Outlier Detection in Multivariate Time Series by Projection Pursuit

Pedro GALEANO, Daniel PEÑA, and Ruey S. TSAY

In this article we use projection pursuit methods to develop a procedure for detecting outliers in a multivariate time series. We show that testing for outliers in some projection directions can be more powerful than testing the multivariate series directly. The optimal directions for detecting outliers are found by numerical optimization of the kurtosis coefficient of the projected series. We propose an iterative procedure to detect and handle multiple outliers based on a univariate search in these optimal directions. In contrast with the existing methods, the proposed procedure can identify outliers without prespecifying a vector ARMA model for the data. The good performance of the proposed method is illustrated in a Monte Carlo study and in a real data analysis.

KEY WORDS: Kurtosis coefficient; Level shift; Multivariate time series; Outlier; Projection pursuit.

1. INTRODUCTION

Outlier detection in time series analysis is an important problem because the presence of even a few anomalous data can lead to model misspecification, biased parameter estimation, and poor forecasts. Several detection methods have been proposed for univariate time series, including those of Fox (1972), Chang and Tiao (1983), Tsay (1986, 1988), Chang, Tiao, and Chen (1988), Chen and Liu (1993), McCulloch and Tsay (1993, 1994), Le, Martin, and Raftery (1996), Luceño (1998), Justel, Peña, and Tsay (2000), Bianco, Garcia Ben, Martínez, and Yohai (2001), and Sánchez and Peña (2003). Most of these methods are based on sequential detection procedures. For multivariate time series, Franses and Lucas (1998) studied outlier detection in cointegration analysis; Tsay, Peña, and Pankratz (2000) proposed a detection method based on individual and joint likelihood ratio statistics; and Lütkephol, Saikkonen, and Trenkler (2004) analyzed the effect of level shifts on the cointegration rank.

Building adequate models for a vector time series is a difficult task, especially when the data are contaminated by outliers. In this article we propose a method for identifying outliers without requiring initial specification of the multivariate model. The method is based on univariate outlier detection applied to some useful projections of the vector time series. The basic idea is simple: A multivariate outlier produces at least a univariate outlier in almost every projected series, and by detecting the univariate outliers, we can identify the multivariate ones. We show that one can often better identify multivariate outliers by applying univariate test statistics to optimal projections than by using multivariate statistics on the original series. We also show that in the presence of an outlier, the directions that maximize or minimize the kurtosis coefficient of the projected series include the direction of the outlier, that is, the direction that maximizes the ratio between the outlier size and the variance of the projected observations. We propose an iterative algorithm based on projections to remove outliers from the observed series.

Pedro Galeano is Teaching Assistant (E-mail: *pedro.galeano@uc3m.es*) and Daniel Peña is Professor (E-mail: *daniel.pena@uc3m.es*), Department of Statistics, Carlos III de Madrid University, Madrid, Spain. Ruey S. Tsay is H. G. B. Alexander Professor of Econometrics and Statistics, Graduate School of Business, University of Chicago, Chicago, IL 60637 (E-mail: *ruey.tsay@gsb.uchicago.edu*). The authors thank the referees and the associate editor for their excellent comments and suggestions. The first two authors acknowledge financial support by MEC grant SEJ2004-03303 and CAM grant HSE/0174/2004. The third author acknowledges financial support by National Science Foundation and the Graduate School of Business, University of Chicago.

The article is organized as follows. In Section 2 we introduce some notation and briefly review the multivariate outlier approach presented by Tsay et al. (2000). In Section 3 we study properties of the univariate outliers introduced by multivariate outliers through projection and discuss some advantages of using projections to detect outliers. In Section 4 we prove that the optimal directions to identify outliers can be obtained by maximizing or minimizing the kurtosis coefficient of the projected series, and in Section 5 we discuss swamping and masking effects. In Section 6 we propose an outlier detection algorithm based on projections. We generalize the procedure to nonstationary time series in Section 7 and investigate the performance of the proposed procedure in a Monte Carlo study in Section 8. Finally, we apply the proposed method to a real data series in Section 9.

2. OUTLIERS IN MULTIVARIATE TIME SERIES

Let $\mathbf{X}_t = (X_{1t}, \ldots, X_{kt})'$ be a k-dimensional vector time series following the vector autoregressive moving average (VARMA) model

$$\boldsymbol{\Phi}(B)\mathbf{X}_t = \mathbf{C} + \boldsymbol{\Theta}(B)\mathbf{E}_t, \qquad t = 1, \ldots, n, \tag{1}$$

where B is the backshift operator such that $B\mathbf{X}_t = \mathbf{X}_{t-1}$, $\boldsymbol{\Phi}(B) = \mathbf{I} - \boldsymbol{\Phi}_1 B - \cdots - \boldsymbol{\Phi}_p B^p$ and $\boldsymbol{\Theta}(B) = \mathbf{I} - \boldsymbol{\Theta}_1 B - \cdots - \boldsymbol{\Theta}_q B^q$ are $k \times k$ matrix polynomials of finite degrees p and q, $\mathbf{C}$ is a k-dimensional constant vector, and $\mathbf{E}_t = (E_{1t}, \ldots, E_{kt})'$ is a sequence of independent and identically distributed (iid) Gaussian random vectors with mean $\mathbf{0}$ and positive-definite covariance matrix $\boldsymbol{\Sigma}$. For the VARMA model in (1), we have the AR representation $\boldsymbol{\Pi}(B)\mathbf{X}_t = \mathbf{C}_{\boldsymbol{\Pi}} + \mathbf{E}_t$, where $\boldsymbol{\Pi}(B) = \boldsymbol{\Theta}(B)^{-1}\boldsymbol{\Phi}(B) = \mathbf{I} - \sum_{i=1}^{\infty} \boldsymbol{\Pi}_i B^i$ and $\mathbf{C}_{\boldsymbol{\Pi}} = \boldsymbol{\Theta}(1)^{-1}\mathbf{C}$ is a vector of constants if $\mathbf{X}_t$ is invertible, and the MA representation $\mathbf{X}_t = \mathbf{C}_{\boldsymbol{\Psi}} + \boldsymbol{\Psi}(B)\mathbf{E}_t$, where $\boldsymbol{\Phi}(1)\mathbf{C}_{\boldsymbol{\Psi}} = \mathbf{C}$ and $\boldsymbol{\Phi}(B)\boldsymbol{\Psi}(B) = \boldsymbol{\Theta}(B)$ with $\boldsymbol{\Psi}(B) = \mathbf{I} + \sum_{i=1}^{\infty} \boldsymbol{\Psi}_i B^i$.

Given an observed time series $\mathbf{Y} = (\mathbf{Y}'_1, \ldots, \mathbf{Y}'_n)'$, where $\mathbf{Y}_t = (Y_{1t}, \ldots, Y_{kt})'$, Tsay et al. (2000) generalized four types of univariate outliers to the vector case in a direct manner using the representation

$$\mathbf{Y}_t = \mathbf{X}_t + \boldsymbol{\alpha}(B)\mathbf{w}I_t^{(h)}, \tag{2}$$

where $I_t^{(h)}$ is a dummy variable such that $I_h^{(h)} = 1$ and $I_t^{(h)} = 0$ if $t \neq h$, $\mathbf{w} = (w_1, \ldots, w_k)'$ is the size of the outlier, and $\mathbf{X}_t$ fol-

Journal of the American Statistical Association
June 2006, Vol. 101, No. 474, Theory and Methods
DOI 10.1198/016214505000001131

lows a VARMA model. The outlier type is defined by the matrix polynomial $\boldsymbol{\alpha}(B)$. If $\boldsymbol{\alpha}(B) = \boldsymbol{\Psi}(B)$, then we have a multivariate innovational outlier (MIO); if $\boldsymbol{\alpha}(B) = \mathbf{I}$, then we have a multivariate additive outlier (MAO); if $\boldsymbol{\alpha}(B) = (1 - B)^{-1}\mathbf{I}$, then we have a multivariate level shift (MLS); and if $\boldsymbol{\alpha}(B) = (\mathbf{I} - \delta\mathbf{I}B)^{-1}$, then we have a multivariate temporary (or transitory) change (MTC), where $0 < \delta < 1$ is a constant. An MIO is an outlier in the innovations and can be interpreted as an internal change in the structure of the series. Its effect on the time series depends on the model, and it can affect several consecutive observations. An MAO can be due to an external cause, such us a typo or measurement error, and affects only a single observation. An MLS changes the mean level of the series and thus, it has a permanent effect. An MTC causes an initial impact, but its effect decreases at a fixed rate in successive observations. In practice, an outlier may produce a complex effect, given by a linear combination of the previously discussed pure effects. Furthermore, different components of $\mathbf{X}_t$ may suffer different outlier effects. For instance, an increase in energy prices can produce a transitory change in one country, without a permanent effect, and a level shift in another country, with a permanent effect. An example of this kind of mixed effects is given in the real data example in Section 9.

