals. BC refers to (median) bias correction. Since the median < mean for both, the estimated sampling distributions are skewed to the right (similar to Tauchen's simulations).

Acknowledgements

I am grateful to Parantap Basu, V. Godambe, Denis Kwiatkowski, Bruce McCullough and two referees for helpful comments. I am especially grateful to Avi Singh for generously taking the time to explain some EF theory, when I presented a version in March 1996 at the Symposium on Estimating Functions at Athens, Georgia.

References

Andrews D. W. K. and J. C. Monahan (1992). An improved heteroscedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* 60 (4), 953-966.

Bhapkar, V. P. (1991). Sufficiency, ancillarity and information in estimating functions: an overview, Ch. 18 in V. P. Godambe (ed.) *Estimating Functions*. Clarendon Press, Oxford.

Christiano, L. J. and M. Eichenbaum (1992). Current real-business-cycle theories and aggregate labor market fluctuations. *Amer. Econ. Rev.* 82,430-450.

Cox, D. R. (1975). Partial likelihood. *Biometrika* 62, 269-276.

Crowder, M. (1986). On consistency and inconsistency of estimating equations. *Econometric Theory* 2, 305-330.

Dhrymes, P.J. (1994). *Topics in Advanced Econometrics.* Vol. II, Springer Verlag, New York.

Dunlop, D. D. (1994). Regression for longitudinal data: A bridge from least squares regression. *Amer. Statistician* 48, 299-303.

Durbin, J. (1960). Estimation of parameters in time-series regression models. *J. of the Roy. Statist. Soc.* Ser. B, 22, 139-153.

Godambe V. P. (1960). An optimum property of regular maximum likelihood estimation. *Ann. of Math. Stat.* 31, 1208-1212.

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

Godambe, V. P. (1985). The foundations of finite sample estimation in stochastic processes. *Biometrika* 72, 419-428.

Godambe, V. P. (1991). Orthogonality of estimating functions and nuisance parameters. *Biometrika* 78, 143-151.

Godambe, V. P. and C.C. Heyde (1987). Quasilikelihood and optimal estimation. *Intern. Statist. Rev.* 55, 231-244.

Godambe, V. P. and B. K. Kale (1991). Estimating functions: an overview, Ch. 1 in V. P. Godambe (ed.) *Estimating Functions.* Clarendon Press, Oxford.

Godambe, V. P. and M. E. Thompson (1989). An extension of quasi-likelihood estimation. *J. Statist. Planning and Inf.* 22, 137-152.

Hall, P. and J. L. Horowitz (1996). Bootstrap critical values for tests based on generalized-method-of-moments estimators. *Econometrica* 64 (4), 891-916.

Hamilton, J. D. (1994). *Time Series Analysis.* Princeton Univ. Press, Princeton, NJ.

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

Hendry, D. F. (1995). *Dynamic Econometrics* Oxford Univ. Press, New York.

Heyde, C. C. and R. Morton (1993). On constrained quasi-likelihood estimation. *Biometrika* 80, 755-761.

Kadane, J. B. (1970). Testing overidentifying restrictions when disturbances are small. *J. Amer. Statist. Assoc.* 65, 182-185.

Kendall, M. and A. Stuart (1979). *The Advanced Theory of Statistics.* MacMillan, Vol. 2, Fourth Ed. New York.

Kianifard, F. and W. H. Swallow (1996). A review of the development and application of recursive residuals in linear models. *J. Amer. Statist. Assoc.* 91, 391-400.

Lele, S. (1991). Resampling using estimating functions. Ch. 22 in V. P. Godambe (ed.) *Estimating Functions.* Clarendon Press, Oxford.

Lindsey, B. (1982). Conditional score functions: some optimality results. *Biometrika* 69, 503-512.

Liang K. and S. L. Zeger (1995). Inference based on estimating functions in the presence of nuisance parameters. *Statis. Sc.* 10, 158-173.

Mehra, R. and E. C. Prescott (1985). The equity premium: A puzzle. *J. of Monetary Econ.* 15, 145-162.

Morton, R. (1981). Efficiency of estimating equations and the use of pivots. *Biometrika* 68, 227-233.

Neyman, J. and E. L. Scott (1948). Consistent estimates based on partially consistent observations. *Econometrica* 16, 1-32.

Ogaki, M. (1993). Generalized method of moments estimation. in G. S. Maddala, C. R. Rao, and H. D. Vinod (eds.) *Handbook of Statistics: Econometrics.* Vol. 11, North Holland Elsevier, New York. Chapter 17, 455-488.

Parzen, M. I., L. J. Wei and Z. Ying (1994). A resampling method based on pivotal estimating functions. *Biometrika* 81, 341-350.

Sawitzski, G. (1994). Testing numerical reliability of data analysis systems. *Computa. Statist. & Data Anal.* 18, 269-286.

Singh, A. C. and R. P. Rao (1997). Optimal instrumental variable estimation for linear models with stochastic regressors using estimating functions. in Godambe V. P (ed) IMS Symposium of Estimating Functions New York.

Small, C. and D.L. McLeish (1988). *The Theory and Applications of Statistical Inference Functions.* Lecture Notes in Statistics No. 44, Springer Verlag, New York.

Sprott, D. A. (1983). Gauss Carl Friedrich, in Kotz and Johnson (eds.) *Encyclopedia of Statistical Sciences.* Vol 3. J. Wiley, New York. 305-308.

Tauchen, G. (1987). Statistical properties of generalized method of moments. *J. Bus. and Econ. Statist.* 4 (4), 397-425.

Thavaneswaran A. (1991). Tests based on an optimal estimate. Ch. 13 in V. P. Godambe (ed.) *Estimating Functions.* Clarendon Press, Oxford.

Vijayan, K. (1991). Estimating functions in survey sampling: Estimation of super-population regression parameters. Ch. 17 in V. P. Godambe (ed.) *Estimating Functions.* Clarendon Press, Oxford.

Vinod, H.D. and Ullah, A. (1981). *Recent Advances in Regression Methods.* Marcel Dekker, New York.

Vinod, H. D. (1984). Distribution of a generalized t-ratio for biased estimators. *Econ. Letters.* 14, 43-52.

Vinod, H. D. (1989). Resampling fuzzy data and Latin Squares: Application to regression. *1989 Proc. Bus. & Econ. Sec.* Amer. Statist. Assoc. Washington, D.C. 304-309.

Vinod, H.D. (1993). Bootstrap methods: Applications in econometrics. in G. S. Maddala, C. R. Rao, and H. D. Vinod (eds.) *Handbook of Statistics: Econometrics.* Vol. 11, North Holland, Elsevier, New York. Chap. 23, 629-661.

Vinod, H.D. (1995). Double bootstrap for shrinkage estimators. *J. of Econometrics* 68, 287-302.

Vinod, H.D. (1996a). Comments on "Bootstrapping time series Models". *Econometric Reviews* 15 (2), 183-190.

Vinod, H.D. (1996b). *Foundations of Asymptotic Inference Using Modern Computers.* Economics Dept., Fordham University, New York, discussion paper, Sept. 1996.

Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models and the Gaussian method. *Biometrika* 61, 439-447.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* 48, 817-838.