Comment

Jianqing FAN and Yingying FAN

We congratulate Koenker and Xiao on their interesting and important contribution to quantile autoregression (QAR). The article provides a comprehensive overview of the QAR model, from probabilistic aspects to model identification, statistical inferences, and empirical applications. The attempt to integrate the quantile regression and the QAR process is intriguing. It demonstrates that, surprisingly, nonparametric coefficient functions can be estimated at a root-n rate for the QAR processes. The authors then put forward some useful tools for testing the significance of lag variables and asymmetric dynamics of time series. We appreciate the opportunity to comment several aspects of this article.

1. CONNECTIONS WITH VARYING–COEFFICIENT MODELS

QAR is closely related to the functional-coefficient autoregressive (FCAR) model. In the time series context, Cai, Fan, and Li (2000b) proposed the following FCAR model for capturing the nonlinearity of a time series:

$$Y_t = \alpha_0(U_t) + \alpha_1(U_t)Y_{t-1} + \cdots + \alpha_p(U_t)Y_{t-p} + \varepsilon_t, \qquad (1)$$

where U_t is a thresholding variable and $\{\varepsilon_t\}$ is a sequence of independent innovations. In particular, when $U_t = Y_{t-d}$ for some lag d, the model was called a functional autoregressive model (FAR) by Chen and Tsay (1993). Varying-coefficient models have been widely used in many aspects of statistical modeling; see, for example, the work of Hastie and Tibshirani (1993), Carroll, Ruppert, and Welsh (1998), and Cai et al. (2000a) for applications to generalized linear models; Brumback and Rice (1998), Fan and Zhang (2000), and Chiang, Rice, and Wu (2001) for analysis of functional data; Lin and Ying (2001) and Fan and Li (2004) for analysis of longitudinal data; Tian, Zucker, and Wei (2005) and Fan, Lin, and Zhou (2006) for applications to the Cox hazards regression model; and Fan, Jiang, Zhang, and Zhou (2003), Hong and Lee (2003), and Mercurio and Spokoiny (2004) for applications to financial modeling. These are just a few examples that testify to the flexibility, popularity, and explanatory power of the varying-coefficient models. In the same vein, they reflect the importance of the QAR model.

What makes QAR different from the FCAR model or, more generally, the varying coefficient model is that the variable U_t is unobservable and $\varepsilon_t = 0$. This makes estimating techniques completely different. For example, in the varying-coefficient model, the coefficient functions in (1) are estimated through localizing on U_t (which are observable), whereas in the QAR model, the coefficient functions are estimated through quantile regression techniques. As a result, two completely different sets of rates of convergence are obtained. The former model admits a nonparametric rate, whereas the latter reveals the parametric rate.

Despite their differences in statistical inferences, QAR is a subfamily of models of FCAR as far as probabilistic aspects are concerned. Hence the stochastic properties established in FCAR are applicable directly to QAR. Chen and Tsay (1993) have given sufficient conditions for the solution to (1) to admit a stationary and ergodic solution. With some modifications of their proof, it can be shown that if $\alpha_j(\cdot)$ is bounded by c_j for all j and if all roots of the characteristic function

$$\lambda^p - c_1\lambda^{p-1} - \cdots - c_p = 0$$

are inside the unit circle, then there exists a stationary solution that is geometrically ergodic.

2. IDENTIFIABILITY OF THE MODEL

An important observation made by Koenker and Xiao is that if, given $Y_{t-p}, \ldots, Y_{t-1}$, the function

$$\beta_t(u) = \theta_0(u) + \theta_1(u)Y_{t-1} + \cdots + \theta_p(u)Y_{t-p} \qquad (2)$$

is strictly increasing in u, then $\beta_t(\tau)$ is the conditional τ-quantile of Y_t given $Y_{t-p}, \ldots, Y_{t-1}$. Because the conditional τ-quantile is identifiable under some mild conditions, the identifiability condition becomes that with probability 1, the QAR model generates at least $(p+1)$ linearly independent vectors of form $\mathbf{Y}_t = (1, Y_{t-1}, \ldots, Y_{t-p})^T$. In other words, letting

$$\mathcal{T} = \{t : \beta_t(u) \text{ is strictly increasing in } u\}, \qquad (3)$$

there are at least $(p+1)$ distinct time points $t_i \in \mathcal{T}$ such that $\mathbf{Y}_{t_i}$ are linearly independent for each realization. A natural and open question is what kind of population would generate, with probability 1, the samples that satisfy the foregoing condition.