The effects of outliers on the innovations are easily obtained when the parameters of the VARMA model for $\mathbf{X}_t$ are known. Using the observed series and the known parameters of the model for $\mathbf{X}_t$, we obtain a series of innovations $\{\mathbf{A}_t\}$ defined by $\mathbf{A}_t = \boldsymbol{\Pi}(B)\mathbf{Y}_t - \mathbf{C}_{\boldsymbol{\Pi}}$, where $\mathbf{Y}_t = \mathbf{X}_t$ and $\mathbf{A}_t = \mathbf{E}_t$ for $t < h$. The relationship between the true white noise innovations, $\mathbf{E}_t$, and the computed innovations, $\mathbf{A}_t$, is given by

$$\mathbf{A}_t = \mathbf{E}_t + \boldsymbol{\Gamma}(B)\mathbf{w}I_t^{(h)}, \tag{3}$$

where $\boldsymbol{\Gamma}(B) = \boldsymbol{\Pi}(B)\boldsymbol{\alpha}(B) = \mathbf{I} - \sum_{i=1}^{\infty} \boldsymbol{\Gamma}_i B^i$. Tsay et al. (2000) showed that when the model is known, the estimation of the size of a multivariate outlier of type i at time h is given by

$$\mathbf{w}_{i,h} = -\left(\sum_{j=0}^{n-h} \boldsymbol{\Gamma}_j'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}_j\right)^{-1}\left(\sum_{j=0}^{n-h} \boldsymbol{\Gamma}_j'\boldsymbol{\Sigma}^{-1}\mathbf{A}_{h+j}\right), \qquad i = I, A, L, T,$$

where $\boldsymbol{\Gamma}_0 = -\mathbf{I}$ and we use the notation $MIO \equiv I$, $MAO \equiv A$, $MLS \equiv L$, and $MTC \equiv T$ for subscripts. The covariance matrix of this estimate is $\boldsymbol{\Sigma}_{i,h} = (\sum_{j=0}^{n-h} \boldsymbol{\Gamma}_j'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}_j)^{-1}$, and it is easy to show that the multivariate test statistic

$$J_{i,h} = \mathbf{w}_{i,h}'\boldsymbol{\Sigma}_{i,h}^{-1}\mathbf{w}_{i,h}, \qquad i = I, A, L, T, \tag{4}$$

will be a noncentral $\chi_k^2(\eta_i)$ random variable with noncentrality parameters $\eta_i = \mathbf{w}'\boldsymbol{\Sigma}_{i,h}^{-1}\mathbf{w}$, for $i = I, A, L, T$. In particular, under the null hypothesis $H_0 : \mathbf{w} = \mathbf{0}$, the distribution of $J_{i,h}$ will be chi-squared with k degrees of freedom. A second statistic proposed by Tsay et al. (2000) is the maximum component statistic defined by $C_{i,h} = \max\{|w_{j,i,h}|/\sqrt{\sigma_{j,i,h}} : 1 \leqslant j \leqslant k\}$, $i = I, A, L, T$, where $w_{j,i,h}$ is the jth element of $\mathbf{w}_{i,h}$ and $\sigma_{j,i,h}$ is the jth element of the main diagonal of $\boldsymbol{\Sigma}_{i,h}$.

In practice, the time index h of the outlier and the parameters of the model are unknown. The parameter matrices are then substituted by their estimates, and the following overall test statistics are defined:

$$J_{\max}(i, h_i) = \max_{1 \le h \le n} J_{i,h} \qquad \text{and}$$
$$C_{\max}(i, h_i^*) = \max_{1 \le h \le n} C_{i,h}, \qquad i = I, A, L, T, \tag{5}$$

where h_i and h_i^* denote the time indices at which the maximum of the joint test statistics and the maximum component statistics occur.

3. OUTLIER ANALYSIS THROUGH PROJECTIONS

In this section we explore the usefulness of projections of a vector time series for outlier detection. First, we study the relationship between the projected univariate models and the multivariate model. Second, we discuss some potential advantages of searching for outliers using the projected series.

3.1 Projections of a VARMA Model

We begin with the properties of a univariate series obtained by the projection of a VARMA series. It is well known that a nonzero linear combination of the components of the VARMA model in (1) follows a univariate ARMA model (see, e.g., Lütkepohl 1993). Let $x_t = \mathbf{v}'\mathbf{X}_t$. If $\mathbf{X}_t$ is a VARMA(p, q) process, then x_t follows an ARMA(p^*, q^*) model with $p^* \leqslant kp$ and $q^* \leqslant (k-1)p + q$. In particular, if $\mathbf{X}_t$ is a VMA(q) series, then x_t is an MA(q^*) with $q^* \leqslant q$, and if $\mathbf{X}_t$ is a VAR(p) process, then x_t follows an ARMA(p^*, q^*) model with $p^* \leqslant kp$ and $q^* \leqslant (k-1)p$. A general form for the model of x_t is

$$\phi(B)x_t = c + \theta(B)e_t, \tag{6}$$

where $\phi(B) = |\boldsymbol{\Phi}(B)|$, $c = \mathbf{v}'\boldsymbol{\Phi}(1)^*\mathbf{C}$, and $\mathbf{v}'\boldsymbol{\Omega}(B)\mathbf{E}_t = \theta(B)e_t$, where $\boldsymbol{\Phi}(B)^*$ is the adjoint matrix of $\boldsymbol{\Phi}(B)$, $\boldsymbol{\Omega}(B) = \boldsymbol{\Phi}(B)^* \times \boldsymbol{\Theta}(B)$, and e_t is a scalar white noise process with variance σ_e^2. The values for $\theta(B)$ and σ_e^2 can be obtained using the algorithm proposed by Maravall and Mathis (1994) that always gives an invertible representation of the univariate process. The AR representation of the univariate model (6) is $\pi(B)x_t = c_\pi + e_t$, where $c_\pi = \theta(1)^{-1}c$ and $\pi(B) = \theta(B)^{-1}\phi(B) = 1 - \sum_{i=1}^{\infty} \pi_i B^i$, and its MA representation is $x_t = c_\psi + \psi(B)e_t$, where $c_\psi = \phi(1)^{-1}c$ and $\psi(B) = \phi(B)^{-1}\theta(B) = 1 + \sum_{i=1}^{\infty} \psi_i B^i$.

When the observed series $\mathbf{Y}_t$ is affected by an outlier, as in (2), the projected series $y_t = \mathbf{v}'\mathbf{Y}_t$ satisfies $y_t = x_t + \mathbf{v}'\boldsymbol{\alpha}(B)\mathbf{w}I_t^{(h)}$. Specifically, if $\mathbf{Y}_t$ has an MAO, then the projected series is $y_t = x_t + \beta I_t^{(h)}$, so that it has an additive outlier of size $\beta = \mathbf{v}'\mathbf{w}$ at $t = h$ provided that $\mathbf{v}'\mathbf{w} \neq 0$. In the same way, the projected series of a vector process with an MLS of size $\mathbf{w}$ will have a level shift with size $\beta = \mathbf{v}'\mathbf{w}$ at time $t = h$. The same result also applies to an MTC. Thus, for the three types of outliers mentioned earlier, the following hypotheses are equivalent,

$$\begin{matrix} H_0 : \mathbf{w} = \mathbf{0} \\ H_A : \mathbf{w} \neq \mathbf{0} \end{matrix} \quad \Leftrightarrow \quad \begin{matrix} H_0^* : \beta = 0 \\ H_A^* : \beta \neq 0 \end{matrix} \qquad \forall\, \mathbf{v} \in S^k - \{\mathbf{v} \perp \mathbf{w}\},$$

because $H_0 = \{\bigcup H_0^* : \mathbf{v} \in S^k - \{\mathbf{v} \perp \mathbf{w}\}\}$, where $S^k = \{\mathbf{v} \in \mathbb{R}^k : \mathbf{v}'\mathbf{v} = 1\}$.

