Durham, Geweke. Published in the Journal of Business & Economic Statistics, 21(4). October 2003. 490-492.

Comment

Garland DURHAM and John GEWEKE
Department of Economics, University of Iowa, Iowa City, IA 52242-1000
({*garland-durham, john-geweke*}*@uiowa.edu*)

This article provides a comprehensive overview and integration of state-of-the-art econometric methods for models that are naturally stated in terms of latent variables but present significant practical problems for inference from data. In so doing it extends these methods in significant ways by incorporating the important concept of backfitting. It shows explicitly how this extension applies to GMM and ML estimators. The specific problems inherent in estimating affine pricing models of the term structure motivate much of the article and provide its illustrative applications. Along the way, the article treats a wide variety of related problems. These comments focus on the main theme leading to the empirical application, which is the development of IS-GMM and the extension to backfitting in Sections 3 and 4.

This article addresses circumstances in which moment conditions

$$E^0[\psi(Y_t^*, \theta)] = 0 \quad \Longleftrightarrow \quad \theta = \theta^0 \tag{1}$$

are most naturally formulated in terms of latent variables Y_t^* rather than observable counterparts Y_t. (As in the article, the superscript "0" indicates the value of the parameter or moment in the data-generating process, which in turn presumes correct

specification of the moment conditions and model.) Motivated by asset pricing models, it takes up the case in which the latent and observable variables are linked by a relation

$$Y_t = g(Y_t^*, \lambda^0), \qquad Y_t^* = g^{-1}(Y_t, \lambda^0). \tag{2}$$

This relation is completely determined by the unknown parameter vector θ, so that $\lambda = \lambda(\theta)$ and, in particular, $\lambda^0 = \lambda(\theta^0)$. Defining

$$\phi[Y_t, \theta, \lambda(\theta)] = \psi\{g^{-1}[Y_t, \lambda(\theta)], \theta\},$$

(1) becomes

$$E^0\{\phi[Y_t, \theta, \lambda(\theta)]\} = 0. \tag{3}$$

An attraction of the article is that it treats extremum estimators in general. This includes the GMM estimator based on (3) set forth at the start of Section 3.2, with the criterion function

$$Q_T(\theta, \lambda) = -\left[T^{-1}\sum_{t=1}^{T}\phi(Y_t, \phi, \lambda)\right]' W_T \left[T^{-1}\sum_{t=1}^{T}\phi(Y_t, \phi, \lambda)\right]. \tag{4}$$

It also includes the MLE that works with the criterion function

$$Q_T(\theta, \lambda) = T^{-1}\sum_{t=1}^{T}\ell^*\{g^{-1}(Y_t, \lambda)|g^{-1}(Y_{t-1}, \lambda); \theta\} + T^{-1}\sum_{t=1}^{T}\log|J_y g^{-1}(Y_t, \lambda)|, \tag{5}$$

in which ℓ^* denotes the log-pdf of the latent Y_t^* and J_y is the Jacobian of transformation. Sections 3 and 4 contrast three different ways of handling the identity $\lambda = \lambda(\theta)$ in (4) or (5). The *infeasible oracle estimator*, introduced at the start of Section 3.2, is

$$\widehat{\theta}_T^* = \arg\max_{\theta\in\Theta} Q_T(\theta, \lambda^0),$$

and the asymptotic variance of this estimator is presented in the same paragraph.

The *IS-GMM estimator* is

$$\widehat{\theta}_T^{IS} = \arg\max_{\theta\in\Theta} Q_T[\theta, \lambda(\theta)], \tag{6}$$

introduced in the next paragraph, and its asymptotic variance is the semiparametric efficiency bound indicated in equation (3.3) in the article.

