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.
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
\[ Y_t = g(Y_t^*, \lambda^0), \quad Y_t^* = g^{-1}(Y_t, \lambda^0). \] (2)
This relation is completely determined by the unknown para-

\[ \text{Definite -} \]

\[ \text{meter vec tor (}, \), so that
\[ \text{by asset pricing models. it takes up the case in which the latent}
\[ \text{different ways of handling the identity}
\[ \text{The formulation (7) is first presented in Section 3.2, after (3.7).}
\[ \text{It also includes the MLE that works with the criterion function}
\[ Q_T(\theta, \lambda) = T^{-1} \sum_{t=1}^{T} \ell^*[g^{-1}(Y_t, \lambda)]g^{-1}(Y_{t-1}, \lambda); \theta] \]
\[ + T^{-1} \sum_{t=1}^{T} \log |J_y g^{-1}(Y_t, \lambda)|. \] (5)
in which \( \ell^* \) denotes the log-pdf of the latent \( Y_t^* \) and \( J_y \)
\[ \text{is the Jacobian of transformation. Sections 3 and 4 contrast three}
\[ \text{different ways of handling the identity } \lambda = \lambda(\theta) \text{ in (4) or (5).}
\[ \text{The infeasible oracle estimator, introduced at the start of Sec-
\[ \text{tion 3.2, is}
\[ \theta_T^* = \arg \max_{\theta \in \Theta} Q_T(\theta, \lambda^0), \]
\[ \text{and the asymptotic variance of this estimator is presented in the}
\[ \text{same paragraph.}
\[ \text{The IS-GMM estimator is}
\[ \hat{\theta}_T^{IS} = \arg \max_{\theta \in \Theta} Q_T(\theta, \lambda(\theta)), \] (6)
\[ \text{introduced in the next paragraph, and its asymptotic variance is}
\[ \text{the semiparametric efficiency bound indicated in equation (3.3)}
\[ \text{in the article.}
\[ \text{The IS-GMM backfitting estimator is } \hat{\theta}_T, \text{ the limit (in } p \text{) of the}
\[ \text{sequence}
\[ \hat{\theta}_T^{(p)} = \arg \max_{\theta \in \Theta} Q_T[\theta, \lambda(\theta_T^{(p-1)})]. \] (7)
\[ \text{The formulation (7) is first presented in Section 3.2, after (3.7).}
\[ \text{If one defines}
\[ \hat{\theta}_T(\lambda) = \arg \max_{\theta \in \Theta} Q_T(\theta, \lambda),
\[ \text{which the article does in (3.7), then (7) can be expressed as}
\[ \hat{\theta}_T^{(p)} = \hat{\theta}_T[\lambda(\theta_T^{(p-1)})]. \] (8)

\[ \lambda \]

\[ \hat{\theta}(\lambda) \]

\[ \theta \]

\[ \lambda(\theta) \]

\[ \hat{\theta}_T^{IS} \]

\[ \hat{\theta}_T \]

\[ \hat{\theta}_T(\lambda) \]

\[ \hat{\theta}_T^{(p)} \]

\[ \hat{\theta}_T[\lambda(\theta_T^{(p-1)})]. \]

\[ \text{Computation of } \hat{\theta}_T^{IS} \text{ entails evaluating derivatives of } \lambda(\theta) \text{ [see (6)] whereas computation of } \hat{\theta}_T \text{ requires only evaluation of}
\[ \text{the function } \lambda(\theta) \text{ itself [see (7)]. Because } \lambda(\theta) \text{ is typically}
\[ \text{complicated, this is a significant practical advantage of back-
\[ \text{fitting. For example, in ML the Jacobian term appearing in the}
\[ \text{last term of (5) involves only } \lambda. \text{ Whereas } \hat{\theta}_T^{IS} \text{ must contend di-
\[ \text{rectly with the shape of this term in each iteration to maximum,}
\[ \hat{\theta}_T \text{ needs to evaluate it only once each step.}
\[ \text{The article defines } \hat{\theta}_T \text{ in (4.1) with reference to an iteration-
\[ \text{stopping rule } p(T), \text{ but it is simpler to work with the limit the}
\[ \text{sequence in (8), and in any event conventional convergence cri-
\[ \text{teria rather than a fixed function } p(T) \text{ are used in the application}
\[ \text{in Section 6. The article develops conditions for the conver-
\[ \text{gence of } \hat{\theta}_T^{(p)} \text{ in } p \text{ to the estimator } \hat{\theta}_T \text{ that in turn guarantee weak}
\[ \text{consistency (Prop. 2). The most important of these, in our view,}
\[ \text{is the contraction mapping of Assumption 6. Proposition 3 pro-
\[ \text{vides the asymptotic variance of } \hat{\theta}_T.\]
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 $\hat{\theta}^{(p)}_T$ 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 $\theta_T^{(0)}$ on the abscissa, the first step in (8) maps vertically to $\lambda_T^{(1)}(\theta_T^{(0)})$ and then horizontally to $\bar{\theta}_T^{(1)} = \theta_T(\lambda_T^{(1)})$. The second step maps vertically to $\lambda_T^{(2)}(\bar{\theta}_T^{(1)})$ and then finds that $\theta_T^{(2)} = \theta_T(\lambda_T^{(2)})$. Because $\theta_T^{(2)} = \bar{\theta}_T^{(1)}$, the iterations have converged, and in fact $\theta_T = \theta_T^{(2)} = \theta_T^{(1)}$. This example illustrates a number of general points about $\bar{\theta}_T^{IS}$ and $\bar{\theta}_T$.

Obviously the estimators are not the same, very special cases aside. If $\lambda(\theta)$ is horizontal, i.e., $\lambda(\theta) = \lambda$ for all $\theta \in \Theta$, then $\bar{\theta}_T^{IS} = \bar{\theta}_T$. 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 $\bar{\theta}_T^{IS}$ and $\bar{\theta}_T$. This very special case suggests that in any application in which $\lambda$ is not very responsive to $\theta$ and there is little interaction between $\theta$ and $\lambda$ 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 $\lambda = \lambda(\theta) = \alpha \theta$, a quadratic criterion function and a Gaussian data-generating process, it is straightforward to show, in the context of Figure 1, that $\text{var}(\bar{\theta}_T) / \text{var}(\bar{\theta}_T^{IS}) = 1 + \alpha^2$. 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, $\bar{\theta}_T^{IS} - \bar{\theta}_T^{P} \rightarrow 0$, but $T^{-1/2}(\bar{\theta}_T^{IS} - \bar{\theta}_T)$ 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 $\alpha = 0$ 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 $\bar{\theta}_T^{IS}$ and $\bar{\theta}_T$ 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 $\tilde{\lambda}(\lambda)$ remains a straight line, but is no longer vertical. As long as $\lambda(\theta)$ and $\tilde{\lambda}(\lambda)$ 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 $\tilde{\lambda}(\lambda)$, these conditions reduce to simple inequalities involving the respective slopes of the two functions.

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 $\bar{\theta}_T^{IS}$ and $\bar{\theta}_T$ could be extremely great, and $\bar{\theta}_T$ 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 $\bar{\theta}_T$ (Tables 5 and 6) bear the interpretation of not differing drastically from $\bar{\theta}_T^{IS}$ (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](image-url)