An MIO produces a more complicated effect. It leads to a patch of consecutive outliers with sizes $\mathbf{v}'\mathbf{w}, \mathbf{v}'\boldsymbol{\Psi}_1\mathbf{w}, \ldots, \mathbf{v}'\boldsymbol{\Psi}_{n-h}\mathbf{w}$, starting with time index $t = h$. Assuming that h is not close to n and because $\boldsymbol{\Psi}_j \to \mathbf{0}$, the size of the outlier in the patch tends to 0. In the particular case where $\mathbf{v}'\boldsymbol{\Psi}_i\mathbf{w} = \psi_i\mathbf{v}'\mathbf{w}$, $\forall i = 1, \ldots, n-h$, y_t has an innovational outlier at $t = h$ with size $\beta = \mathbf{v}'\mathbf{w}$. But if $\mathbf{v}'\boldsymbol{\Psi}_i\mathbf{w} = 0$, $i = 1, \ldots, n-h$, then y_t has an additive outlier at $t = h$ with size $\mathbf{w}$, and if $\mathbf{v}'\boldsymbol{\Psi}_i\mathbf{w} = \mathbf{v}'\mathbf{w}$, $i = 0, \ldots, n-h$, then y_t has a level shift at $t = h$ with size $\beta = \mathbf{v}'\mathbf{w}$. Therefore, the univariate series y_t obtained by the projection can be affected by an additive outlier, a patch of outliers, or a level shift.

3.2 Some Advantages of Projection Methods

The first advantage of using projections to search multivariate outliers is simplicity. By using univariate series, we do not need to prespecify a multivariate model for the underlying series in outlier detection. Second, if the model parameters are known, then a convenient projection direction will lead to test statistics that are more powerful than the multivariate ones. Third, as we show later in a Monte Carlo study, the same conclusion continues to hold when the parameters are estimated from the observed series.

To illustrate the second advantage, consider a k-dimensional time series $\mathbf{Y}_t$ generated from the VARMA model in (1) and affected by an MAO, MLS, or MTC at $t = h$. Let $\mathbf{V}$ be a $k \times k$ matrix such that the first column is $\mathbf{w}/\|\mathbf{w}\|$ and the remaining columns consist of $k-1$ vectors that are orthogonal to $\mathbf{w}$ and have unit length. The multivariate series $\mathbf{V}'\mathbf{Y}_t$ is affected by an outlier of size $(\|\mathbf{w}\|, 0, \ldots, 0)'$ at time $t = h$. Note that the outlier affects only the first component of the transformed series. Because the multivariate test statistic $J_{i,h}$ of (4) is invariant to linear transformations, its value is the same for both $\mathbf{Y}_t$ and $\mathbf{V}'\mathbf{Y}_t$ series. Thus all of the information concerning the outlier is in the first component of $\mathbf{V}'\mathbf{Y}_t$, which is the projection of the vector time series in the direction of the outlier. The remaining components of $\mathbf{V}'\mathbf{Y}_t$ are irrelevant for detecting the outlier. Moreover, because the test statistic $J_{i,h}$ is distributed as a noncentral $\chi^2_k(\eta_i)$ with noncentrality parameter $\eta_i = \mathbf{w}'\boldsymbol{\Sigma}^{-1}_{i,h}\mathbf{w}$, its power is given by $Pow(M) = \Pr(J_{i,h} > \chi^2_{k,\alpha})$, where $\chi^2_{k,\alpha}$ is the 100αth percentile of the chi-squared distribution with k degrees of freedom. In contrast, projecting the series $\mathbf{Y}_t$ on a given direction $\mathbf{v}$, we obtain a series y_t affected by an outlier at time $t = h$ and the univariate test statistic $j_{i,h} = \beta^2_{i,h}/\sigma^2_{i,h}$, where $\beta_{i,h}$ is a consistent estimate of β, is distributed as a noncentral $\chi^2_1(\nu_i)$ with noncentrality parameter $\nu_i = \beta^2/\sigma^2_{i,h}$, where $\beta = \mathbf{v}'\mathbf{w}$ and $\sigma^2_{i,h} = \mathrm{var}(\beta_{i,h})$. The power of this test statistic is $Pow(U) = \Pr(j_{i,h} > \chi^2_{1,\alpha})$. Because the detection procedure that we propose is affine equivariant, for simplicity, we assume that $\mathbf{Y}_t = \mathbf{E}_t$ is white noise and that $\boldsymbol{\Sigma} = \mathbf{I}$. If $\mathbf{v} = \mathbf{w}/\|\mathbf{w}\|$, then it is easy to see that for every $\mathbf{w}$, $\eta_i = \nu_i = \mathbf{w}'\mathbf{w}$ for $i = I, A$, $\eta_L = \nu_L = (n-h+1)\mathbf{w}'\mathbf{w}$, and $\eta_T = \nu_T = (1-\delta^{2(n-h+1)})/(1-\delta^2)\mathbf{w}'\mathbf{w}$. The powers, $Pow(U)$ and $Pow(M)$, and their differences $Pow(U) - Pow(M)$ for the case of an MAO are shown in Figure 1 for different values of $\mathbf{w}'\mathbf{w}$. The figure shows that the larger the number of components, the larger the advantage of the projection test over the multivariate one. When the size of the outlier increases, both tests have power close to 1, and hence the difference goes to 0 for large outliers. In Section 8 we show by a simulation study that for correlated series, the same conclusion will continue to hold when the parameters are estimated from the data, although the power will depend on the model.

4. FINDING THE PROJECTION DIRECTIONS

The objective of projection pursuit algorithms is to find interesting features of high-dimensional data in low-dimensional spaces via projections obtained by maximizing or minimizing an objective function termed the *projection index*, which depends on the data and the projection vector. It is commonly

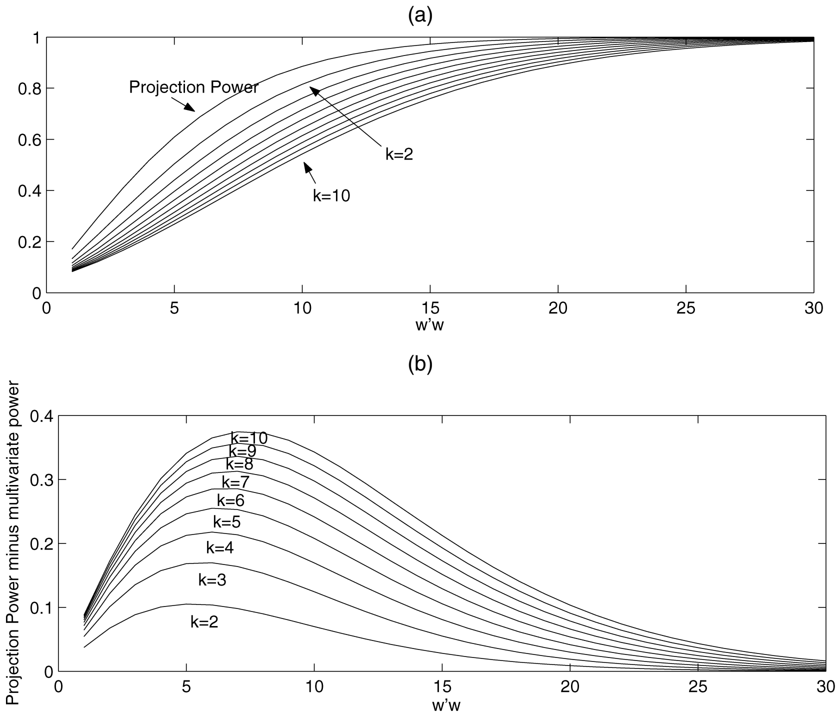

Figure 1. Powers of the Multivariate and the Projection Statistics as a Function of the Outlier Size. (a) Absolute powers; (b) difference of powers.

assumed that the most interesting projections are the farthest ones from normality, showing some unexpected structure such as clusters, outliers or nonlinear relationships among the variables. General reviews of projection pursuit techniques have been given by Huber (1985), Jones and Sibson (1987), and Posse (1995). Peña and Prieto (2001a) proposed a procedure for multivariate outlier detection based on projections that maximize or minimize the kurtosis coefficient of the projected data. Peña and Prieto (2001b) showed that these projected directions are also useful to identify clusters in multivariate data. Pan, Fung, and Fang (2000) suggested using projection pursuit techniques to detect high-dimensional outliers and showed that the projected outlier identifier is a centered Gaussian process on a high-dimensional unit sphere. Pan and Fang (2002) suggested looking for outliers in the projection directions given by the directions that maximizes the sample kurtosis and skewness.

