Threshold Autoregressions for Strongly Autocorrelated Time Series

Markku LANNE
Department of Economics, University of Helsinki, Finland (*markku.lanne@helsinki.fi*)

Pentti SAIKKONEN
Department of Statistics, University of Helsinki, Finland (*pentti.saikkonen@helsinki.fi*)

In some cases the unit root or near unit root behavior of linear autoregressive models fitted to economic time series is not in accordance with the underlying economic theory. To accommodate this feature we consider a threshold autoregressive (TAR) process with the threshold effect only in the intercept term. Although these processes are stationary, their realizations switch between different regimes and can therefore closely resemble those of (near) integrated processes for sample sizes relevant in many economic applications. Estimation and inference of these TAR models are discussed, and a specification test for testing their stability is derived. Testing is based on the idea that if (near) integratedness is really caused by level shifts, the series purged of these shifts should be stable so that known stationarity tests can be applied to this series. Simulation results indicate that in certain cases these tests, like several linearity tests, can have low power. The proposed model is applied to interest rate data.

KEY WORDS: Near unit root; Stability test; Threshold models.

Many economic time series are strongly autocorrelated and can be modeled by linear (near) unit root or I(1) processes. However, implications of such linear models may not be in accordance with economic theory or "stylized facts." For instance, the estimated impulse response function may imply very slow mean-reversion inconsistent with the properties of many economic and financial variables. In the long run, (near) unit root behavior is also unthinkable for such variables as nominal interest rates and unemployment rate that do not take negative values. In such cases the occurrence of a (near) unit root may be an indication of factors not accounted for by the employed linear model. In this article we shall therefore consider nonlinear processes which are stationary but, due to stochastic level shifts, whose linear properties are very similar to those of I(1) or nearly I(1) processes. Specifically, we consider threshold autoregressive (TAR) models in which the threshold effect only appears in the intercept term.

Recently, González and Gonzalo (1998) and Caner and Hansen (2001) have also considered TAR models as alternatives to linear (nearly) I(1) models. Their models can have a (near) unit root in some of the regimes. This is contrary to our models, which have constant stable roots clearly smaller than 1 in all the regimes and in which only the intercept term switches between different regimes. González and Gonzalo (1998) and Caner and Hansen (2001) also derived statistical tests for testing their TAR models against linear (nearly) I(1) models. These linearity tests assume a stationary threshold variable which in practice is typically a lagged difference of the series. Because the motivation of our model assumes that the threshold variable is a lagged level of the series, these previous tests are not suitable for us.

Deriving a linearity test in the context of our TAR model appears very difficult. To help model selection, we therefore instead consider testing the null hypothesis that our TAR model is stable or stationary. A similar approach has previously been employed by Corradi, Swanson, and White (2000), who tested the stability of a general nonlinear process by applying the stationarity test of Kwiatkowski, Phillips, Schmidt, and Shin (1992) directly to the observed series. According to Monte Carlo simulations, this is not a reasonable approach in our context. It turns out to be much better to make use of the specified TAR model and employ a series purged of the estimated level shifts. If our stationary TAR model is appropriate, this series should be stable, and we use the stationarity test of Leybourne and McCabe (1994) for testing whether this is really the case. Visual inspection of this series can also be very helpful in selecting an appropriate model. Unfortunately, however, simulation results indicate that our stability test can have low power in certain cases, but this seems to be inherent in the considered testing problem because a similar result is obtained for several linearity tests presented in the previous literature.

Nominal interest rates are a potentially useful field of application of our TAR model. We illustrate this by using the three-month Swiss Franc Eurorate and UK Treasury bill rate data. Our results show that at least these interest rate series can be successfully modeled by clearly stable TAR models considered in this article. This is in contrast to previous linear interest rate models, which have typically contained a (near) unit root and therefore produced implications not in accordance with the observed behavior of interest rates (see, e.g., Ball and Torous 1996).

The plan of the article is as follows. In Section 1 our TAR model is presented, and in Section 2 parameter estimation and statistical inference are briefly discussed. Section 3 introduces the stability test. The empirical applications are presented in Section 4. Section 5 provides simulation results of the size and power of the stability test and some previously suggested linearity tests. Finally, Section 6 concludes.

Journal of Business & Economic Statistics
April 2002, Vol. 20, No. 2

1. MODELS

As an alternative to a linear (nearly) I(1) model, we consider the TAR model

$$y_t = \nu + \sum_{k=1}^{m} \delta_k I(y_{t-d} \geq c_k) + \sum_{j=1}^{p} \phi_j y_{t-j} + \sigma \varepsilon_t, \qquad (1)$$

where $I(\cdot)$ is the indicator function and ε_t is a sequence of continuous iid(0, 1) random variables with a density function positive everywhere. The threshold parameters are ordered as $c_1 < \cdots < c_m$ and the scale parameter σ satisfies $\sigma > 0$. Moreover, the zeros of the polynomial $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ are assumed to lie outside the unit circle. With these assumptions the process y_t is geometrically ergodic and strong mixing with geometrically decaying mixing numbers. This can be deduced from Masry and Tjøstheim (1995, lemma 3.1) by taking $\underline{d} = 0$, $(a'_1, \ldots, a'_{i_{g+1}-i_1}) = (\phi_p, \ldots, \phi_1)$, and using well-known results about eigenvalues of companion matrices. For the type of applications we have in mind, it is reasonable to assume that $\delta_k \geq 0$ $(k = 1, \ldots, m)$, where the inequality is strict unless otherwise stated. It is worth noting, however, that this assumption is only used to motivate the model. It is not necessary for the theoretical results to be discussed later.