The aforementioned identifiability conditions are hard to check. However, they are needed not only for connections to the quantile regression, but also for identifiability. To see this, look at the specific case where $p = 0$, in which $Y_t = \theta_0(U_t)$. Clearly, $\theta_0(\cdot)$ is the quantile function of Y_t only when $\theta_0(\cdot)$ is monotone increasing. When this condition is violated, the model is not necessarily identifiable. For example, $Y_t = |U_t - .5|$ and $Z_t = U_t - .5I(U_t > .5)$ have identically the same distribution but very different $\theta_0(\cdot)$.

We note that the QAR(p) model

$$Y_t = \theta_0(U_t) + \theta_1(U_t)Y_{t-1} + \cdots + \theta_p(U_t)Y_{t-p} \equiv \boldsymbol{\theta}(U_t)^T\mathbf{X}_t$$

is not differentiable from the model

$$Y_t = \boldsymbol{\theta}(1 - U_t)^T\mathbf{X}_t,$$

where $\mathbf{X}_t = (Y_{t-1}, \ldots, Y_{t-p})^T$. Thus if $\boldsymbol{\theta}(\tau)$ is a solution, then so is $\boldsymbol{\theta}(1-\tau)$.

Jianqing Fan is Professor (E-mail: *jqfan@princeton.edu*) and Yingying Fan is a Doctoral Candidate (E-mail: *yingying@princeton.edu*), Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08544. The authors thank the associate editor for helpful comments. The work was supported in part by National Science Foundation grant DMS-05-32370 and National Institutes of Health grant R01 GM072611.

Journal of the American Statistical Association
September 2006, Vol. 101, No. 475, Theory and Methods
DOI 10.1198/016214506000000681

3. FITTING AND DIAGNOSTICS

Koenker and Xiao estimate the coefficient functions $\boldsymbol{\theta}(\tau)$ with the quantile regression

$$\min_{\boldsymbol{\theta}} \sum_t \rho_\tau(Y_t - \mathbf{X}_t^T \boldsymbol{\theta}). \tag{4}$$

This convex optimization usually exists. The resulting estimates $\hat{\boldsymbol{\theta}}(\tau)$ are consistent estimates of the parameter

$$\boldsymbol{\theta}^*(\tau) = \arg\min_{\boldsymbol{\theta}} E\rho_\tau(Y_t - \mathbf{X}_t^T \boldsymbol{\theta}) \tag{5}$$

under some mild conditions. Without some technical conditions, $\boldsymbol{\theta}^*(\tau)$ and $\boldsymbol{\theta}(\tau)$ are not necessarily the same. This is evidenced by the example given in the previous section in which $\theta^*(\tau) = \tau/2$ with no ambiguity, whereas $\theta(\tau) = |\tau - .5|$ or $\tau - .5I(\tau > .5)$.

The foregoing argument suggests that the results of Koenker and Xiao should replace $\boldsymbol{\theta}(\tau)$ by $\boldsymbol{\theta}^*(\tau)$ unless the conditions under which they are identical are clearly imposed. If the primary interest is really on $\boldsymbol{\theta}(\tau)$, then the conditional quantile regression should be replaced by the restricted conditional quantile regression (RCQR),

$$\min_{\boldsymbol{\theta}} \sum_t I(t \in \mathcal{T}) \rho_\tau(Y_t - \mathbf{X}_t^T \boldsymbol{\theta}). \tag{6}$$

This avoids some samples in which the monotonicity condition is violated that create inconsistent estimators. However, the set $\mathcal{T}$ is unknown and depends on the value at another quantile τ. This creates some difficulties in implementation.

One possible way out is to replace $\mathcal{T}$ by one of its subsets. For example, if all $\theta_j(\cdot)$'s are monotonically increasing, then we can replace $\mathcal{T}$ by the subset in which all components of $\mathbf{X}_t$ are nonnegative. Another possibility is to use (4) to get an initial estimate and then check whether the functions $\{\hat{\beta}_t(\tau), t = 1, \ldots, T\}$ are strictly increasing at some percentiles (e.g., $\tau = .05, .1, .15, \ldots, .95$). Delete the cases in which the monotonicity is violated and use RCQR (6). The process can be iterated.