In this section we generalize the application of projections to multivariate time series analysis and define a maximum discrimination direction as the direction that maximizes the size of the univariate outlier, $\mathbf{v}'\mathbf{w}$, with respect to the variance of the projected series. We show that for an MAO, MLS, and MTC, the direction of the outlier is a direction of maximum discrimination, which can be obtained by finding the extremes of the kurtosis coefficient of the projected series. For an MIO, we prove that the direction of the outlier is a maximum discrimination direction for the innovations series, which can be obtained by projecting the innovations.

In what follows, for a time series z_t, we define $\bar{z} = \frac{1}{n}\sum_{t=1}^{n} z_t$ and $\widetilde{z}_t = z_t - \bar{z}$, where n is the sample size. Let $\mathbf{Y}_t$ and $\mathbf{A}_t$ be the observed series and innovations in (2) and (3). For ease in presentation and without loss of generality, we assume that $E(\mathbf{X}_t) = \mathbf{0}$ and $\boldsymbol{\Sigma}_{\mathbf{X}} = \operatorname{cov}(\mathbf{X}_t) = \mathbf{I}$, and define the deterministic variable $\mathbf{R}_t = \boldsymbol{\alpha}(B)\mathbf{w}I_t^{(h)}$. Projecting $\mathbf{Y}_t$ on the direction $\mathbf{v}$, we obtain $y_t = x_t + r_t$, where $r_t = \mathbf{v}'\mathbf{R}_t$. In addition, we have $E[\frac{1}{n}\sum_{t=1}^{n}\mathbf{Y}_t] = E(\overline{\mathbf{Y}}) = \overline{\mathbf{R}}$ and

$$\boldsymbol{\Sigma}_{\mathbf{Y}} = E\left[\frac{1}{n}\sum_{t=1}^{n}(\mathbf{Y}_t - \overline{\mathbf{Y}})(\mathbf{Y}_t - \overline{\mathbf{Y}})'\right] = E\left[\frac{1}{n}\sum_{t=1}^{n}\widetilde{\mathbf{Y}}_t\widetilde{\mathbf{Y}}_t'\right]$$
$$= \mathbf{I} + \boldsymbol{\Sigma}_{\mathbf{R}},$$

where $\boldsymbol{\Sigma}_{\mathbf{R}} = \frac{1}{n}\sum_{t=1}^{n}\widetilde{\mathbf{R}}_t\widetilde{\mathbf{R}}_t'$. Using the results of Rao (1973, p. 60), the maximum of $(\mathbf{v}'\mathbf{w})^2/(\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{Y}}\mathbf{v})$ under the constraint $\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{Y}}\mathbf{v} = 1$ is $\mathbf{v} = \boldsymbol{\Sigma}_{\mathbf{Y}}\mathbf{w}$. For the cases of MAO, MLS, and MTC, we have $\boldsymbol{\Sigma}_{\mathbf{Y}} = \mathbf{I} + \beta_i\mathbf{w}\mathbf{w}'$ with β_i given by

$$\beta_A = \frac{n-1}{n^2},$$
$$\beta_L = \frac{n-h+1}{n}\left(\frac{h-1}{n}\right),$$
$$\beta_T = \frac{1}{n}\left[\left(\frac{1-\delta^{2(n-h+1)}}{1-\delta^2}\right) - \frac{1}{n}\left(\frac{1-\delta^{(n-h+1)}}{1-\delta}\right)^2\right],$$

and $\mathbf{v} = (1 + \beta_i\mathbf{w}'\mathbf{w})\mathbf{w}$. Thus the interesting direction $\mathbf{v}$ of projection is proportional to $\mathbf{w}$. The same result also holds in the MIO case for the maximum of $(\mathbf{v}'\mathbf{w})^2/(\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{A}}\mathbf{v})$ under the constraint $\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{A}}\mathbf{v} = 1$, where $\boldsymbol{\Sigma}_{\mathbf{A}}$ is the expected value of the covariance matrix of the innovations $\mathbf{A}_t$.

We prove next that the direction of the outlier, $\mathbf{w}$, can be found by maximizing or minimizing the kurtosis coefficient of the projected series. Toward this end, we need some preliminary results, the proofs of which are given in the Appendix.

Lemma 1. The kurtosis coefficient $\gamma_y(\mathbf{v})$ of the project series $y_t = \mathbf{v}'\mathbf{Y}_t$, under the restriction $\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{Y}}\mathbf{v} = 1$, is

$$\gamma_y(\mathbf{v}) = 3 - 3(\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{R}}\mathbf{v})^2 + \omega_r(\mathbf{v}), \tag{7}$$

where $\omega_r(\mathbf{v}) = \frac{1}{n}\sum_{t=1}^{n}\widetilde{r}_t^4$.

Lemma 2. The extreme directions of the kurtosis coefficient of y_t under the constraint $\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{Y}}\mathbf{v} = 1$ are given by the eigenvectors of the matrix $\sum_{t=1}^{n}\beta_t(\mathbf{v})\widetilde{\mathbf{R}}_t\widetilde{\mathbf{R}}_t'$ associated with eigenvalues $\mu(\mathbf{v}) = n(\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{R}}\mathbf{v})^2(\gamma_r(\mathbf{v}) - 3)$, where $\beta_t(\mathbf{v}) = (\mathbf{v}'\widetilde{\mathbf{R}}_t)^2 - 3(\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{R}}\mathbf{v}) - \mu(\mathbf{v})/n$ and $\gamma_r(\mathbf{v})$ is the kurtosis coefficient of $r_t = \mathbf{v}'\mathbf{R}_t$. Moreover, the directions that maximize or minimize the kurtosis coefficient are given by the eigenvectors associated with the largest and the smallest eigenvalues $\mu(\mathbf{v})$.

The following theorem shows the usefulness of the extreme directions of the kurtosis coefficient of y_t.

Theorem 1. Suppose that $\mathbf{X}_t$ is a stationary VARMA(p,q) process and $\mathbf{Y}_t = \mathbf{X}_t + \boldsymbol{\alpha}(B)\mathbf{w}I_t^{(h)}$ as in (2), then we have the following results:

a. For an MAO, the kurtosis coefficient of y_t is maximized when $\mathbf{v}$ is proportional to $\mathbf{w}$ and is minimized when $\mathbf{v}$ is orthogonal to $\mathbf{w}$.

b. For an MTC, the kurtosis coefficient of y_t is maximized or minimized when $\mathbf{v}$ is proportional to $\mathbf{w}$ and is minimized or maximized when $\mathbf{v}$ is orthogonal to $\mathbf{w}$.

c. For an MLS, the kurtosis coefficient of y_t is minimized when $\mathbf{v}$ is proportional to $\mathbf{w}$ and is maximized when $\mathbf{v}$ is orthogonal to $\mathbf{w}$ if

$$h \in \left(\left(1 + \frac{1}{2}\left(1 - \frac{1}{\sqrt{3}}\right)n, 1 + \frac{1}{2}\left(1 + \frac{1}{\sqrt{3}}\right)n\right)\right).$$

Otherwise, the kurtosis coefficient of y_t is maximized when $\mathbf{v}$ is proportional to $\mathbf{w}$ and is minimized when $\mathbf{v}$ is orthogonal to $\mathbf{w}$.

Theorem 1 has two important implications. First, for an MAO, MLS, or MTC, one of the directions obtained by maximizing or minimizing the kurtosis coefficient is the direction of the outlier. Second, these directions are obtained without the information of the time index at which the outlier occurs. Given the characteristics of innovational outliers, it is natural to think that the direction of the outlier can be obtained by focusing on the innovations series. This is indeed the case.

Corollary 1. If $\mathbf{X}_t$ is a stationary VARMA(p,q) process and $\mathbf{Y}_t = \mathbf{X}_t + \boldsymbol{\Psi}(B)\mathbf{w}I_t^{(h)}$ as in (2) with $\mathbf{A}_t = \mathbf{E}_t + \mathbf{w}I_t^{(h)}$, then the kurtosis coefficient of $a_t = \mathbf{v}'\mathbf{A}_t$ is maximized when $\mathbf{v}$ is proportional to $\mathbf{w}$ and is minimized when $\mathbf{v}$ is orthogonal to $\mathbf{w}$.