The *IS-GMM backfitting estimator* is $\widehat{\theta}_T$, the limit (in p) of the sequence

$$\theta_T^{(p)} = \arg\max_{\theta\in\Theta} Q_T\left[\theta, \lambda\left(\theta_T^{(p-1)}\right)\right]. \tag{7}$$

The formulation (7) is first presented in Section 3.2, after (3.7). If one defines

$$\overline{\theta}_T(\lambda) = \arg\max_{\theta\in\Theta} Q_T(\theta, \lambda),$$

which the article does in (3.7), then (7) can be expressed as

$$\theta_T^{(p)} = \overline{\theta}_T\left[\lambda\left(\theta_T^{(p-1)}\right)\right]. \tag{8}$$

Computation of $\widehat{\theta}_T^{IS}$ entails evaluating derivatives of $\lambda(\theta)$ [see (6)] whereas computation of $\widehat{\theta}_T$ requires only evaluation of the function $\lambda(\theta)$ itself [see (7)]. Because $\lambda(\theta)$ is typically complicated, this is a significant practical advantage of backfitting. For example, in ML the Jacobian term appearing in the last term of (5) involves only λ. Whereas $\widehat{\theta}_T^{IS}$ must contend directly with the shape of this term in each iteration to maximum, $\widehat{\theta}_T$ needs to evaluate it only once each step.

The article defines $\widehat{\theta}_T$ in (4.1) with reference to an iteration-stopping rule $p(T)$, but it is simpler to work with the limit the sequence in (8), and in any event conventional convergence criteria rather than a fixed function $p(T)$ are used in the application in Section 6. The article develops conditions for the convergence of $\theta_T^{(p)}$ in p to the estimator $\widehat{\theta}_T$ that in turn guarantee weak consistency (Prop. 2). The most important of these, in our view, is the contraction mapping of Assumption 6. Proposition 3 provides the asymptotic variance of $\widehat{\theta}_T$.

The contrast between $\widehat{\theta}_T^{IS}$ and $\widehat{\theta}_T$ is important in understanding the article's contribution, in interpreting either estimate in practice, and in choosing an estimator in an empirical application. The article states that, at least typically, $\widehat{\theta}_T^{IS}$ and $\widehat{\theta}_T$ coincide so long as (8) has a fixed point [Sec. 3.2, end of the paragraph including (3.7)]. If that were the case, then these estimators would have the same asymptotic variance, but the article demonstrates that these variances are different.

We find that a geometric interpretation of the estimators $\widehat{\theta}_T^{IS}$ and $\widehat{\theta}_T$ is helpful in appreciating these differences. Figure 1 portrays what is perhaps the simplest possible situation. The concentric circles in the figure represent level curves of $Q_T(\theta, \lambda)$; in the case portrayed, there is no interaction between θ and λ in $Q_T(\theta, \lambda)$. The positively sloped straight line represents the identity $\lambda = \lambda(\theta) = a\theta$. The function $\overline{\theta}_T(\lambda)$ is the