To see the motivation of the above model, suppose that $y_{t-d} < c_1$. Then the process evolves in the lowest regime according to a linear AR(p) process whose level is determined by the parameter $\nu/\phi(1)$. However, when $y_{t-d} \geq c_1$ occurs, the value of the intercept term of the process changes from ν to a larger value $\nu + \delta_1$ and the process starts evolving around a higher level $(\nu + \delta_1)/\phi(1)$. Except for a new level, the properties of the process are the same as before the level shift. The larger is the value of the parameter δ_1, the larger is the difference between the two levels. After evolving in the second regime, the process may return to the first regime or, if $m \geq 2$, it can shift to an even higher level. In this way the process wanders between the m regimes and it is this wandering that, for sample sizes typical in many economic applications, can make realizations of the process look very similar to those of a linear I(1) or nearly I(1) process. In standard unit root and stationarity testing, such series can easily be confounded with I(1) series. This can happen even if the roots of the polynomial $\phi(z)$ are reasonably far away from unity and the number of regimes is only two or three (for examples see Section 4).

Because y_t can be treated as a stationary process and the zeros of the polynomial $\phi(z)$ lie outside the unit circle, Equation (1) implies the representation

$$y_t = \mu + \sum_{i=1}^{m} \frac{\delta_i}{\phi(L)} I(y_{t-d} \geq c_i) + z_t, \qquad (2)$$

where $\mu = \nu/\phi(1)$, $z_t = \phi(L)^{-1}\sigma\varepsilon_t$, and L stands for the usual lag operator. Clearly, z_t is a linear zero mean stationary AR(p) process. It represents the part of y_t from which both current and lagged effects of the stochastic level shifts have been removed. Thus, time series plots of y_t and z_t can be used as an informal practical aid to assess the significance of the estimated level shifts.

For some cases it is reasonable to extend the preceding model and allow for the possibility that the conditional variance may vary between regimes. For instance, in interest rate applications this may be needed to model the "stylized fact" that the variance is higher at higher levels. Therefore, we extend model (1) to

$$y_t = \nu + \sum_{k=1}^{m} \delta_k I(y_{t-d} \geq c_k) + \sum_{j=1}^{p} \phi_j y_{t-j} + \sigma(y_{t-d}) \varepsilon_t, \qquad (3)$$

where $\sigma(y_{t-d}) = \sigma + \sum_{k=1}^{m} \omega_k I(y_{t-d} \geq c_k)$ and the values of the parameters σ and $\omega_1, \ldots, \omega_m$ are such that $\sigma(y_{t-d})$ is positive everywhere. Thus, model (3) can also capture changes in the error variance dependent on the level of the series. That Equation (3) also defines a stationary process can again be deduced from Masry and Tjøstheim (1995, lemma 1). Hence, a representation, analogous to (2), applies with $z_t = \phi(L)^{-1}\sigma(y_{t-d})\varepsilon_t$, a zero mean stationary process.

2. PARAMETER ESTIMATION AND STATISTICAL INFERENCE

Given an observed time series $y_{r+1}, \ldots, y_T$, where $r = \max\{p, d\}$, the parameters in the TAR model (1) can be estimated by least squares (LS). Suppose the values of the integers p, m, and d are known. Then, for given values of the threshold parameters $c_1, \ldots, c_m$, the regression coefficients $\nu, \phi_1, \ldots, \phi_p$, and $\delta_1, \ldots, \delta_m$ in (1) can be estimated by ordinary LS. Because the sum of squares function in this LS estimation only takes a finite number of different values and the number of thresholds is usually one or two, it is possible and also quite feasible to use grid search over possible values of the threshold parameters to obtain the final LS estimators. Alternatively, one could consider the sequential procedure of Bai and Perron (1998) and estimate the thresholds one at a time. This would be especially convenient in the case of very long series or many thresholds. In practice the possible values of the threshold parameters are usually restricted in such a way that a prescribed minimum portion of observations are guaranteed to belong to each regime (cf., Caner and Hansen 2001). Since the LS estimators of the threshold parameters are not unique, we choose each of them as the smallest possible value such that the sum of squares function is minimized.

This estimation procedure is straightforward to extend to the case of the heteroskedastic TAR model (3). For given values of the threshold parameters, one can first use ordinary LS estimates of the regression coefficients $\nu, \phi_1, \ldots, \phi_p$, and $\delta_1, \ldots, \delta_m$ and associated residuals to estimate the unknown parameters σ and $\omega_1, \ldots, \omega_m$ in the function $\sigma(y_{t-d})$. This latter estimation is simply a reparameterized form of estimating the error variances in each regime. Using the estimate of $\sigma(y_{t-d})$, one can next obtain obvious weighted LS estimates of the regression coefficients $\nu, \phi_1, \ldots, \phi_p$, and $\delta_1, \ldots, \delta_m$ as well as the corresponding residual sum of squares function. This residual sum of squares function, which depends on the given values of the threshold parameters, can then be treated in the same way as its counterpart in model (1) so that the estimation procedure differs from the previous one only in that the ordinary LS estimation is replaced by a weighted LS estimation. This weighted LS estimation is based on a two-step approach and, if desired, it can be iterated until convergence to obtain proper weighted LS estimates. In this article proper weighted LS estimates are employed because the needed extra computations are quite modest.

Consistency and limiting distributions of LS estimators and maximum likelihood (ML) estimators in general TAR models with one threshold have recently been obtained by Chan (1993) and Qian (1998). The results presented in these articles imply that, under appropriate regularity conditions, the estimates of threshold parameters can be treated as if they were the true parameter values when hypotheses on other parameters are tested. Thus, except for the threshold parameters and with one exception to be discussed shortly, approximate standard errors of estimators and standard large sample tests with asymptotic chi-squared distributions under the null hypothesis can be constructed in a straightforward way. As for the LS estimators of the threshold parameters, complicated nonstandard limiting distributions result (see Chan 1993, Hansen 1997, and Qian 1998). Therefore, inference on the threshold parameters is difficult and will not be considered in this article.

