Comment on "Iterative and Recursive Estimation in Structural Non-Adaptive Models" by Sergio Pastorello, Valentin Patilea and Eric Renault

Jun Pan*

May 29, 2003

This paper constitutes a serious attempt to develop a new inference method for structural models with latent variables. In addition to tackling some difficult econometric issues, it is carefully executed with thoughtful discussions on many detailed issues. Overall, I see this article as a nice addition to the literature that focuses on the development of efficient and feasible econometric methods for structural models. My comments will revolve around one very basic question: as a practicing econometrician, under what situation will I choose the proposed *latent backfitting method* over existing approaches?

1 The Setup

I will start with a relatively general setup that includes some, if not all, of the motivating examples used in this paper. In many ways, I am simply re-stating the general framework

*MIT Sloan School of Management, junpan@mit.edu.

of the paper, but with one difference that will become important later. Let X be a vector of state variables, and let θ be a set of constant parameters associated with the data-generating process of the state variables X. Let λ be a set of constant parameters that control the market prices of risk present in X, and let Y be a vector of market observables. Given that this paper puts a lot of emphasis on efficiency gain due to the one-to-one relationship between latent state variables X and observed asset prices Y, I will focus most of my discussion on the case where the number of state variables equals the number of observables. More specifically, I'll assume that conditions hold so that there is a one-to-one mapping between Y and X:

$$Y_t^{(i)} = g^{(i)}(X_t, \theta, \lambda) \,, \quad i = 1, 2, \ldots, n \,, \tag{1}$$

where n is the dimension of X.

This setup is of interest for a variety reasons, the most interesting (at least in my opinion) of which originates from the increasing availability of time-series data on derivative securities such as options, swaps, swaptions, and credit derivatives. For every derivative security, we also have time-series data on the underlying or related securities: stocks for stock options, swaps for swaptions, corporate bonds for credit default swaps, etc. The above setup provides an ideal framework to incorporate time-series data from various markets that are exposed to the same fundamental risk factors. For example, we can place both stock and option prices in Y, and model the risk factors affecting the stock market explicitly in X. By incorporating them in one integrated system, we can ask some important economic questions. For example, how are the risk factors in X priced in the financial assets Y? In other words, in addition to knowing θ, which characterizes the dynamics of the fundamental risk factors, we also know λ, which tells us about the market price of such risk factors. By reason of familiarity, one such example that comes to mind is Pan (2002), which incorporates the time-series data on the S&P 500 index and options in a unified framework and estimates stock-price dynamics and market prices of associated risk factors.

2 MLE and IS-GMM

Now let's think about the best estimation method for such a setup. As usual, maximum-likelihood estimation (MLE) is the gold. If the dynamics of X are specified in such a way that the conditional transition density $f(X_t \,|\, X_{t-1})$ and the pricing relation in equation (1) are known in closed form, then the conditional transition density of Y is readily available and MLE is our last stop.

Here is where feasibility becomes an issue. For a wide range of empirical problems, the demand for rich and flexible dynamics of the state variables typically results in sacrificing tractability to the extent that the transition density cannot be derived in closed form. This is the case for the general affine jump-diffusion models (Duffie and Kan (1996)), which have found a wide range of applications in asset pricing. There is, however, still tractability. For example, although the conditional density is not known explicitly, the conditional transform (including characteristic function and moment-generating function) of affine jump-diffusions can be calculated with considerable tractability (Duffie, Pan, and Singleton (2000)). This naturally leads one to focus more directly on the dynamics of X and the particular tractability it offers. In other words, given that MLE is not feasible, let's settle for the second best and write down moment conditions (as efficient as possible) using the analytical tractability in X.

To a large extent, this is the motivative behind the "implied-state" generalized method of moments (IS-GMM) proposed by Pan (2002). Suppose, for the moment, that the state variables X can be observed directly, and let's start with moment conditions of the form:

$$E_{t-1}^{\theta_0}\left[h\left(X_t,\theta_0\right)\right] = 0\,, \tag{2}$$

where θ_0 are the set of true parameters that govern the actual dynamics of X, and where the moment conditions are built to take advantage of the analytical tractability in X. For example, if the joint moment-generating function of X is known in closed form, then the test function h can be chosen to include the first n moments of X. Moreover, efficient instruments, which typically involve even higher moments, can be calculated using Hansen (1985).

