Statistical Science
2011, Vol. 26, No. 1, 47–48
DOI: 10.1214/11-STS345A
Main article DOI: 10.1214/10-STS345

Discussion of "Feature Matching in Time Series Modeling" by Y. Xia and H. Tong

Bruce E. Hansen

1. INTRODUCTION

Xia and Tong have written a provocative and stimulating paper. Among the many topics raised in their paper, I would like in particular to endorse several of their postulates:

1. All models are wrong.
2. Observations are not error-free.
3. Estimation needs to account for the above two issues.

As described in the paper, suppose that we observe a process $\{y_t : t = 1, \ldots\}$ for which we have a model $\{x_t(\theta) : t = 1, \ldots\}$ which depends upon an unknown parameter θ. Let $F_x(\theta)$ denote the joint distribution of the $x_t(\theta)$ process and F_y the joint distribution of the observables. When we say that the model is wrong, we mean that there is no θ such that $F_x(\theta) = F_y$. If we think of the distribution F_y as a member of a large space of potential joint distributions, then the set of joint distributions $F_x(\theta)$ constitutes a low-dimensional subspace of this larger space. While there is no true θ, we can define the pseudo-true θ as the value which makes $F_x(\theta)$ as close as possible to F_y. This requires specifying a distance metric between the joint distributions

$$d(\theta) = d(F_x(\theta), F_y)$$

and then we can define the best-fitting model $F_x(\theta)$ by selecting θ to minimize $d(\theta)$. The relevant question is then: what is the appropriate distance metric?

2. CATCH-ALL ESTIMATION

Xia and Tong recommend what they call a "catch-all" approach, where the distance metric is a weighted sum of squared k-step forecast residuals. They show that in some situations this criterion allows consistent estimation of the parameters of the true latent process. Their Theorem C requires that the latent process is deterministic, but the result might hold more broadly.

This can be illustrated in a very simple example of a latent AR(1) with additive measurement error. Suppose that the latent process is

$$x_t = \theta x_{t-1} + \varepsilon_t$$

and the observed process is

$$y_t = x_t + \eta_t,$$

where ε_t and η_t are independent white noise. In this case, it is well known that y_t has an ARMA(1, 1) representation

$$y_t = \theta y_{t-1} + u_t - \alpha u_{t-1}, \tag{1}$$

where u_t is white noise and $0 \le \alpha < 1$.

Xia and Tong propose estimation based on k-step forecast errors. The k-step forecast equation for the observables is

$$y_{t-1+k} = \theta^k y_{t-1} + e_t(k), \tag{2}$$

where

$$e_t(k) = \sum_{j=0}^{k-1} \theta^j (u_{t+k-j-1} - \alpha u_{t+k-j-2}).$$

Xia and Tong's estimator is based on a weighted average of squared forecast errors. For simplicity, suppose all the weight is on the kth forecast error. The estimator is

$$\hat{\theta}_{\{k\}} = \arg\min_{\theta} \sum_{t=1}^{T} (y_{t-1+k} - \theta^k y_{t-1})^2$$

which has the explicit solution

$$\hat{\theta}_{\{k\}} = \left(\frac{\sum_{t=1}^{T} y_{t-1} y_{t-1+k}}{\sum_{t=1}^{T} y_{t-1}^2} \right)^{1/k}.$$

We calculate that as $n \to \infty$

$$\hat{\theta}_{\{k\}} \xrightarrow{p} \theta_{\{k\}} = \theta(1-c)^{1/k},$$

where $c = \alpha\sigma_u^2/\theta\sigma_y^2$, $\sigma_u^2 = Eu_t^2$ and $\sigma_y^2 = Ey_t^2$.

Bruce E. Hansen is Professor, Department of Economics, University of Wisconsin, 1180 Observatory Drive, Madison, Wisconsin 53706, USA (e-mail: behansen@wisc.edu).