The null hypothesis, which cannot be tested in a simple way, implies that the number of regimes is smaller than specified. In particular, the null hypothesis may imply a linear AR(p) model or that $\delta_1 = \cdots = \delta_m = 0$ holds in (1). Being able to test this null hypothesis and its extensions would be of major interest. Unfortunately, however, this testing problem is nonstandard and very difficult. An obvious difficulty is that under the null hypothesis the threshold parameters are not identified. In the present context further difficulties arise because the null hypothesis is supposed to imply an I(1) or nearly I(1) process. This means that the recent tests of González and Gonzalo (1998) and Caner and Hansen (2001) do not apply. Trying to extend the tests of these authors to the case of an I(1) or nearly I(1) threshold variable appears difficult and will not be attempted here. Some alternative approaches for testing the null hypothesis $\delta_1 = \cdots = \delta_m = 0$ in (1) will be discussed later.

Finally, note that, in addition to the value of the parameter m, also the values of the parameters p and d are unknown in practice so that experimentation with different values and some kind of model selection are needed. A consistent estimator of the delay parameter d can be found in Chan (1993) and Qian (1998).

3. STABILITY TEST

To assess whether our stationary TAR model can really account for the assumed high persistence of an observed series, a test procedure for testing it against a conventional linear I(1) model will be developed. The application of this test procedure together with available unit root and stationarity tests can give useful information about the properties of the considered time series and the relevance of our TAR model.

First consider the null hypothesis that the observed series is generated by a homoskedastic TAR model (1). As noticed in Section 1, then the process z_t defined in (2) has the stationary AR(p) representation $z_t = \phi(L)^{-1}\sigma\varepsilon_t$. Under the alternative, z_t is assumed to be an unstable I(1) process. Thus, if the process z_t were observable, the stationarity tests of Saikkonen and Luukkonen (1993) and Leybourne and McCabe (1994) could be used directly to test the desired hypotheses. Because this is not the case, we simply replace z_t by a sample analog based on the LS estimation of the null model and proceed as if z_t were observable. The same idea can be extended to the heteroskedastic TAR model (3) in a straightforward manner.

Our stationarity test is formulated in the framework of Leybourne and McCabe (1994), although the approach of Saikkonen and Luukkonen (1993) would lead to a similar test procedure. Thus, consider the unobserved component model

$$\phi(L)z_t = \sum_{j=1}^{t} \eta_j + \sigma\varepsilon_t, \qquad t = 1, 2, \ldots, \tag{4}$$

where the iid(0, σ_η^2) sequence η_t is totally independent of the sequence ε_t. The null hypothesis states that $\sigma_\eta^2 = 0$, whereas $\sigma_\eta^2 > 0$ under the alternative. The process defined by (4) is second order equivalent in moments to the ARIMA(p,1,1) process

$$\Delta z_t = \phi(L)^{-1}(1-\theta L)\sigma\varepsilon_t, \qquad t = 1, 2, \ldots, \tag{5}$$

where $\Delta = 1 - L$ is the difference operator. The moving average parameter θ is a function of the variances σ_η^2 and σ^2 and satisfies $0 < \theta \leq 1$ with $\theta = 1$ under the null hypothesis. Thus, z_t is supposed to be a stationary AR(p) process under the null hypothesis and an integrated ARIMA(p,1,1) process under the alternative.

Now consider the observed series y_t, which is highly persistent whether the null hypothesis holds or not. Under the null hypothesis this high persistence is modeled by the stochastic level shifts implied by the TAR process (1) or (3), whereas under the alternative a linear I(1) model is used for the same purpose. Thus, when we use the representation (2) with (4) to construct our stability test, we assume that under the null hypothesis the values of the parameters $\delta_1, \ldots, \delta_m$ are nonzero but under the alternative they are zero. Under the alternative the values of the parameters $\omega_1, \ldots, \omega_m$ in the TAR model (3) are similarly assumed to be zero.

In the case of the TAR model (1), our test statistic can also be based on an empirical counterpart of the process $u_t = z_t/\sigma$. The advantage of this formulation is that it extends to the case of the heteroskedastic TAR model (3). We use a circumflex to signify LS estimators and residuals based on the TAR model (1) or (3). Using the LS estimators $\hat{\phi}_1, \ldots, \hat{\phi}_p$ and the residuals $\hat{\varepsilon}_t$, we first compute recursively

$$\hat{u}_t = \sum_{j=1}^{p} \hat{\phi}_j \hat{u}_{t-j} + \hat{\varepsilon}_t, \quad t = 1, \ldots, T,$$

where $\hat{u}_t = 0$ for $t \leq 0$. An explicit expression of the residual $\hat{\varepsilon}_t$ is given by

$$\hat{\sigma}(y_{t-d})\hat{\varepsilon}_t = y_t - \hat{\nu} - \sum_{k=1}^{m} \hat{\delta}_k I(y_{t-d} \geq \hat{c}_k) - \sum_{j=1}^{p} \hat{\phi}_j y_{t-j}, \quad t = 1, \ldots, T, \tag{6}$$

where $\hat{\sigma}(y_{t-d}) = \hat{\sigma}$ in the case of model (1).

Our test procedure obviously assumes that $\hat{u}_t$ should behave like a stationary series under the null hypothesis and like an

I(1) series under the alternative. That this happens under the null hypothesis and for T large follows from the consistency of the LS estimators in (6). As for the alternative, recall that $\delta_k = \omega_k = 0$ $(k = 1, \ldots, m)$ by assumption so that z_t and, consequently, y_t is an unstable ARIMA process. For simplicity, consider the homoskedastic TAR model (1). Then, for $\hat{u}_t$ to behave like an I(1) series, we should have $\hat{\delta}_k = o_p(1)$ $(k = 1, \ldots, m)$ and $\hat{\phi}(1) = o_p(1)$. The fact that the threshold parameters are not identified complicates demonstrating this. However, the situation is simplified if we make the commonly used assumption and require that a fixed minimum portion of observations must belong to each regime. Then it is not difficult to use known limit theorems on stationary and integrated processes to show that $\hat{\delta}_k = o_p(1)$ $(k = 1, \ldots, m)$ and $\hat{\phi}(1) = o_p(1)$ hold [see Tsay (1998, theorem 1) for a similar result]. Thus, the series $\hat{u}_t$ behaves in the way the stationarity test of Leybourne and McCabe (1994) assumes.