To illustrate the problem using the conditional quantile regression (4) and to address the issue of identifiability, we generate 2,000 data points from the QAR(1) model

$$Y_t = \Phi^{-1}(U_t) + (1.8U_t - 1.7)Y_{t-1}. \tag{7}$$

Hence we have $\theta_0(\tau) = \Phi^{-1}(\tau)$ and $\theta_1(\tau) = 1.8\tau - 1.7$. Fit the data using (4) and (6). The resulting estimates are depicted in Figure 1. The estimates (dot-dashed curve) obtained using RCQR (6) are very close to the true coefficient functions (thin solid curve), whereas the conditional quantile method (4) results in estimates (dashed curve) far away from the true functions. Indeed, the latter estimates are for the functions $\theta_0^*(\tau)$ and $\theta_1^*(\tau)$ defined by (5), which were computed numerically and depicted in Figure 1 by the thick solid curve. This example shows that even if monotonicity conditions are not fulfilled at all t, the coefficient functions can still be identifiable and consistently estimated, but the conditional quantile regression estimate, defined by maximizing (4), can be inconsistent.

A related question is how robust the fitting techniques are to model misspecification. For example, if the data-generating process is FCAR (1) without observing U_t, but we still use the conditional quantile regression (4) or its modification (6) to fit the data, how robust is the quantile estimate? To quantify this, we simulate the 2,000 data points from the model

$$Y_t = \Phi^{-1}(U_t) + (.85 + .25U_t)Y_{t-1} + \varepsilon_t, \tag{8}$$

where $\varepsilon_t \sim \mathrm{N}(0, \sigma^2)$. Figure 2 shows the plots for small noise, $\sigma = 0$; moderate noise, $\sigma = .8$, and relatively large noise, $\sigma = 2$. The fitting techniques are very sensitive to the noise level. The estimates differ substantially from the true coefficient functions for even moderate $\sigma = .8$.

Checking monotonicity of $\hat{\beta}_t(\tau)$ is one aspect of model diagnostics. Another aspect is to check whether the distribution of $\hat{U}_t = \hat{\beta}_t^{-1}(Y_t)$ for $t \in \mathcal{T}$ is uniform. There are many approaches to this kind of testing problem, including the Kolmogorov–Smirnov test and visual inspection of the estimated density. For example, one can create the normally transformed data $\hat{Z}_t = \Phi^{-1}(\hat{U}_t)$, then use the normal-reference rule of the kernel density estimate to see whether, the transformed residuals $\{\hat{Z}_t\}$ are normally distributed. Alternatively, one can use the quantile–quantile plot to accomplish a similar task.

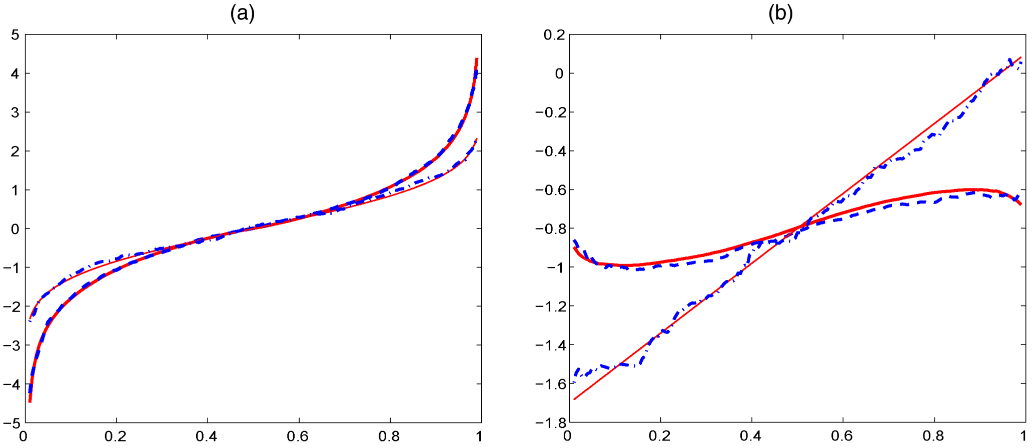

Figure 1. Estimates of $\theta_0(\tau) = \Phi^{-1}(\tau)$ (a) and $\theta_1(\tau) = 1.8\tau - 1.7$ (b) in Model (7). The thin curves are the true coefficient functions $\theta_0(\tau)$ and $\theta_1(\tau)$, the dashed curves are the estimates obtained by using conditional quantile regression (4), the dot-dashed curves are the estimates obtained by using restricted conditional quantile regression (6), and the thick solid curves are $\theta_0^*(\tau)$ and $\theta_1^*(\tau)$.