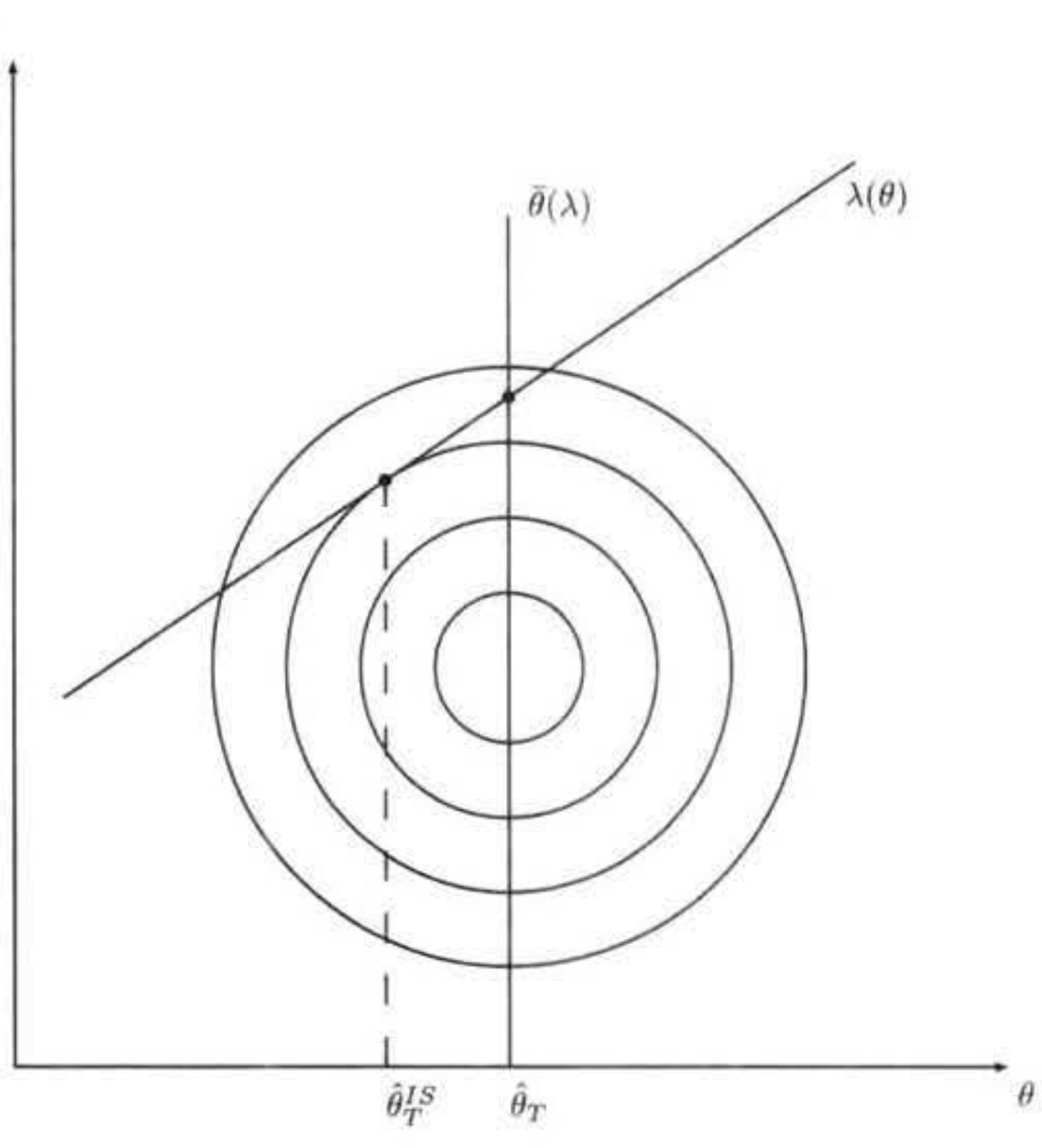

Figure 1.

locus of tangencies of horizontal lies to the concentric circle level curves, and therefore is the vertical line through the center of these circles portrayed in Figure 1. The estimator $\widehat{\theta}_T^{IS}$ is the abscissa of the point of tangency between this line and the level curves. The sequence $\theta_T^{(p)}$ in this example converges in exactly two steps. From any initial value $\theta^{(0)}$ on the abscissa, the first step in (8) maps vertically to $\lambda_T^{(1)} = \lambda(\theta^{(0)})$ and then horizontally to $\theta_T^{(1)} = \overline{\theta}_T(\lambda_T^{(1)})$. The second step maps vertically to $\lambda_T^{(2)} = \lambda(\theta_T^{(1)})$ and then finds that $\theta_T^{(2)} = \overline{\theta}_T(\lambda_T^{(1)})$. Because $\theta_T^{(2)} = \theta_T^{(1)}$, the iterations have converged, and in fact $\widehat{\theta}_T = \theta_T^{(2)} = \theta_T^{(1)}$. This example illustrates a number of general points about $\widehat{\theta}_T^{IS}$ and $\widehat{\theta}_T$.

Obviously the estimators are not the same, very special cases aside. If $\lambda(\theta)$ were horizontal, [i.e., $\lambda(\theta) = \lambda^* \, \forall \theta \in \Theta$], then $\widehat{\theta}_T^{IS} = \widehat{\theta}_T$. This, in turn, is a case in which there is no nonadaptivity problem and illustrates the article's points about the importance of nonadaptivity in comparing the asymptotic distributions of $\widehat{\theta}_T^{IS}$ and $\widehat{\theta}_T$. This very special case suggests that in any application in which λ is not very responsive to θ and there is little interaction between θ and λ in the criterion function (4) or (5), backfitting may achieve results close to IS-GMM, with the advantage of substantial computational efficiency. On the other hand, given $\lambda = \lambda(\theta) = a\theta$, a quadratic criterion function and a Gaussian data-generating process, it is straightforward to show, in the context of Figure 1, that $\text{var}(\widehat{\theta}_T)/\text{var}(\widehat{\theta}_T^{IS}) = 1 + a^2$. This raises general questions about efficiency loss due to backfitting that might be investigated in more detail in future work.