Treating $\hat{u}_t$ $(t = 1, \ldots, T)$ as an observed series, we now proceed in exactly the same way as in Leybourne and McCabe (1994). The first step is to obtain suitable estimators for the parameters $\phi_1, \ldots, \phi_p$ and to form an empirical counterpart for the series $\phi(L)\hat{u}_t$. In the same way as in Saikkonen and Luukkonen (1993) and Leybourne and McCabe (1994), the use of the LS estimators $\hat{\phi}_1, \ldots, \hat{\phi}_p$ renders the resulting test inconsistent, and the same thing happens if one uses the LS estimators from a regression of $\hat{u}_t$ on $\hat{u}_{t-1}, \ldots, \hat{u}_{t-p}$. Thus, following these previous authors, we consider (quasi) ML estimators based on the auxiliary ARMA(p,1) model

$$\phi(L)\Delta\hat{u}_t = (1 - \theta L)a_t, \qquad (7)$$

where a_t is treated as Gaussian white noise and θ is as in (5). Notice that if $\hat{u}_t$ on the left side is replaced by u_t, we have $a_t = \varepsilon_t$. Let $\phi_1^*, \ldots, \phi_p^*$ be the ML estimators of $\phi_1, \ldots, \phi_p$ based on (7). Using these estimators we next form the series

$$\hat{e}_t^* = \hat{u}_t - \sum_{j=1}^{p} \phi_j^* \hat{u}_{t-j},$$

and its demeaned version

$$\hat{\varepsilon}_t^* = \hat{e}_t^* - (T - p)^{-1} \sum_{t=p+1}^{T} \hat{e}_t^*, \qquad t = p+1, \ldots, T.$$

As in Leybourne and McCabe (1994), we can finally introduce the test statistic

$$S = \hat{\varepsilon}^{*\prime} V \hat{\varepsilon}^* / (T - p)\hat{\varepsilon}^{*\prime}\hat{\varepsilon}^*,$$

where $\hat{\varepsilon}^* = [\hat{\varepsilon}_{p+1}^* \cdots \hat{\varepsilon}_T^*]'$ and V is a $(T-p) \times (T-p)$ matrix with ijth element equal to the minimum of i and j. It can be shown that under the null hypothesis and appropriate regularity conditions, test statistic S has the same limiting distribution as obtained in Leybourne and McCabe (1994). A formal proof of this result can be found in a discussion paper version of this article and is available upon request. Thus, under the null hypothesis,

$$S \xrightarrow{d} \int_0^1 [B(r) - rB(1)]^2 \, dr, \qquad (8)$$

where $B(r)$ is a standard Brownian motion and consequently $B(r) - rB(1)$ is a standard Brownian bridge. Large values of test statistic S are critical. Critical values can be found in Kwiatkowski et al. (1992).

Recently, Caner and Kilian (2001) and Lanne and Saikkonen (in press) have pointed out that, in finite samples, the stationarity test of Leybourne and McCabe (1994) suffers from a serious over-rejection problem when applied to stationary but strongly autocorrelated series. The latter authors developed a modification of the original test which works considerably better and will be used in the applications of this article. In our context, the reason for the overrejection problem is that estimation errors in the ML estimates $\phi_1^*, \ldots, \phi_p^*$ are totally ignored when the series $\hat{e}_t^*$ is constructed. Therefore, following the idea of Lanne and Saikkonen (in press), we replace the demeaned version of $\hat{e}_t^*$ by residuals from the auxiliary regression model

$$\hat{e}_t^* = \kappa - \beta(L)\varphi(L)\Delta\hat{u}_{t-1} + w_t + \beta(L)\varrho_{t-1}, \qquad t = 2p+1, \ldots, T, \qquad (9)$$

where κ is an intercept term, $\beta(L) = \sum_{j=1}^{p} \beta_j L^{j-1} = \sum_{j=1}^{p} \phi(1)^{-1}(\phi_j^* - \phi_j)L^{j-1}$, $\varphi(L) = \sum_{j=0}^{p-1} \varphi_j L^j$ with $\varphi_j = \sum_{i=j+1}^{p} \phi_i$, and $\varrho_t = \sum_{j=1}^{t} \eta_j + \varepsilon_t$. When the null hypothesis holds, $\varrho_t = \varepsilon_t$ and (9) can be treated as a regression model with moving average errors and multiplicative constraints between the regression coefficients and moving average parameters. Thus, we treat ϱ_t in (9) as Gaussian white noise, estimate the parameters by ML, and use the resulting residuals $\tilde{\varepsilon}_t^*$ $(t = 2p+1, \ldots, T)$, say, instead of $\hat{\varepsilon}_t^*$. In the same way as in Lanne and Saikkonen (in press), one can also show here that this modification does not change the asymptotic distribution of the test.

4. EMPIRICAL RESULTS

4.1 Swiss Franc Eurorate

As a first empirical illustration, we consider the monthly Swiss Franc three-month Eurorate during the period 1978:1–1997:9 (237 observations) extracted from the International Financial Statistics published by the IMF. This series serves as an example of a case where a homoskedastic TAR model (1) is sufficient. The solid line in Figure 1 depicts the three-month Swiss Franc Eurorate. The series looks very persistent and, in fact, the Dickey–Fuller test does not reject the null of a unit root at the 5% level (the value of the test statistic is −1.92, and the 5% critical value is −2.86), while the modified Leybourne–McCabe test does reject the null of stationarity at the 10% level (the value of the test statistic is .42, and the 10% and 5% critical values are .35 and .46, respectively). The estimation results of the linear AR(2) model and the level-shift TAR(2) models with one and two thresholds are presented in Table 1. The linear model corresponds to our expectations based on the stationarity test: the sum of the AR coefficients is high, indicating (near) unit root behavior of the series. In the one-threshold model, the threshold is estimated at 7.1%. Taking this level shift into account results in a "z" series for which the null of stationarity is rejected at the 1% level. Also, the visual inspection of this series in Figure 1 reveals some potentially remaining level shifting. According to the test for