The problem is that the state variables X are not directly observable. At the same time, however, our setup tells us that X_t are not really latent in the sense that they can be backed out from the observables Y through the pricing relation (1), given that we know the true parameters θ_0 and λ_0. Of course, the problem is that we don't know θ_0 or λ_0, which are, in fact, the objects of our investigation. The idea of IS-GMM comes from the observation that for any given set of θ and λ, we can always try to invert X_t from Y_t using the pricing relation (1). So assuming the inversion holds, for every θ and λ, we have a vector of implied state variables $X_t^{\theta,\lambda}$, which might have nothing to do with the true state variables X_t. When evaluated at the true parameters θ_0 and λ_0, however, the implied state variables become the true state variables:

$$X_t^{\theta_0,\lambda_0} = X_t\,. \tag{3}$$

Taking advantage of the unique information embedded in the implied state variables, Pan (2002) proposes IS-GMM estimators by replacing the state variables X_t with the implied state variables $X_t^{\theta,\lambda}$. Clearly, the moment condition is still of the form (2), but the sample analogue of the moment condition becomes

$$G_N(\theta,\lambda) = \frac{1}{N}\sum_{t=1}^{N} h\left(X_t^{\theta,\lambda},\theta\right) , \tag{4}$$

which involves implied state variables $X_t^{\theta,\lambda}$, not the true state variables.[1]

3 Latent Backfitting Method

The latent backfitting method proposed by Sergio Pastorello, Valentin Patilea and Eric Renault is designed more or less for the same setup summarized in Section 1. The application of this method could be more general, but my comment will focus on the stated setup as I found it to be the most interesting application.

My first observation is the restriction imposed on λ. To be more specific, this paper assumes that the set of parameters λ is a function of θ. I found this to be quite puzzling because this restriction excludes most of the motivating examples discussed in the paper. For example, in the option pricing application, the market prices of various factors (volatility risk or jumps) do not show up in the stock price dynamics. So by definition, these important parameters do not belong to θ, nor can they be any functions of θ. We have a similar situation in the state-space term-structure estimation: the λ's are only present in the pricing relation (1) but not in the data-generating process of X. Given the severity of this restriction, I am willing to give the authors the benefit of doubt and assume that their θ is big enough to include both the θ and λ. In other words, let's assume that there are parameters λ that show up only in the pricing relation but not the data-generating process of X.

[1] See Pan (2002) for the large-sample properties of IS-GMM estimators.

With this assumption in mind, I am puzzled by my second observation: how can the latent backfitting method identify λ? If I understand the paper correctly, the backfitting method adopts a recursive scheme. In the first step, choose some initial values θ^1 and λ^1 and use them to back out the state variables $X_t^{\theta^1,\lambda^1}$. In the second step, take the state variables backed out from the first step to the moment condition (2) and obtain GMM (or MLE) estimators for θ, with the assumption that the backed-out state variables are the true state variables. The recursion happens when one feeds the second-stage estimates θ^2 to the pricing relation to back out an updated version of the state variables $X_t^{\theta^2,\lambda^1}$. My problem with this approach is that λ never gets updated in this recursive scheme because it does not show up in the moment condition.[2]

This leads to my third puzzle: how can IS-GMM estimators be the limiting case of the latent backfitting estimators? In IS-GMM, the identification of λ comes from the its influence on the implied state variables $X^{\theta,\lambda}$, which in turn show up in the sample analogue $G_N^{\theta,\lambda}$ of the moment condition. To be more explicit, the IS-GMM obtains identification of λ through

$$\frac{\partial}{\partial \lambda} G_N^{\theta,\lambda} = \frac{1}{N} \sum_{t=1}^{N} h_x \left(X_t^{\theta,\lambda}, \theta \right) \frac{\partial X_t^{\theta,\lambda}}{\partial \lambda}\,, \tag{5}$$

where h_x denotes the derivative of h with respect to the state variables X. The key to identifying λ in IS-GMM is $\partial X_t^{\theta,\lambda}/\partial\lambda$: the parameter dependence of the implied state variables. In latent backfitting, however, the backed-out state variables are assumed to be observed. As a result, there is no $\partial X_t^{\theta,\lambda}/\partial\lambda$ in latent backfitting. Unless I am missing something important, I am not convinced that the latent backfitting method converges to IS-GMM in the limit.

[2] Of course, this is not a problem if λ is indeed a function of θ. This, however, is not the case for most of the interesting empirical applications considered in literature, including the empirical implementation considered in this paper!

At least, this does not seem to hold true for the case where λ shows up only in the pricing relation but not the data-generating process of X.

My last puzzle comes from reading Section 6 on empirical implementation issues. There, we have an example where λ does not show up in data-generating process. So according to my puzzles 2 and 3, I would expect λ to be unidentifiable using the latent backfitting method. Nevertheless, we have results showing that the latent backfitting method produces satisfactory results for λ. In the next section, I will try to reconcile these two conflicting observations.

4 Empirical Implementation Issues

In the empirical implementation, X is a one-factor CIR or Vasicek process, and Y are zero-yields of varying maturities. The setup we've considered so far does not hold in this example because there are four observables, but only one latent state variable.