5. MASKING AND SWAMPING EFFECTS

In the presence of multiple outliers, it would be of limited value if one considered only the projections that maximize or minimize the kurtosis coefficient because of the potential problem of masking effects. For instance, a projection might effectively reveal one outlier but almost eliminate the effects of other outliers. To overcome such a difficulty, we present an iterative procedure for analyzing a set of $2k$ orthogonal directions consisting of (a) the direction that maximizes the kurtosis coefficient, (b) the direction that minimizes the kurtosis coefficient, and (c) two sets of $k-1$ directions orthogonal to (a) and (b). Our motivation for using these orthogonal directions is twofold. First, the results of Section 4 reveal that in some cases the directions of interest are orthogonal to those that maximize or minimize the kurtosis coefficient of the projected series; second, these directions ensure nonoverlapping information, so that if the effect of an outlier is almost hidden in one direction, then it may be revealed by one of the orthogonal directions. Furthermore, after removing the effects of outliers detected in the original set of $2k$ orthogonal directions, we propose to iterate the analysis using new directions until no more outliers are detected. Therefore, if a set of outliers are masked in one direction, then they may be revealed either in one of the orthogonal directions or in a later iteration after removing detected outliers. To illustrate, we analyze in detail the cases of a series with two MAOs and an MAO and an MLS.

Theorem 2. Suppose that $\mathbf{X}_t$ is a stationary VARMA(p,q) process, $\mathbf{Y}_t$ is the observed vector series, and $y_t = \mathbf{v}'\mathbf{Y}_t$ is a projected scalar series.

a. Let $\mathbf{Y}_t = \mathbf{X}_t + \mathbf{w}_1 I_t^{(h_1)} + \mathbf{w}_2 I_t^{(h_2)}$, with $h_1 < h_2$. There are three possibilities as follows:

1. If $\mathbf{w}_1$ and $\mathbf{w}_2$ are proportional to each other, then the kurtosis coefficient of y_t is maximized when $\mathbf{v}$ is proportional to $\mathbf{w}_i$ and is minimized when $\mathbf{v}$ is orthogonal to $\mathbf{w}_i$.
2. If $\mathbf{w}_1$ and $\mathbf{w}_2$ are orthogonal, then the kurtosis coefficient is maximized when $\mathbf{v}$ is proportional to the outlier with larger Euclidean norm and minimized when $\mathbf{v}$ is in one of the orthogonal directions.
3. Let φ be the angle between $\mathbf{w}_1$ and $\mathbf{w}_2$, then the kurtosis coefficient is approximately maximized when $\mathbf{v}$ is the direction of the outlier that gives the maximum of $\{\|\mathbf{w}_1\|/\|\mathbf{w}_2\|\cos\varphi, \|\mathbf{w}_2\|/\|\mathbf{w}_1\|\cos\varphi\}$, where the quantity denotes the ratio between the norm of outlier $\mathbf{w}_i$ and the length of the projection of outlier $\mathbf{w}_j$ on $\mathbf{w}_i$, where $j \neq i$.

b. Let $\mathbf{Y}_t = \mathbf{X}_t + \mathbf{w}_1 I_t^{(h_1)} + \mathbf{w}_2 S_t^{(h_2)}$ with $h_1 < h_2$. Then the kurtosis coefficient of y_t is maximized or minimized when $\mathbf{v}$ is proportional to $\mathbf{w}_2$, and it is minimized or maximized when $\mathbf{v}$ is orthogonal to $\mathbf{w}_2$.

In the case of two MAOs, Theorem 2 shows that if both outliers are proportional, then the maximum of the kurtosis coefficient is obtained in the direction of the outliers, and if the outliers are orthogonal, then the maximum of the kurtosis is obtained in the direction of the outlier with larger Euclidean norm and the minimum is obtained in an orthogonal direction. Thus an orthogonal direction of the one that gives the maximum of the kurtosis coefficient will reveal the presence of the outlier with smaller norm. If the outliers are neither proportional nor orthogonal to each other, then the direction of the outlier with the largest projection on the direction of the other outlier will produce a maximum kurtosis coefficient. Note that the projection of $\mathbf{w}_2$ on the direction of $\mathbf{w}_1$ is given by $\|\mathbf{w}_2\|\cos\varphi$. The ratio $\|\mathbf{w}_1\|/\|\mathbf{w}_2\|\cos\varphi$ is large if $\|\mathbf{w}_1\|$ is large enough compared with $\|\mathbf{w}_2\|$, in which case the kurtosis coefficient will be maximized in the direction of $\mathbf{w}_1$, or if $\cos\varphi$ is small enough, in which case an orthogonal direction to $\mathbf{w}_1$ will reveal $\mathbf{w}_2$.

Second, the direction of the MLS gives the maximum or the minimum of the kurtosis coefficient of the projected series. Thus the MLS will be revealed in the directions that maximize or minimize the kurtosis coefficient of y_t. Projecting in the orthogonal directions will eliminate the effect of this level shift and reveal the second outlier. If the statistics for this second outlier are not significant, then we remove the effect of the level shift, and a new set of directions may reveal the second outlier.

In contrast, outlier detection procedures are sometimes affected by swamping effects; that is, one outlier affects the series in such a way that other "good" data points appear like outliers. The procedure that we propose in the next section includes several steps to avoid swamping effects, which can appear in the univariate searches using the projection statistics. The idea is to delete nonsignificant outliers after a joint estimation of the parameters and detected outliers. We clarified this in the next section.

6. ALGORITHMS FOR OUTLIER DETECTION

Here we propose a sequential procedure for outlier detection based on the directions that mimimize and maximize the kurtosis coefficient of the projections. We refer to these directions as the *optimal projections*. The procedure is divided into four steps: (1) Obtain these directions, (2) search for outliers in the projected univariate time series, (3) remove the effect of all detected outliers by using an approximated multivariate model, and (4) iterate the previous steps applied to the cleaned series until no more outliers are found. Note that in step 2, the detection is carried out in two stages: First, MLSs are identified, and second, MIOs, MAOs, and MTCs are found. Finally, a vector model is identified for the cleaned time series, and the outlier effects and model parameters are jointly estimated. The fitted model is refined if necessary, for example, removing insignificant outliers, if any.

6.1 Computation of the Projection Directions

We use the procedure of Peña and Prieto (2001b) to construct the $2k$ projection directions of interest. For an observed vector series $\mathbf{Y}_t$, our goal here is to obtain the maximum and minimum of the kurtosis coefficient of the projected series and the orthogonal directions of the optimal projections. Toward this end, consider the following procedure:

1. Let $m = 1$ and $\mathbf{Z}_t^{(m)} = \mathbf{Y}_t$.
2. Let $\mathbf{\Sigma}_{\mathbf{Z}}^{(m)} = \frac{1}{n}\sum_{t=1}^{n} \widetilde{\mathbf{Z}}_t^{(m)} \widetilde{\mathbf{Z}}_t^{(m)\prime}$, and find $\mathbf{v}_m$ such that
$$\mathbf{v}_m = \mathop{\arg\max}_{\mathbf{v}_m' \mathbf{\Sigma}_{\mathbf{Z}}^{(m)} \mathbf{v}_m = 1} \frac{1}{n}\sum_{t=1}^{n} \left(\mathbf{v}_m' \widetilde{\mathbf{Z}}_t^{(m)}\right)^4. \tag{8}$$
3. If $m = k$, then stop; otherwise, define $\mathbf{Z}_t^{(m+1)} = (\mathbf{I} - \mathbf{v}_m \mathbf{v}_m' \mathbf{\Sigma}_{\mathbf{Z}}^{(m)})\mathbf{Z}_t^{(m)}$; that is, $\mathbf{Z}_t^{(m+1)}$ is the projection of the observations in an orthogonal direction to $\mathbf{v}_m$. Let $m = m+1$ and go to step 2.
4. Repeat the same procedure to minimize the objective function in (8), to obtain another set of k directions, namely $\mathbf{v}_{k+1}, \ldots, \mathbf{v}_{2k}$.

A key step of the foregoing algorithm is to solve the optimization problem in (8). Toward this end, we use a modified Newton method to solve the system given by the first-order optimality conditions, $\nabla\gamma_z(\mathbf{v}) - 2\lambda\boldsymbol{\Sigma}_{\mathbf{Z}}^{(m)}\mathbf{v} = 0$ and $\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{Z}}^{(m)}\mathbf{v} - 1 = 0$, by means of linear approximations. (See Peña and Prieto 2001b for technical details of the method.) One relevant issue is that the proposed procedure is affine equivariant; that is, the method selects equivalent directions for series modified by an affine transformation.