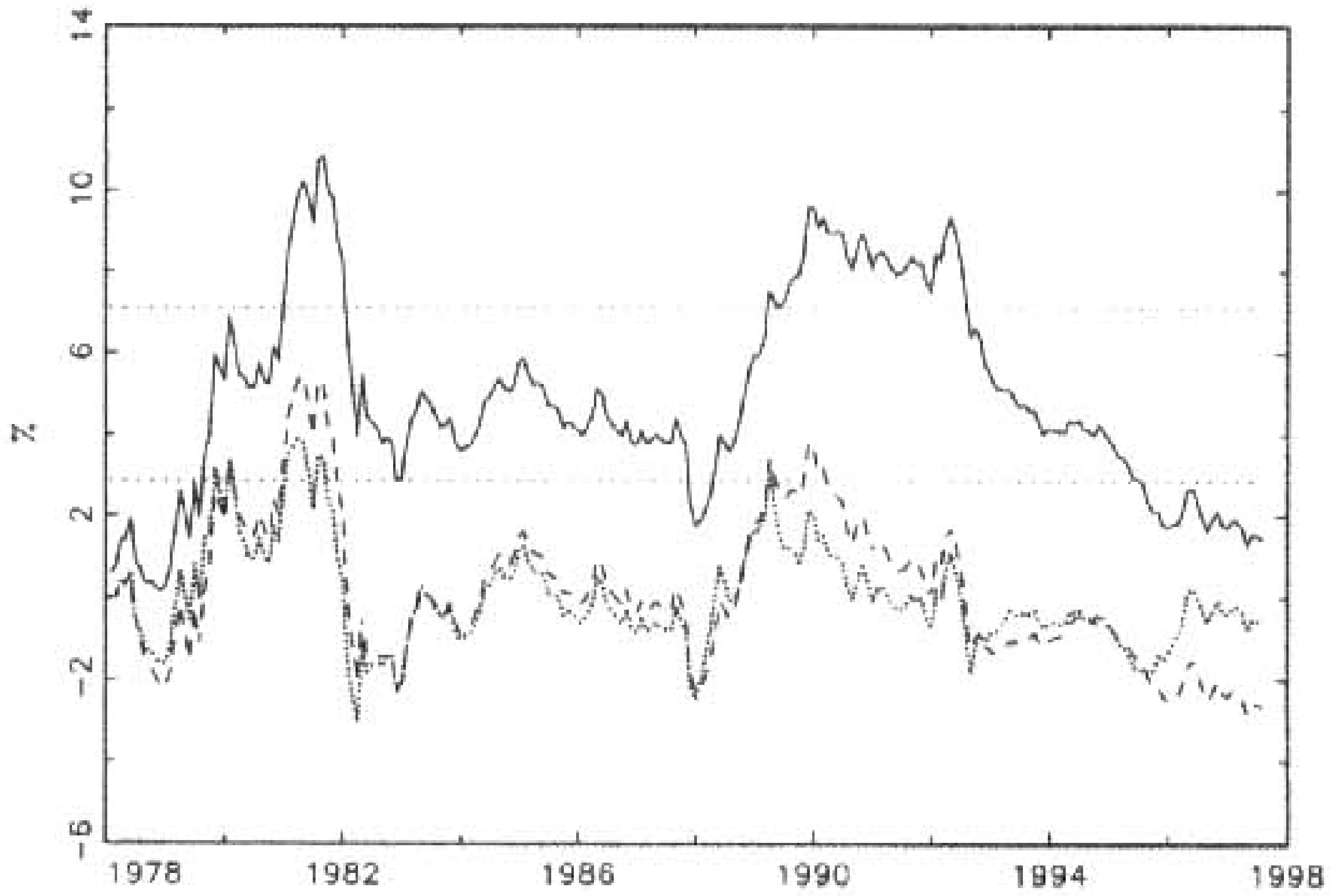

Figure 1. The Level (solid line) and "z" Series Computed from the One-Threshold (dashes) and Two-Threshold (dots) Models of the Three-Month Swiss Franc Eurorate. The dotted straight lines indicate the estimated thresholds.

a general TAR model, the level-shift model is not sufficient. The rejection could, however, also indicate some other kind of misspecification, such as the need for an additional threshold.

In the two-threshold model, the upper threshold is estimated at 7.1% and the lower threshold at 2.8%. The corresponding levels around which the process evolves in the three regimes are about 8.3, 4.6, and 2.0%, respectively. The stationarity of the "z" series cannot be rejected at the 10% level, and it has a narrower range than the "z" series for the one-threshold model. Compared to the linear AR and one-threshold level-shift TAR models, the sum of the autoregressive coefficients and the residual sum of squares are considerably lower and, according to diagnostic tests, there is no reason to suspect the adequacy of the model. As a further informal check, we simulated long realizations from the estimated linear AR and TAR models. It turned out that the AR model tended to produce excessively erratic realizations with a large number of negative values, whereas the range of the values in the realizations from the two-threshold TAR model more closely corresponded to that of the observed series.

4.2 UK Treasury Bill Rate

The second example is the monthly series of the UK Treasury Bill rate during the period 1964:1–1997:9 (405 observations) also extracted from the International Financial Statistics of the IMF. For this series the heteroskedastic specification (3) is required. The solid line in Figure 2 depicts the interest rate series, and the estimation results are presented in Table 2. The estimated linear model is characterized by high persistence, and the Dickey–Fuller test cannot reject the null of a unit root at the 5% level (the value of the test statistic is −2.78, and the critical value is −2.86 at the 5% level), whereas the modified Leybourne–McCabe test rejects the null of stationarity

Table 1. Estimation Results: Swiss Franc Eurorate

	Model		
	AR	TAR (one threshold)	TAR (two thresholds)
ν	.140	.241	.284
	(.074)	(.098)	(.098)
δ_1		.221	.354
		(.139)	(.144)
δ_2			.510
			(.181)
c_1		7.103	2.845
c_2			7.103
ϕ_1	1.141	1.113	1.045
	(.064)	(.067)	(.071)
ϕ_2	−.169	−.172	−.183
	(.064)	(.064)	(.064)
Sum of AR coefficients	.972	.941	.861
Residual sum of squares	60.822	60.167	58.620
Stationarity test of the "z" series[a]	.420	.883	.192
AR(1)[b]	.164	.285	.534
AR(2)	.139	.427	.136
AR(3)	.220	.192	.114
AR(4)	.354	.253	.170
AR(5)	.457	.389	.275
AR(6)	.487	.435	.327
LM test for homoskedasticity[c]		.636	.680
LM test for general TAR[d]		.036	.195
Number of observations by regime:			
Upper		54	54
Middle			133
Lower		181	48

NOTE: The TAR models are estimated under the restriction that at least 15% of the observations belong to each regime. The delay parameter $d = 1$. The figures in the parentheses are standard errors. For all tests except the stationarity test, marginal significance levels are reported.