To understand the source of identification of λ, let's consider an illustrative example. Let's assume that only two yields $Y_t^{(1)}$ and $Y_t^{(2)}$ are used for estimation. The first yield is used to back out the state variable X while the second one is assumed to be measured with error. The following linear pricing relation holds for zero yields in a one-factor CIR or Vasicek model,

$$Y_t = A + BX_t \, ,$$

where A and B are constant coefficients that depend on the maturity of the yield, θ, and, most importantly, λ.

In the first step of the latent backfitting method, we fix some parameters θ^1 and λ^1 to calculate the associated A and B and then back out the state variable X by

$$X_t = \frac{Y_t^{(1)} - A_1}{B_1} ,$$

where we use A_1 and B_1 to denote the coefficients for the maturity associated with $Y_t^{(1)}$. In the second step, we assume that the backed-out X_t is actually observable and obtain second-stage estimates θ^2 and λ^2 using the moment condition that derives from the likelihood function

$$f(X_t|X_{t-1},\theta)\frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left(-\frac{\left(Y_t^{(2)} - A_2 - B_2 X_t\right)^2}{2\sigma_e^2}\right) , \tag{6}$$

where f is the conditional density of X (Gaussian for Vasicek and non-central χ^2 for CIR), and where σ_e is the standard deviation of the pricing error associated with $Y_t^{(2)}$.

It is important to note that in equation (6), the first piece, $f(X_t|X_{t-1},\theta)$, comes from the actual dynamics of the state variable X, and the second piece comes from the pricing error associated with $Y^{(2)}$. To be more specific, the first piece takes advantage of the dynamic property of X, which is the key motivation behind an approach like IS-GMM, while the second piece has nothing to do with the dynamic property of X. Instead, it takes advantage of the cross-sectional information in the data, which is valuable when the number of observables is greater than the number of state variables.[3] As discussed in Section 3, the first piece does not provide identification for λ in the latent backfitting approach. By introducing the pricing relation through the pricing error, the second piece, however, does provide identification for λ. Specifically, A_2 depends on λ in a Vasicek model, and both A_2 and B_2 depend on λ in a CIR model.

[3]This pricing error approach has been used in a wide range of applications. See, for example, Chen and Scott (1993) for a term-structure application and Pan (2002) for an option application.

I guess this is how the conflicting observations are reconciled. The empirical implementation of this paper uses one yield (or linear combination of the four yields) to back out the state variable X, and then takes advantage of the cross-sectional information embedded in the other yields through pricing errors. The moment condition (to be more precise, the score of the likelihood function) is built not only from the dynamics of the state variable X, but also from the pricing errors associated with the "surplus" yields. The identification of λ is achieved from the latter addition to the moment condition.

It is clearly a good practice to use as many liquid market prices as possible, and to use both the time-series and cross-sectional information in the data. I would like to point out, however, the potential pitfall of the proposed recursive approach. In particular, when the number of usable market observables just equals the number of state variables, the latent backfitting method has an identification problem for parameters like λ, which shows up only in the pricing relation but not in the data-generating process. At least for this special case, the latent backfitting method does not converge to IS-GMM, which does not suffer from the identification problem of λ.

Let me close by answering the initial question I raised: under which condition will the latent backfitting method be the chosen estimation method? Clearly, we would have to exclude the case we just mentioned. Putting this aside, my view of this recursive scheme is that it does provide some computational tractability by focusing one issue at a time. When I was working on Pan (2002), in my computer programs I had a flag that would turn off the parameter dependence of X, which corresponds to the second step of latent backfitting method. As an intermediate try, I would use this flag to get a quick estimation of θ and then use it as initial parameter values for the IS-GMM estimation. At no point, however, did I trust the estimation results with the parameter dependence of X turned off. First, there is identification problem for λ for the case where the number of observables equals the number of state variables. Second, the desired fixed point is ever so eluding when not much pressure is enforced. If IS-GMM is the end point where things converge, why not use IS-GMM (when MLE is not feasible)?

References

Chen, R. and L. Scott (1993). Maximum Likelihood Estimation for a Multifactor Equilibrium Model of the Term Structure of Interest Rates. *The Journal of Fixed Income 3*, 14–31.

Duffie, D. and R. Kan (1996). A Yield-Factor Model of Interest Rates. *Mathematical Finance 6*(4), 379–406.

Duffie, D., J. Pan, and K. Singleton (2000). Transform Analysis and Asset Pricing for Affine Jump-Diffusions. *Econometrica 68*, 1343–1376.

Hansen, L. P. (1985). A Method for Calculating Bounds on the Asymptotic Covariance Matrices of Generalized Method of Moments Estimators. *Journal of Econometrics 30*, 203–238.

Pan, J. (2002). The Jump-Risk Premia Implicit in Option Prices: Evidence from an Integrated Time-Series Study. *Journal of Financial Economics 63*, 3–50.