If the model underlying the hypothetical situation in Figure 1 is specified correctly, then the center of the concentric circles will move, stochastically, toward the line $\lambda(\theta)$ as sample size T increases. Because both estimators are consistent, $\widehat{\theta}_T^{IS} - \widehat{\theta}_T \xrightarrow{p} 0$, but $T^{-1/2}(\widehat{\theta}_T^{IS} - \widehat{\theta}_T)$ will have a nondegenerate limiting distribution whose variance will depend on the relative orientation of $\lambda(\theta)$ and the level curves, as discussed in the previous paragraph. We would expect differences relative to standard errors to persist. (In the application in the article, it seems to us that this is the case, but the differences are not large, suggesting that the limiting case $a = 0$ might be an idealized, rough approximation in this application.) On the other hand, if the model underlying the situation portrayed in Figure 1 is misspecified, then the center of the circles will not, in general, converge to a point on $\lambda(\theta)$. The estimators $\widehat{\theta}_T^{IS}$ and $\widehat{\theta}_T$ will converge to *different* pseudotrue values, and in the metric of the standard error of either one, differences between them will grow.

Generalizing Figure 1 to the case in which level curves are ellipses rather than circles is informative in illustrating some of the other points in Sections 3 and 4 of the article. The function $\overline{\theta}(\lambda)$ remains a straight line, but is no longer vertical. As long as $\lambda(\theta)$ and $\overline{\theta}(\lambda)$ have different slopes, there will be exactly one fixed point. The sequence $\{\theta_T^{(p)}\}$ either converges toward this point or diverges from it, depending on whether the contraction mapping conditions of Assumption 6 are violated. With linear $\lambda(\theta)$ and $\overline{\theta}(\lambda)$, these conditions reduce to simple inequalities involving the respective slopes of the two functions.

Relevant applications, including those in asset pricing, do not, of course, have the simplicity of either Figure 1 or this mild extension. The parameters θ and λ are vectors, not scalars, and so two-dimensional diagrams cannot represent the situation adequately. More important, in our view, is the fact that the functions $\lambda(\theta)$ and $\overline{\theta}(\lambda)$ are nonlinear in interesting applications, including those presented in Section 6 of the article. This opens up possibilities like the very simplified one portrayed in Figure 2. There are two fixed points, (θ_1, λ_1) and (θ_2, λ_2). The first point satisfies the contraction mapping condition, but the second point does not. Over the range portrayed in Figure 2, if $\theta_T^{(0)} > \theta_2$ then $\lim_{p\to\infty} \theta_T^{(p)} = \theta_1$, whereas if $\theta_T^{(0)} < \theta_2$, then $\theta_T^{(p)}$ rapidly diverges downward (to the left in Fig. 2).

More important in Figure 2 is the fact that no level contours of Q_T have been included. In part, this was done to keep that illustration simple, but it is also to allow the reader to sketch level curves of his or her own, verifying that in this situation, differences between $\widehat{\theta}_T^{IS}$ and $\widehat{\theta}_T$ could be extremely great, and $\widehat{\theta}_T$ could be misleading. A reliable empirical application of backfitting must rule out situations like the one portrayed in Figure 2. In the empirical work reported in Section 6 of the article, the contraction mapping conditions of Assumption 6 do not hold, and multiple stationary points of $\{\theta_T^{(p)}\}$ were found. The selected $\widehat{\theta}_T$ (Tables 5 and 6) bear the interpretation of not differing drastically from $\widehat{\theta}_T^{IS}$ (Tables 3 and 4). Such comparisons will not always be possible, however, if this research program realizes the goal of using backfitting for inferences in situations impracticable for IS-GMM. Identification of this critical area is one of the article's important contributions.