[a] The 1%, 5%, and 10% critical values are .739, .463, and .347, respectively.

[b] AR(k) is the LM test for autocorrelation of orders 1 to k.

[c] LM test for equality of error variance across the regimes.

[d] LM test for the same autoregressive polynomial in all the regimes.

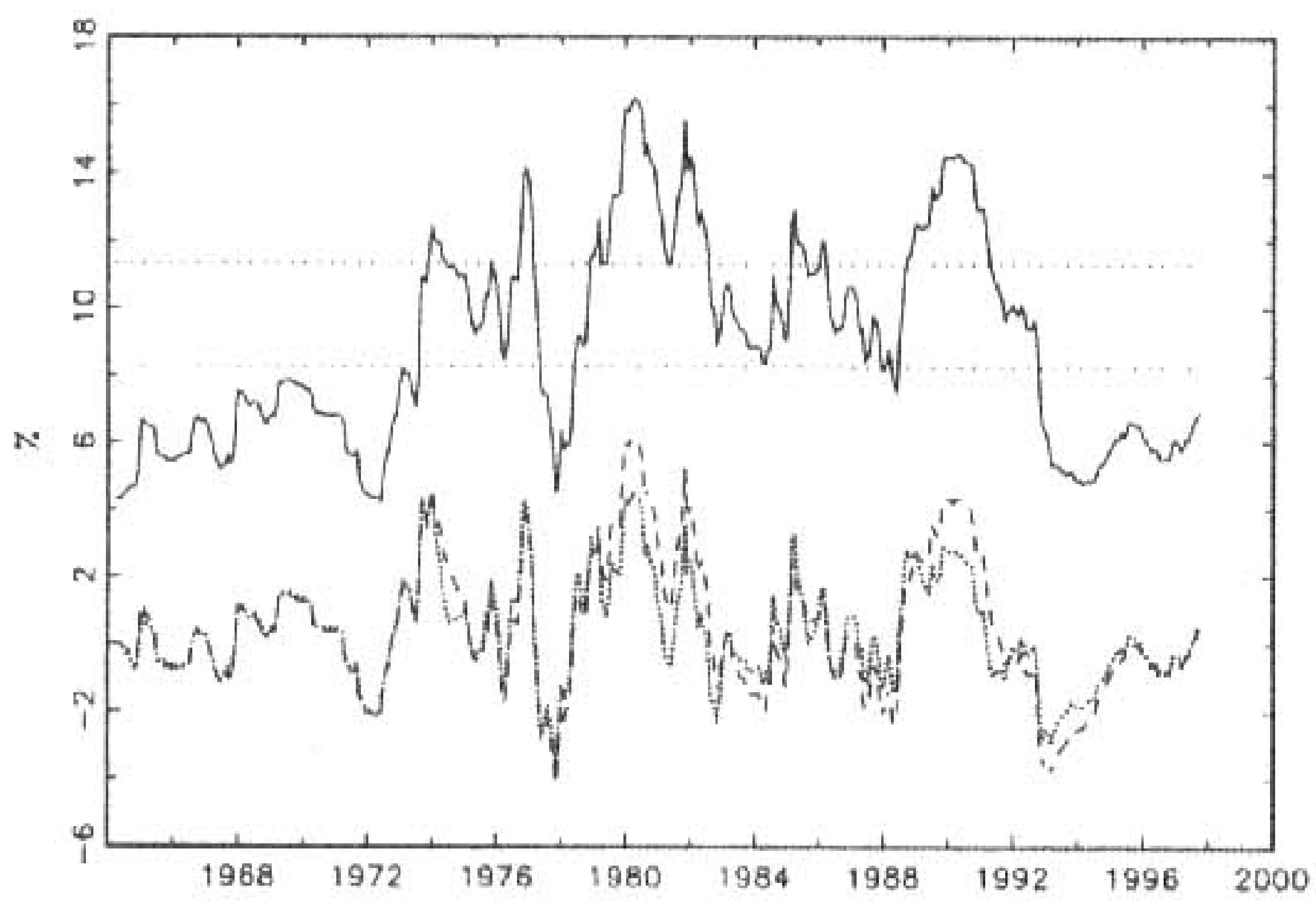

Figure 2. The Level (solid line) and "z" Series Computed from the One-Threshold (dashes) and Two-Threshold (dots) Models of the UK Treasury Bill Rate. The dotted straight lines indicate the estimated thresholds.

(the value of the test statistic is 3.43, and the 5% critical value is .46).

As already indicated, the homoskedasticity of the error term was clearly rejected when one- and two-threshold TAR models (1) were tried, so results of the heteroskedastic model (3) are only reported. In the one-threshold model, the threshold is estimated at 8.3%. The sum of the autoregressive coefficients is not much lower for this model than for the linear model, and the stationarity of the "z" series is rejected at the 5% level. It is also noteworthy that the estimated residual variance in the upper regime is almost twice the estimate in the lower regime. The second threshold is estimated at 11.3%. The levels corresponding to these thresholds, around which the process evolves in the upper, middle, and lower regimes, are about 12.0, 9.4, and 6.3%, respectively. The "z" series in Figure 2 for the two-threshold model shows no clear signs of level shifts, and also the stability test fails to reject at the 10% level. The decrease in the sum of the AR coefficients compared to the one-threshold model is considerable, albeit not as dramatic as in the model for the Swiss Franc Eurorate. The estimated residual variances in the three regimes also seem to be clearly different, lending further support to the three-regime against two-regime specification. Simulations of long realizations provided informal evidence in favor of the two-threshold TAR model against the linear AR model also in this case.