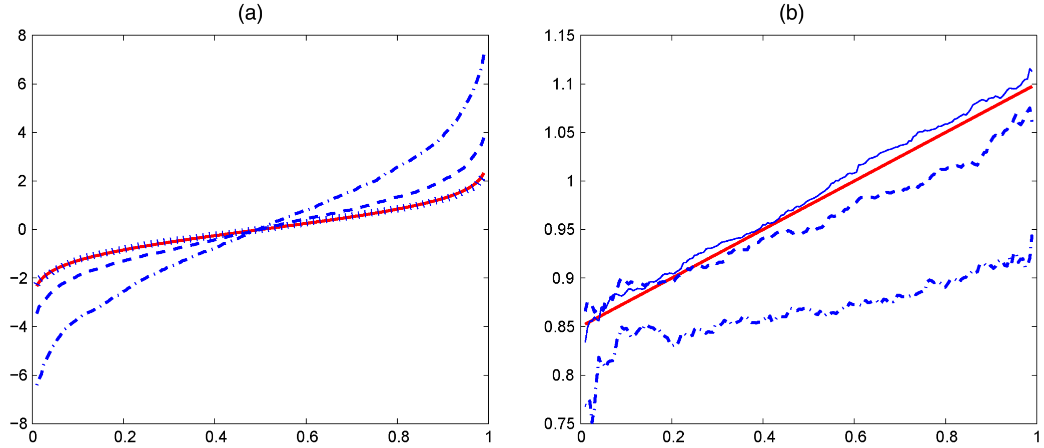

Figure 2. The Influences of the Error ε_t on the Estimation of $\theta_0(\tau) = \Phi^{-1}(\tau)$ (a) and the Estimation of $\theta_1(\tau) = .85 + .25\tau$ (b) in Model (8) for Different Noise Levels σ [(a) ——, true; ıııııı, $\sigma = 0$; - - -, $\sigma = .8$; -·-·-, $\sigma = 2$. (b) ——, true; ——, $\sigma = 0$; - - -, $\sigma = .8$; -·-·-, $\sigma = 2$].

For the data generated from (7) used in Figure 1, Figure 3 presents the histograms of $\hat{U}_t$ and the quantile–quantile plots of $\hat{Z}_t$. From these plots, we can see that the distribution of $\hat{U}_t = \hat{\beta}_t^{-1}(Y_t)$ for $t \in \mathcal{T}$ by using the conditional quantile regression method (4) is not uniform, whereas $\hat{U}_t$ obtained using the RCQR method (6) is uniformly distributed. These results are also supportive of our previous points.

Note that when the data are generated from model (1) without observing U_t, the model still can be identifiable when the density f_ε of innovation ε_t is known. The problem is more com-