6.2 Searching for Univariate Outliers

The most commonly used tests for outlier detection in univariate time series are the likelihood ratio test (LRT) statistics, $\lambda_{i,h}$, $i = I, A, L, T$ (see Chang and Tiao 1983; Tsay 1988). Because the location of the outlier and the parameters of the model are unknown, the estimated parameters are used to define the overall test statistics $\Lambda(i, h_i) = \max\{|\lambda_{i,t}|, 1 \leq t \leq n\}$, $i = I, A, L, T$. Using these statistics, Chang and Tiao (1983) proposed an iterative algorithm for detecting innovational and additive outliers. Tsay (1988) generalized the algorithm to detect level shifts and transitory changes (see Chen and Liu 1993; Sánchez and Peña 2003 for additional extensions).

In this article we consider a different approach. There is substantial evidence that using the same critical values for all the LRT statistics can easily misidentify a level shift as an innovative outlier (see Balke 1993; Sánchez and Peña 2003). The latter authors showed that the critical values for the LRT statistic for detecting level shifts are different from those for testing additive or innovative outliers. Therefore, we propose to identify the level shifts in a series before checking for other types of outlier. Toward this end, it is necessary to develop a procedure that is capable of detecting level shifts in the presence of the other types of outliers. Using the notation of Section 3.1, Bai (1994) proposed the cusum statistic,

$$C_t = \frac{t}{\sqrt{n}\psi(1)\sigma_e}\left(\frac{1}{t}\sum_{i=1}^{t} y_i - \bar{y}\right), \tag{9}$$

to test for a level shift at $t = h+1$ in a linear process and showed that the statistic converges weakly to a standard Brownian bridge on $[0, 1]$. In practice, the quantity $\psi(1)\sigma_e$ is replaced by a consistent estimator,

$$\widehat{\psi(1)\sigma_e} = \left[\widehat{\gamma(0)} + 2\sum_{i=1}^{K}\left(1 - \frac{|i|}{K}\right)\widehat{\gamma(i)}\right]^{1/2},$$

where $\widehat{\gamma(i)} = \mathrm{cov}(y_t, y_{t-i})$ and K is a quantity such that $K \to \infty$ and $K/n \to 0$ as $n \to \infty$. Under the assumption of no level shifts in the sample, the statistic $\max_{1\leq t\leq n}|C_t|$ is asymptotically distributed as the supremum of the absolute value of a Brownian bridge with cumulative distribution function $F(x) = 1 + 2\sum_{i=1}^{\infty}(-1)^i e^{-2i^2x^2}$, for $x > 0$. The cusum statistic (9) has several advantages over the LRT statistic for detecting level shifts. First, it is not necessary to specify the order of the ARMA model, which can be difficult in the presence of level shifts. Second, as shown in Section 8, this statistic seems to be more powerful than the LRT in all the models considered. Third, the statistic (9) seems to be robust to the presence of other outliers whereas the LRT statistic is not.

6.2.1 Level Shift Detection. Given the $2k$ projected univariate series $y_{t,i} = \mathbf{v}_i'\mathbf{Y}_t$ for $i = 1, \ldots, 2k$, we propose an iterative procedure to identify level shifts based on the algorithm proposed by Inclán and Tiao (1994) and Carnero, Peña, and Ruiz (2003) for detecting variance changes and level shifts in a white noise series. Let H be the prespecified minimum distance between two level shifts. The proposed algorithm divides the series into pieces after detecting a level shift and proceeds as follows:

1. Let $t_1 = 1$ and $t_2 = n$. Obtain
$$D_L = \max_{1\leq i\leq 2k}\max_{t_1\leq t\leq t_2}|C_t^i|, \tag{10}$$
where C_t^i is the statistic (9) applied to the ith projected series for $i = 1, \ldots, 2k$. Let
$$(i_{\max}, t_{\max}) = \arg\max_{1\leq i\leq 2k}\arg\max_{t_1\leq t\leq t_2}|C_t^i|. \tag{11}$$
If $D_L > D_{L,\alpha}$, where $D_{L,\alpha}$ is the critical value for the significance level α, then there is a possible level shift at $t = t_{\max} + 1$, and we go to step 2a. If $D_L < D_{L,\alpha}$, then there is no level shift in the series, and the algorithm stops.

2a. Define $t_2 = t_{\max}$ of step 1, and obtain new values of D_L and $(i_{\max}, t_{\max})$ of (10) and (11). If $D_L > D_{L,\alpha}$, and $t_2 - t_{\max} > H$, then we redefine $t_2 = t_{\max}$ and repeat step 2a until $D_L < D_{L,\alpha}$ or $t_2 - t_{\max} \leq H$. Define $t_{\text{first}} = t_2$, where t_2 is the last time index that attains the maximum of the cusum statistic that is larger than $D_{L,\alpha}$ and satisfies $t_2 - t_{\max} > H$. The point $t_{\text{first}} + 1$ is the first time point with a possible level shift.

2b. Define $t_1 = t_{\max}$ of step 1 and $t_2 = n$, and obtain new values of D_L and $(i_{\max}, t_{\max})$ of (10) and (11). If $D_L > D_{L,\alpha}$ and $t_{\max} - t_1 > H$, then we redefine $t_1 = t_{\max}$ and repeat step 2b until $D_L < D_{L,\alpha}$ or $t_{\max} - t_1 \leq H$. Define $t_{\text{last}} = t_1$, where t_1 is the last time index that attains the maximum of the cusum statistics that is larger than $D_{L,\alpha}$ and satisfies $t_{\max} - t_1 > H$. The point $t_{\text{last}} + 1$ is the last time point with a possible level shift.

2c. If $t_{\text{last}} - t_{\text{first}} < H$, then there is just a level shift, and the algorithm stops. If not, then keep both values as possible changepoints and repeat steps 2a and 2b for $t_1 = t_{\text{first}}$ and $t_2 = t_{\text{last}}$ until no more possible changepoints are detected. Then go to step 3.

3. Define a vector $\mathbf{h}^L = (h_0^L, \ldots, h_{r_L+1}^L)$, where $h_0^L = 1$, $h_{r_L+1}^L = n$ and $h_1^L < \cdots < h_{r_L}^L$ are the changepoints detected in step 2. Obtain the statistic D_L in each subinterval (h_i^L, h_{i+2}^L) and check its statistical significance. If it is not significant, then eliminate the corresponding possible changepoint. Repeat step 3 until the number of possible changepoints remains unchanged and the time indexes of changepoints are the same between iterations. Removing $h_0^L = 1$ and $h_{r_L+1}^L = n$ from the final vector of time indexes, we obtain r_L level shifts in the series at time indexes $h_i^L + 1$ for $i = 1, \ldots, L$.

4. Let $\{h_1^L, \ldots, h_{r_L}^L\}$ be the time indexes of r_L detected level shifts. To remove the impacts of level shifts, we fit the model
$$(\mathbf{I} - \boldsymbol{\Pi}_1 B - \cdots - \boldsymbol{\Pi}_{\widehat{p}} B^{\widehat{p}})\mathbf{Y}_t^* = \mathbf{A}_t^*, \tag{12}$$

where $\mathbf{Y}_t^* = \mathbf{Y}_t - \sum_{i=1}^{r_L} \mathbf{w}_i S_t^{(h_i^L)}$ and the order $\widehat{p}$ is chosen such that

$$\widehat{p} = \underset{0 \le p \le p_{\max}}{\arg\max}\, AIC(p) = \underset{0 \le p \le p_{\max}}{\arg\max} \left\{ \log|\widehat{\boldsymbol{\Sigma}}_p| + 2\frac{k^2 p}{n} \right\},$$

where $\widehat{\boldsymbol{\Sigma}}_p = \frac{1}{n-2p-1} \sum_{t=p+1}^{n} \mathbf{A}_t^* \mathbf{A}_t^{*\prime}$ and $p_{\max}$ is a prespecified upper bound. If some of the effects of level shifts are not significant, then we remove the least significant one from the model in (12) and reestimate the effects of the remaining $r_L - 1$ level shifts. This process is repeated until all of the level shifts are significant.

Some comments on the proposed procedure are in order. First, the statistic D_L is the maximum of dependent random variables and has an intractable distribution. We obtain critical values by simulation in the next section. Second, the test statistics (9) are highly correlated for close observations. Thus consecutive large values of $\mathbf{C}_t$ might be caused by a single level shift. To avoid overdetection, we do not allow two level shifts to be too close by using the number H in steps 2 and 3. In the simulations and real data example, we chose $H = 10$ and found that it works well.