5. SIMULATION RESULTS

This section presents limited simulation results on the performance of our stability test and some alternative tests in distinguishing between a stationary TAR model and a linear (near) unit root process. The estimated two-threshold TAR

Table 2. Estimation Results: UK Treasury Bill Rate

	Model		
	AR	*TAR (one threshold)*	*TAR (two thresholds)*
μ	.217	.363	.538
	(.080)	(.101)	(.129)
δ_1		.220	.262
		(.095)	(.096)
δ_2			.217
			(.110)
c_1		8.270	8.270
c_2			11.320
ϕ_1	1.341	1.335	1.307
	(.047)	(.047)	(.049)
ϕ_2	−.365	−.391	−.391
	(.046)	(.045)	(.045)
Sum of AR coefficients	.976	.944	.915
Residual sum of squares	105.220	103.825	103.272
Stationarity test of the "z" series[+]	3.434	.490	.134
AR(1)[a]	.388	.428	.352
AR(2)	.694	.523	.601
AR(3)	.795	.723	.798
AR(4)	.573	.814	.885
AR(5)	.710	.108	.229
AR(6)	.787	.145	.312
LM test for general TAR[*]		.831	.063
Number of observations by regime:			
Upper		215	96
Middle			119
Lower		188	188
Residual variance by regime:			
Upper		.615	.655
Middle			.576
Lower		.344	.342

NOTE: The TAR models are estimated by iterated weighted least squares under the restriction that at least 15% of the observations belong to each regime. The delay parameter $d = 1$. The figures in the parentheses are standard errors. For all tests except the stationarity test, marginal significance levels are reported.

*See notes to Table 1.

models of Section 4 were used as the data generating processes (DGP's). This choice was motivated by our observation that generating representative realizations of TAR models is not a straightforward task because complicated relations prevail between the parameters, making it almost impossible to concentrate on examining the effect of changing only the value of one parameter at a time. For instance, changing the values of AR parameters also changes the variance of the process, and unless the threshold parameters are changed accordingly, most of the simulated realizations need not have any observations in some of the regimes. However, finding the required adjustments is laborious at the very least and makes reporting the results inconvenient. Of course, this difficulty depends on the length of the series and disappears when sufficiently long series are simulated. All the following results are based on 1,000 replications. Throughout, the lag length is assumed known, and a two-threshold TAR model is fitted to the generated realizations. While this conforms with the DGP's in the size simulations, we have made this choice in the power simulations as well because our experience with these models suggests that two thresholds are, in general, required to capture strong autocorrelation.

The empirical sizes of the stability test are presented in Table 3 along with the rejection rates of the modified Leybourne–McCabe test and augmented Dickey–Fuller (ADF) test based on a model with an intercept but no time trend. The rejection rates of the stability test indicate that the test controls size relatively well already with 100 observations. As expected, the size is closer to the nominal size for the Swiss Franc Eurorate model, which has a lower sum of the autoregressive coefficients than the UK Treasury bill rate model. Somewhat surprisingly, the sample size seems to have a negligible effect on the empirical size. The results of the modified Leybourne–McCabe test and the ADF test imply that both of these tests can easily fail to reveal the stationarity of the considered TAR processes. In particular, using the modified Leybourne–McCabe test for the original series directly results in severe overrejection. It may be noted that the same result was obtained when the test of Kwiatkowski et al. (1992) was tried (the rejection rates varied between .30 and .48). This implies that, in general, the recent suggestion of Corradi et al. (2000) to use this test in the case of a nonlinear stationary DGP should be applied with caution.

For power simulations we used the same DGP as Leybourne and McCabe (1994):

$$y_t = \alpha_t + \varepsilon_t, \qquad \alpha_t = \alpha_{t-1} + \eta_t,$$

where $\varepsilon_t \sim \text{nid}(0, 1)$ and $\eta_t \sim \text{nid}(0, \sigma_\eta^2)$ are independent and the process α_t is initialized at zero. Note that this linear I(1) process can equivalently be expressed as

$$y_t = y_{t-1} + \zeta_t - \theta\zeta_{t-1},$$

where $\zeta_t \sim \text{nid}(0, \sigma_\zeta^2)$ and $0 < \theta \leq 1$. In terms of this representation, the closer θ is to unity the closer y_t is to a stationary process. The rejection rates are presented in Table 4. For $T = 300$, the power is good for some values of σ_η^2. Initially it increases with σ_η^2 but decreases as σ_η^2 increases from .1 to 8.0. The relatively low power at very small values of σ_η^2 follows from the fact that in this case the random walk component of y_t is only weak. On the other hand, as σ_η^2 becomes large enough, the realizations start resembling those from a TAR model because with increased variance of the random walk component large shifts from one level to another become more common. As a result, fitting a TAR model to the realizations becomes increasingly successful, which tends to decrease the power of the stability test. This can also be seen from the values of the moving average parameter θ. Initially the power improves with diminishing values of θ but begins to decline as θ approaches zero and the process approaches a random walk. This is a disappointing result because it means that the power of the test is low against alternatives that are of primary interest for us.

Because our stability test seemed to lack power in cases of interest, some simulation experiments were conducted to examine whether some known linearity tests would be more successful. The estimated two-threshold TAR model for the Swiss Franc Eurorate was used as the DGP, and in each experiment the nominal size was 5%. Although the following tests have been derived assuming a strictly stationary threshold variable, limited simulation experiments suggested that they control size well. Tsay's (1989) test based on arranging the observations according to the threshold variable and using predictive residuals from the arranged autoregression can have

Table 3. Rejection Rates of the Modified Leybourne–McCabe (L–M) and Augmented Dickey–Fuller (ADF) and Stability Tests With Nominal Size 5%

	T	
	100	300
	Swiss Franc Eurorate model	
Modified L–M test	.414	.322
ADF	.238	.667
Stability test	.068	.072
	U.K. Treasury bill rate model	
Modified L–M test	.361	.350
ADF	.371	.641
Stability test	.080	.083

NOTE: The figures are based on 1,000 replications of the two-threshold TAR models of the previous section. The ADF test is based on a regression model including an intercept but no time trend.