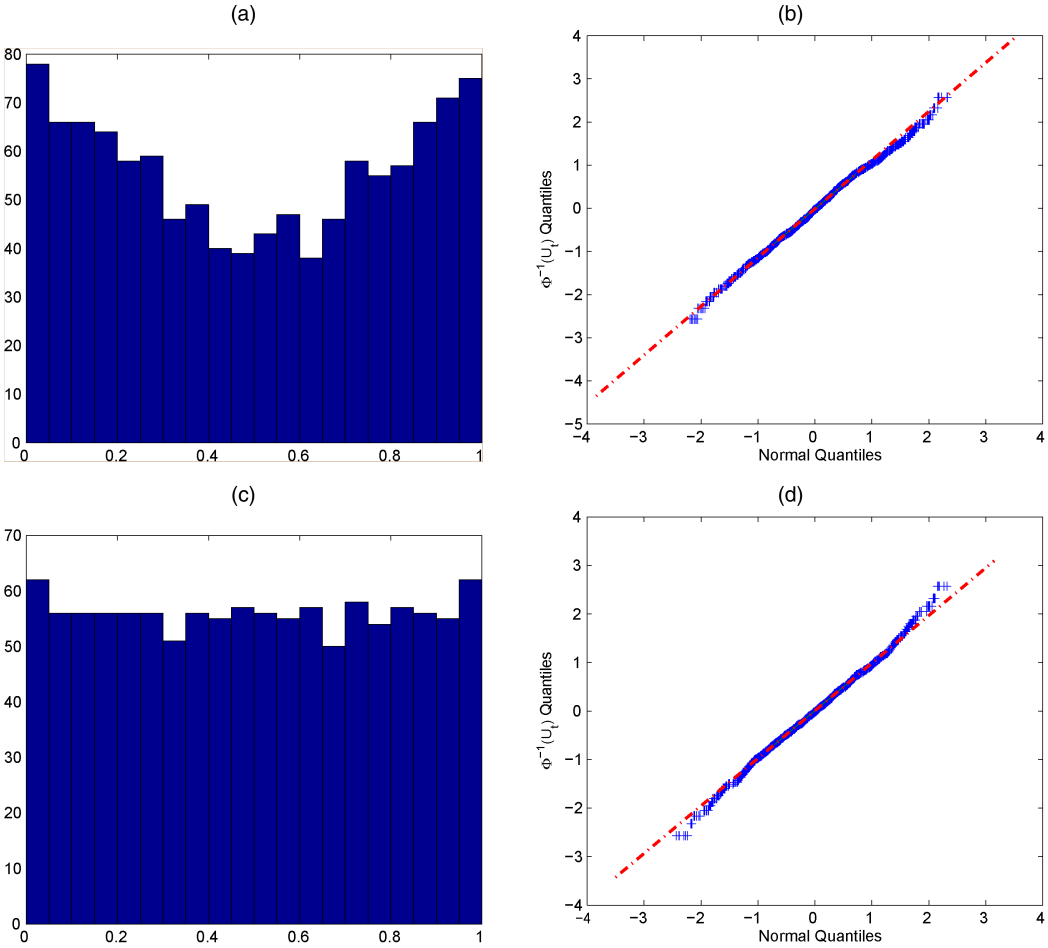

Figure 3. Histograms of $\hat{U}_t$ and Quantile–Quantile Plots of $\hat{Z}_t$. (a) Histogram plot of $\hat{U}_t = \hat{\beta}_t^{-1}(Y_t)$ for $t \in \mathcal{T}$, where $\hat{\beta}_t(\tau)$ is estimated using the conditional quantile regression (4). (b) Quantile–quantile plot of $\Phi^{-1}(\hat{U}_t)$ versus the standard normal distribution, where $\hat{U}_t$ is the same as in (a). (c) and (d) The same as (a) and (b), but with restricted conditional quantile regression (6) as the estimation method.

plicated but similar to a deconvolution problem. The estimation procedure can be quite complicated. To see this, note that

$$P(Y_t \le a | Y_{t-1}, \ldots, Y_{t-p}) = P(\beta_t(U_t) + \varepsilon_t \le a) \\ = \int \beta_t^{-1}(a - x) f_\varepsilon(x)\, dx.$$

Thus letting $Q_t(\tau)$ be the conditional τ-quantile, we have

$$\int \beta_t^{-1}(Q_t(\tau) - x) f_\varepsilon(x)\, dx = \tau.$$

Let us denote the solution by $Q_t(\tau) = h(\beta_t(\cdot), \tau)$ for some function h. Then the coefficient functions can be estimated by minimizing a quantity similar to (6),

$$\min_{\boldsymbol\theta} \int_0^1 \sum_t I(t \in \mathcal{T}) \rho_\tau\bigl(Y_t - h(\beta_t(\cdot), \tau)\bigr)\, d\tau,$$

where $\boldsymbol\theta(\cdot)$ and $\beta_t(\cdot)$ are related through (2). This is indeed a complicated optimization problem. The method of Koener and Xiao is a specific case of this method with $\varepsilon_t = 0$.

4. CONCLUDING REMARKS

Koenker and Xiao have developed a nice scheme for conditional quantile inference and made insightful connections with the QAR model. However, the issues of identifiability and possible misspecification of models suggest that extra care should be taken in making this kind of link. In particular, the conditional quantile method does not always produce a consistent estimate for the random-coefficient functions $\boldsymbol\theta(\cdot)$ when the monotonicity conditions are not satisfied. Further studies are needed.

ADDITIONAL REFERENCES

Brumback, B., and Rice, J. A. (1998), "Smoothing Spline Models for the Analysis of Nested and Crossed Samples of Curves" (with discussion), *Journal of the American Statistical Association*, 93, 961–994.

Cai, Z., Fan, J., and Li, R. (2000a), "Efficient Estimation and Inferences for Varying-Coefficient Models," *Journal of the American Statistical Association*, 95, 888–902.