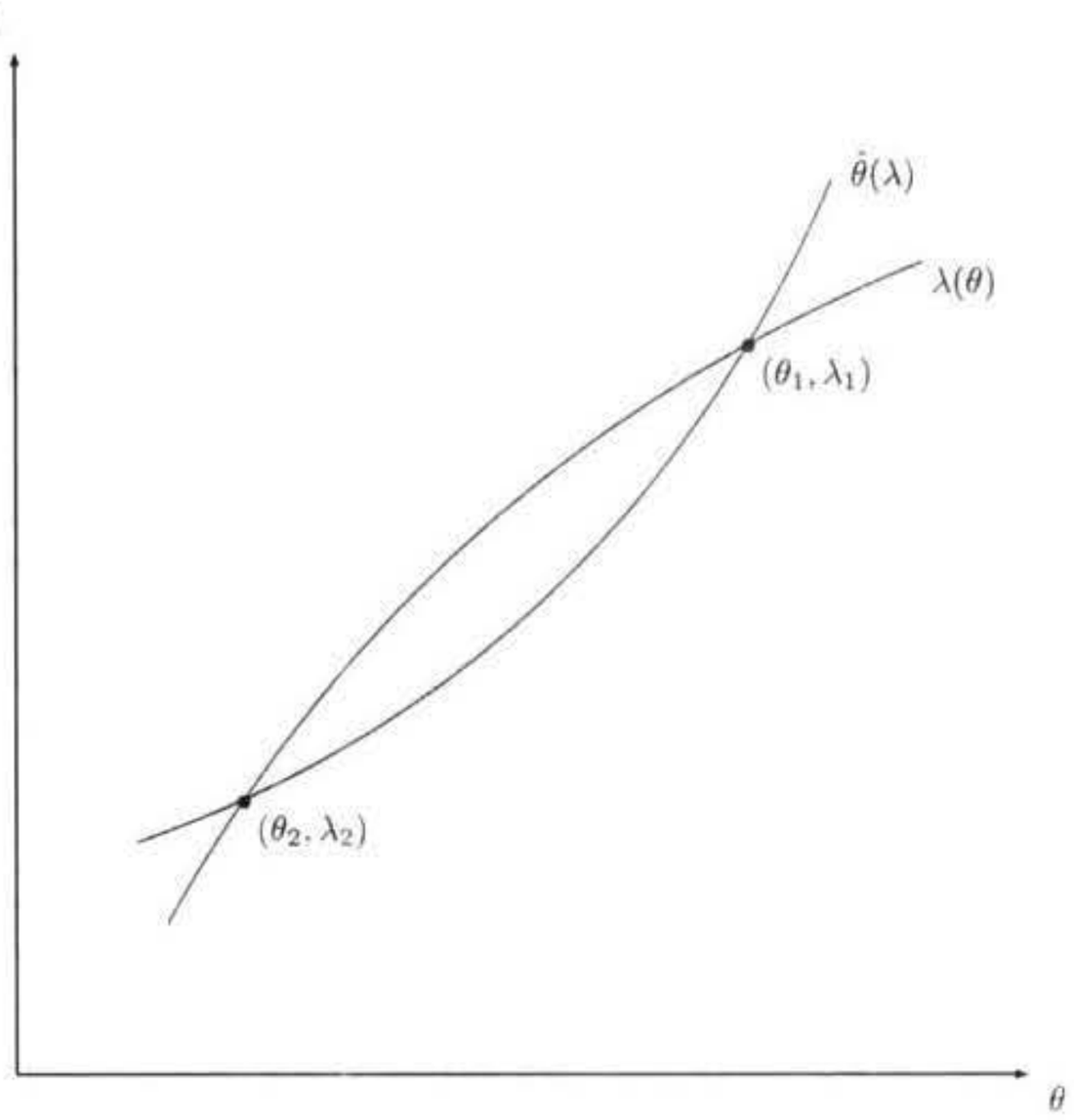

Figure 2.