Table 4. Power of the Stability Test With Nominal Size 5%

		T	
σ_η^2	θ	100	300
.001	.969	.128	.539
.010	.905	.486	.897
.100	.730	.698	.960
1.000	.382	.332	.823
2.000	.268	.236	.697
4.000	.172	.172	.444
8.000	.101	.142	.265

NOTE: The DGP is the following: $y_t = \alpha_{t-1} + \varepsilon_t$, $\alpha_t = \alpha_{t-1} + \eta_t$ with $\varepsilon_t \sim \text{nid}(0, 1)$ and $\eta_t \sim \text{nid}(0, \sigma_\eta^2)$. Equivalently, the DGP can be expressed as $y_t = y_{t-1} + \zeta_t - \theta\zeta_{t-1}$ with $\zeta_t \sim \text{nid}(0, \sigma_\zeta^2)$. The number of replications is 1,000.

very low power: it was just 11% with even as many as 500 observations. Caner and Hansen (2001) suggested bootstrapping the critical values of the Wald test for no thresholds. The empirical power turned out to be only about 45% for 300 observations. Because the TAR model is a limiting case of the smooth transition autoregression, we also considered the test of Luukkonen, Saikkonen, and Teräsvirta (1988) against this type of nonlinearity (using critical values from the asymptotic distribution derived under the assumption of an integrated transition variable). Also this test lacked power: it was about 22% for 300 observations. Thus, all these linearity tests have difficulties with distinguishing the considered TAR models from their linear approximations.

6. CONCLUSION

In this article we have argued that fairly simple stationary nonlinear TAR models may be useful alternatives for linear (nearly) I(1) models in cases where the long-run properties of the latter are not in accordance with underlying economic theory or observed properties of the series. We also developed a stability test that, together with available stationarity and unit root tests, can be used to assess the applicability of our TAR models over conventional I(1) alternatives. Two empirical examples on nominal interest rates demonstrated the potential usefulness of these ideas.

From the empirical examples and limited simulation results, the following conclusions emerge. First, clearly stationary TAR processes considered in this article can easily be deemed as nonstationary unit root processes in standard unit root and stationarity testing. Second, for sample sizes of a few hundred observations or less, our stability test, as well as some recent linearity tests, can have low power. This implies that, in addition to statistical tests, also other procedures are worth using in model selection. In this respect, visual inspection of the graph of the "z" series on which our stability test is based, is a convenient possibility. Long realizations of the chosen linear model and our nonlinear TAR model can also be simulated to visually inspect their resemblance to the observed series. At least in our empirical examples, this procedure seemed to provide corroborating evidence in favor of the TAR specifications.

ACKNOWLEDGMENTS

We thank Dag Tjøstheim for his helpful advice and two anonymous referees for useful comments. Financial support from the Yrjö Jahnsson Foundation is gratefully acknowledged.

[Received May 1999. Revised May 2001.]

REFERENCES

Bai, J., and Perron, P. (1998), "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 66, 47–78.

Ball, C. A., and Torous, W. N. (1996), "Unit Roots and the Estimation of Interest Rate Dynamics," *Journal of Empirical Finance*, 3, 215–238.

Caner, M., and Hansen, B. E. (2001), "Threshold Autoregression with a Unit Root," *Econometrica*, 69, 1555–1596.

Caner, M., and Kilian, L. (2001), "Size Distortions of Tests of the Null Hypothesis of Stationarity: Evidence and Implications for the PPP Debate," *Journal of International Money and Finance*, 20, 639–657.

Chan, K. S. (1993), "Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model," *Annals of Statistics*, 21, 520–533.

Corradi, V., Swanson, N. R., and White, H. (2000), "Testing for Stationary-Ergodicity and for Comovements Between Nonlinear Discrete Time Markov Processes," *Journal of Econometrics*, 96, 39–73.

González, M., and Gonzalo, J. (1998), "Threshold Unit Root Models," unpublished manuscript, U. Carlos III de Madrid.

Hansen, B. E. (1997), "Inference in TAR Models," *Studies in Nonlinear Dynamics and Econometrics*, 1, 119–131.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., and Shin, Y. (1992), "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure Are We That Economic Time Series Have a Unit Root?," *Journal of Econometrics*, 154, 159–178.

Lanne, M., and Saikkonen, P. (in press), "Reducing Size Distortions of Parametric Stationarity Tests," *Journal of Time Series Analysis*.

Leybourne, S. J., and McCabe, B. P. M. (1994), "A Consistent Test for a Unit Root," *Journal of Business & Economic Statistics*, 12, 157–166.

Luukkonen, R., Saikkonen, P., and Teräsvirta, T. (1988), "Testing Linearity Against Smooth Transition Autoregressive Models," *Biometrika*, 75, 491–499.

Masry, E., and Tjøstheim, D. (1995), "Nonparametric Estimation and Identification of Nonlinear ARCH Time Series," *Econometric Theory*, 11, 258–289.

Qian, L. (1998), "On Maximum Likelihood Estimators for a Threshold Autoregression," *Journal of Statistical Planning and Inference*, 75, 21–46.

Saikkonen, P., and Luukkonen, R. (1993), "Testing for a Moving Average Unit Root in Autoregressive Integrated Moving Average Models," *Journal of the American Statistical Association*, 88, 596–601.

Tsay, R. S. (1989), "Testing and Modeling Threshold Autoregressive Processes," *Journal of the American Statistical Association*, 84, 231–240.

——— (1998), "Testing and Modeling Multivariate Threshold Models," *Journal of the American Statistical Association*, 93, 1188–1202.