——— (2000b), "Functional-Coefficient Regression Models for Nonlinear Time Series," *Journal of the American Statistical Association*, 95, 941–956.

Carroll, R. J., Ruppert, D., and Welsh, A. H. (1998), "Nonparametric Estimation via Local Estimating Equations," *Journal of the American Statistical Association*, 93, 214–227.

Chen, R., and Tsay, R. J. (1993), "Functional-Coefficient Autoregressive Models," *Journal of the American Statistical Association*, 88, 298–308.

Chiang, C.-T., Rice, J. A., and Wu, C. O. (2001), "Smoothing Spline Estimation for Varying Coefficient Models With Repeatedly Measured Dependent Variables," *Journal of the American Statistical Association*, 96, 605–619.

Fan, J., Jiang, J., Zhang, C., and Zhou, Z. (2003), "Time-Dependent Diffusion Models for Term Structure Dynamics and the Stock Price Volatility," *Statistica Sinica*, 13, 965–992.

Fan, J., and Li, R. (2004), "New Estimation and Model Selection Procedures for Semiparametric Modeling in Longitudinal Data Analysis," *Journal of the American Statistical Association*, 99, 710–723.

Fan, J., Lin, H., and Zhou, Y. (2006), "Local Partial-Likelihood Estimation for Life Time Data," *The Annals of Statistics*, 34, 290–325.

Fan, J., and Zhang, J. T. (2000), "Functional Linear Models for Longitudinal Data," *Journal of the Royal Statistical Society*, Ser. B, 62, 303–322.

Hastie, T. J., and Tibshirani, R. J. (1993), "Varying-Coefficient Models" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 55, 757–796.

Hong, Y., and Lee, T.-H. (2003), "Inference on Predictability of Foreign Exchange Rates via Generalized Spectrum and Nonlinear Time Series Models," *Review of Economics and Statistics*, 85, 1048–1062.

Lin, D. Y., and Ying, Z. (2001), "Semiparametric and Nonparametric Regression Analysis of Longitudinal Data" (with discussion), *Journal of the American Statistical Association*, 96, 103–126.

Mercurio, D., and Spokoiny, V. (2004), "Statistical Inference for Time-Inhomogeneous Volatility Models," *The Annals of Statistics*, 32, 577–602.

Tian, L., Zucker, D., and Wei, L. J. (2005), "On the Cox Model With Time-Varying Regression Coefficients," *Journal of the American Statistical Association*, 100, 172–183.

Comment

Keith Knight

First, I would like to congratulate the authors for a truly interesting and stimulating article. They have presented a very elegant methodology that should prove useful in many disciplines in which time series analysis is used. I find myself with really nothing to criticize; however, I comment on two issues: (a) the relationship between QAR(p) and AR(p) processes and (b) asymptotics for estimation in the infinite-variance case.

1. QAR VERSUS AR

QAR processes appear to be a very useful complement to AR processes, particularly in identifying local behavior in time series having different structure than the global behavior. However, viewed under the lens of classical time series analysis, the two processes are quite similar. In particular, it is interesting to note that the autocovariance function of a stationary QAR(p) process is simply that of a stationary (fixed) parameter AR(p) process; we have

$$y_t = \theta_0(U_t) + \theta_1(U_t) y_{t-1} + \cdots + \theta_p(U_t) y_{t-p} \\ = E[\theta_1(U_t)] y_{t-1} + \cdots + E[\theta_p(U_t)] y_{t-p} + V_t,$$

where

$$V_t = \theta_0(U_t) + \bigl\{\theta_1(U_t) - E[\theta_1(U_t)]\bigr\} y_{t-1} \\ + \cdots + \bigl\{\theta_p(U_t) - E[\theta_p(U_t)]\bigr\} y_{t-p}.$$

It is easy to verify that $\{V_t\}$ is a sequence of uncorrelated (but not independent) random variables. Thus classical model identification techniques based on autocorrelations, partial autocorrelations, and so on will tend to identify data generated from a QAR(p) process as a (fixed-parameter) AR(p) model. Although this is probably acceptable from certain viewpoints (e.g., prediction), it would also fail to identify structure in the

Keith Knight is Professor, Department of Statistics, University of Toronto, Toronto, ON M5S 3G3, Canada (E-mail: *keith@utstat.toronto.edu*).

Journal of the American Statistical Association
September 2006, Vol. 101, No. 475, Theory and Methods
DOI 10.1198/016214506000000690