6.2.2 An Algorithm for Outlier Detection. Using the level-shift adjusted series, we use the following procedure to detect additive outliers, transitory changes, and innovative outliers in the $2k$ projected univariate series $y_{t,i} = \mathbf{v}_i' \mathbf{Y}_t^*$ and their associated innovational series $a_{t,i} = \mathbf{v}_i' \mathbf{A}_t^*$ for $i = 1, \dots, 2k$.

1. For each projected series $y_{t,i}$, fit an AR(p) with p selected by the Akaike information criterion (AIC) and compute the LRT statistics $\lambda_{A,t}^i$ and $\lambda_{T,t}^i$. In addition, compute the LRT statistics $\lambda_{I,t}^i$ using the associated innovational series $a_{t,i}$. This leads to the maximum statistics

$$\Lambda_A = \max_{1 \le i \le 2k} \max_{1 \le t \le n} |\lambda_{A,t}^i|,$$
$$\Lambda_T = \max_{1 \le i \le 2k} \max_{1 \le t \le n} |\lambda_{T,t}^i|, \quad \text{and} \tag{13}$$
$$\Lambda_I = \max_{1 \le i \le 2k} \max_{1 \le t \le n} |\lambda_{I,t}^i|.$$

2. For $i = A, T$, and I, let $\Lambda_{A,\alpha}$, $\Lambda_{T,\alpha}$, and $\Lambda_{I,\alpha}$ be the critical values for a predetermined significance level α. If $\Lambda_i < \Lambda_{i,\alpha}$ for $i = I, A, T$, then no outliers are found, and the algorithm stops. If $\Lambda_i > \Lambda_{i,\alpha}$ for only one i, then identify an outlier of type i and remove its effect using the multivariate parameter estimates. If $\Lambda_i > \Lambda_{i,\alpha}$ for more than one i, then identify the outlier based on the most significant test statistic and remove its effect using the multivariate parameter estimates. Repeat steps 1 and 2 until no more outliers are detected.
3. Let $\{h_1^A, \dots, h_{r_A}^A\}$, $\{h_1^T, \dots, h_{r_T}^T\}$, and $\{h_1^I, \dots, h_{r_I}^I\}$ be the time indexes of the r_A, r_T, and r_I detected additive outliers, transitory changes, and innovative outliers. Estimate jointly the model parameters and the detected outliers for the series $\mathbf{Y}_t^*$, $(\mathbf{I} - \boldsymbol{\Pi}_1 B - \cdots - \boldsymbol{\Pi}_{\widehat{p}} B^{\widehat{p}}) \mathbf{Y}_t^{**} = \mathbf{A}_t^{**}$, where

$$\mathbf{Y}_t^{**} = \mathbf{Y}_t^* - \sum_{i_A=1}^{r_A} \mathbf{w}_{i_A} I_t^{(h_{i_A}^A)} - \sum_{i_T=1}^{r_T} \frac{\mathbf{w}_{i_T}}{1 - \delta B} I_t^{(h_{i_T}^T)}$$

and

$$\mathbf{A}_t^{**} = \mathbf{A}_t^* - \sum_{i_I=1}^{r_I} \mathbf{w}_{i_I} I_t^{(h_{i_I}^I)}.$$

If some of the outlier effects become insignificant, then remove the least significant outlier and reestimate the model. Repeat this process until all of the remaining outliers are significant.

In Section 8 we obtain critical values for the test statistics $\lambda_{A,t}^i$, $\lambda_{T,t}^i$, and $\lambda_{I,t}^i$ through simulation.

6.3 Final Joint Estimation of Parameters, Level Shifts, and Outliers

Finally, we perform a joint estimation of the model parameters, the level shifts, and the outliers detected using the equation $(\mathbf{I} - \boldsymbol{\Pi}_1 B - \cdots - \boldsymbol{\Pi}_{\widehat{p}} B^{\widehat{p}}) \mathbf{Z}_t = \mathbf{D}_t$, where

$$\mathbf{Z}_t = \mathbf{Y}_t - \sum_{i_L=1}^{r_L} \mathbf{w}_{i_L} S_t^{(h_{i_L}^L)} - \sum_{i_A=1}^{r_A} \mathbf{w}_{i_A} I_t^{(h_{i_A}^A)} - \sum_{i_T=1}^{r_T} \frac{\mathbf{w}_{i_T}}{1 - \delta B} I_t^{(h_{i_T}^T)}$$

and

$$\mathbf{D}_t = \mathbf{A}_t - \sum_{i_I=1}^{r_I} \mathbf{w}_{i_I} I_t^{(h_{i_I}^I)},$$

and $\{h_1^L, \dots, h_{r_L}^L\}$, $\{h_1^A, \dots, h_{r_A}^A\}$, $\{h_1^T, \dots, h_{r_T}^T\}$, and $\{h_1^I, \dots, h_{r_I}^I\}$ are the time indexes of the r_L, r_A, r_T, and r_I detected level shifts, additive outliers, transitory changes and innovative outliers. If some effect is found to be not significant at a given level, then we remove the least significant one and repeat the joint estimation until all of the effects are significant.

Some comments on the proposed procedure are as follows. First, as mentioned in Section 2, an outlier can be a combination of different effects. This does not cause any problem for the proposed procedure, because it allows for multiple outlier detections at a given time point either in different projection directions or in successive iterations. The real data example of Section 9 demonstrates how the proposed procedure handles such a situation. Second, the procedure includes several steps to avoid swamping effects. Specifically, after detecting level shifts, we fit an autoregression and remove any nonsignificant level shifts. We repeat the same step after detecting the other types of outlier. Finally, we perform a joint estimation of model parameters and detected level shifts and outliers to remove any nonsignificant identification of level shift or outliers.

7. THE NONSTATIONARY CASE

In this section we study the case when the time series is unit-root nonstationary. Assume that $\mathbf{X}_t \sim I(d_1, \dots, d_k)$, where the d_i's are nonnegative integers denoting the degrees of differencing of the components of $\mathbf{X}_t$. Let $d = \max(d_1, \dots, d_k)$ and consider first the case of $d = 1$, which we denote simply by $\mathbf{X}_t \sim I(1)$. For such a series, in addition to the outliers introduced by Tsay et al. (2000), we also entertain the multivariate ramp shift (MRS) defined as $\mathbf{Y}_t = \mathbf{X}_t + \mathbf{w} R_t^{(h)}$, where $R_t^{(h)} = (I - B)^{-1} S_t^{(h)}$ with $S_t^{(h)}$ being a step function at the time index h (i.e., $S_t^{(h)} = 1$ if $t \ge h$ and $= 0$ otherwise). This outlier implies a slope change in the multivariate series, and it may occur in an $I(1)$ series. It is not considered in the stationary case because the series has no time slope. Consequently, for an MRS, we assume that it applies only to the components of $\mathbf{Y}_t$ with $d_j = 1$; that is, the size of the outlier $\mathbf{w} = (w_1, \ldots, w_k)'$ satisfies $w_j = 0$ if $d_j = 0$.

The series $\mathbf{X}_t$ can be transformed into stationarity by taking the first difference of its components even though d_j might be zero for some j. As we show later, this is not a drawback for our method. The first differencing affects the existing outliers as follows. In the MIO case, $(\mathbf{I} - \mathbf{B})\mathbf{Y}_t = (\mathbf{I} - \mathbf{B})\mathbf{X}_t + \widetilde{\boldsymbol{\Psi}}(B)\mathbf{w}I_t^{(h)}$, where $\widetilde{\boldsymbol{\Psi}}(B) = (\mathbf{I} - \mathbf{B})\boldsymbol{\Psi}(B)$. Therefore, an MIO produces an MIO in the differenced series. In the MAO case, $(\mathbf{I} - \mathbf{B})\mathbf{Y}_t = (\mathbf{I} - \mathbf{B})\mathbf{X}_t + \mathbf{w}(I_t^{(h)} - I_{t-1}^{(h)})$, producing two consecutive MAOs with the same size but opposite signs. In the MLS case, $(\mathbf{I} - \mathbf{B})\mathbf{Y}_t = (\mathbf{I} - \mathbf{B})\mathbf{X}_t + \mathbf{w}I_t^{(h)}$, resulting in an MAO of the same size. In the MTC case, $(\mathbf{I} - \mathbf{B})\mathbf{Y}_t = (\mathbf{I} - \mathbf{B})\mathbf{X}_t + \zeta(B)\mathbf{w}I_t^{(h)}$, where $\zeta(B) = 1 + \zeta_1 B + \zeta_2 B^2 + \cdots$ such that $\zeta_j = \delta^{j-1}(1 - \delta)$. Thus an MTC produces an MTC with decreasing coefficients ζ_j. In the MRS case, $(\mathbf{I} - \mathbf{B})\mathbf{Y}_t = (\mathbf{I} - \mathbf{B})\mathbf{X}_t + \mathbf{w}S_t^{(h)}$, which produces an MLS with same size.

The results of Section 4 can be extended to include the aforementioned outliers induced by differencing. For instance, Theorem 1 shows that the directions that maximize or minimize the kurtosis coefficient of the projected series under the presence of two consecutive MAOs with the same size but opposite signs are the direction of the outlier or a direction orthogonal to it. Therefore, in the $I(1)$ case, we propose a procedure similar to that of the stationary case for the first differenced series. This procedure consists of the following steps:

1. Take the first difference of all of the components of $\mathbf{Y}_t$. Check for MLS as in Section 6.2.1. All of the level shifts detected in the differenced series are incorporated as ramp shifts in the original series and are estimated jointly with the model parameters. If any ramp shift is not significant, then remove it from the model, and repeat the detecting process until all of the ramp shifts are significant. This leads to a series, $\mathbf{Y}_t^* = \mathbf{Y}_t - \sum_{i=1}^{r_R} \mathbf{w}_i R_t^{(h)}$, that is free of ramp shifts.
2. Take the first difference of all of the components of $\mathbf{Y}_t^*$. The series $(\mathbf{I} - \mathbf{B})\mathbf{Y}_t^*$ may be affected by MIOs, two consecutive MAOs, MAOs, and MTCs. Then proceed as in Section 6.2.2. All of the outliers detected in the differenced series are incorporated by the corresponding effects in the original series and are estimated jointly with the model parameters. If any of the outliers becomes insignificant, then remove it from the model. Repeat the process until all of the outliers are statistically significant.

The procedure can also be applied to series that have $d_j = 0$ for some components and to cointegrated series. In these cases $\nabla\mathbf{Y}_t$ is overdifferenced, implying that its MA component contains some unit roots. Nevertheless, this is not a problem for the proposed procedure. If the series is not cointegrated, then any projection direction provides an univariate series without unit roots in its MA component. If the series is cointegrated, then the directions of the outliers will in general be different from the directions of cointegration. In other words, if $\mathbf{v}$ is a vector obtained by maximizing or minimizing the kurtosis coefficient, then it is unlikely to be a cointegration vector, and $\mathbf{v}'\nabla\mathbf{Y}_t = \nabla(\mathbf{v}'\mathbf{Y}_t)$ is stationary and invertible because $\mathbf{v}'\mathbf{Y}_t$ is a nonstationary series. However, if the series are cointegrated, then the final estimation should be carried out using the error correction model of Engle and Granger (1987). Note that if $\mathbf{v}$ is the cointegration vector, then $\mathbf{v}'\mathbf{Y}_t$ is stationary and $\nabla\mathbf{v}'\mathbf{Y}_t$ is overdifferenced. Although no relationship is expected between the outlier directions and the cointegration vector, we have verified, by using Monte Carlo simulations, that the probability of finding the cointegration relationship as a solution of the optimization algorithm is very low. Specifically, we generated 10,000 series from a vector AR(1) model with two components and a cointegration relationship and found the directions in (8). To compare the directions with the cointegration vector, we calculated the absolute value of the cosine of the angle between these two directions. The average value of this cosine is .62 with variance .09. It is easy to show that if the angle has a uniform distribution in the interval $(0, \pi)$, then the distribution of the cosine of the angle has mean .63 and variance .09. Next, we repeated the same experiment with the same series but affected by outliers, level shifts, or transitory changes, and obtained in every case that the mean of the angles between the direction found and the cointegrating direction is the one that exits between the direction of the outlier and the cointegration direction. Therefore, we conclude that there should be no confusion between the cointegration vectors and the directions that maximize or minimize the kurtosis coefficient of the projected series.

Consider next the case where $d = 2$. Define a multivariate quadratic shift (MQS) as $\mathbf{Y}_t = \mathbf{X}_t + \mathbf{w}Q_t^{(h)}$ where $Q_t^{(h)} = (\mathbf{I} - \mathbf{B})^{-1}R_t^{(h)}$. This outlier introduces a change in the quadratic trend of the multivariate series. The series $\mathbf{X}_t$ can be transformed into a stationary one by taking the second differences. Hence an MQS is transformed into an MLS, an MRS is transformed into an MAO, and so on. A similar procedure as that proposed for the $I(1)$ case applies. In fact, the discussion can be generalized to handle outliers in a general $I(d)$ series.

8. SIMULATIONS AND COMPUTATIONAL RESULTS

In this section we investigate the computational aspects of the proposed procedures through simulation. First, we obtain critical values for all of the test statistics; second, we compare the power of the multivariate and projection procedures for detecting outliers. We begin with the power of detecting level shifts, followed by power of identifying other types of outlier. To save space, we show only the results for the stationary case.

8.1 Critical Values

Critical values of the test statistics for outlier detection for univariate and multivariate time series are usually obtained through simulation using a numerous series from different models. For instance, the outlier detection routines in the programs TRAMO and SCA use critical values obtained by such a simulation study. We follow this stream and consider eight VARMA(p, q) models to generate the critical values. The dimensions used in the simulation are $k = 2, 3, 5$, and 10, and the parameter matrices used are given in Table 1. The constant term of the models is always the vector $\mathbf{1}_k$, and the innovations covariance matrix is the identity matrix. For cases of k = 2 and 3, the AR parameter matrices have eigenvalues of approximately

Table 1. Vector Time Series Models Used in Simulation Study

	Model 1	Model 2	Model 3
Dimension $k=2$			
$\mathbf{\Phi}$	$\begin{bmatrix} .6 & .2 \\ .2 & .4 \end{bmatrix}$		$\begin{bmatrix} .6 & .2 \\ .2 & .4 \end{bmatrix}$
$\mathbf{\Theta}$		$\begin{bmatrix} -.7 & 0 \\ -.1 & -.3 \end{bmatrix}$	$\begin{bmatrix} -.7 & 0 \\ -.1 & -.3 \end{bmatrix}$
	Model 4	Model 5	Model 6
Dimension $k=3$			
$\mathbf{\Phi}$	$\begin{bmatrix} .6 & .2 & 0 \\ .2 & .4 & 0 \\ .6 & .2 & .5 \end{bmatrix}$		$\begin{bmatrix} .6 & .2 & 0 \\ .2 & .4 & 0 \\ .6 & .2 & .5 \end{bmatrix}$
$\mathbf{\Theta}$		$\begin{bmatrix} -.7 & 0 & 0 \\ -.1 & -.3 & 0 \\ -.7 & 0 & -.5 \end{bmatrix}$	$\begin{bmatrix} -.7 & 0 & 0 \\ -.1 & -.3 & 0 \\ -.7 & 0 & -.5 \end{bmatrix}$
	Model 7		
Dimension $k=5$			
$\mathbf{\Phi}$	$\operatorname{diag}(\mathbf{\Phi})=(.9,.7,.5,.3,.1)$, and $\Phi(i,i+1)=-.5$, $i=1,\dots,4$, $\Phi(i,j)=0$, elsewhere		
	Model 8		
Dimension $k=10$			
$\mathbf{\Phi}$	$\operatorname{diag}(\mathbf{\Phi})=(.9,.8,\dots,.1,0)$, and $\Phi(i,i+1)=-.5$, $i=1,\dots,9$, $\Phi(i,j)=0$, elsewhere		

.276 and .724, and .276, .5, and .724, whereas the MA parameter matrices have eigenvalues $-.3$ and $-.7$, and $-.3, -.5$, and $-.7$. For the cases where $k=5$ and 10, the AR parameter matrices have nonzero elements in their main and first upper diagonals only, and their eigenvalues are the elements of their respective main diagonals. Using the eight models, we generated critical values of the test statistics Λ_I, Λ_A, Λ_L, and Λ_T in (13) and D_L in (10). The LRT statistic for detecting level shifts is included for comparison purposes.

The sample sizes used are $n = 50, 100, 200, 300, 400$, and 500, covering most of the cases encountered in practice. For a given model and sample size, we generated 10,000 series and computed the test statistics. Table 2 summarizes the empirical critical values of the simulation. From this table, we see only minor differences in the critical values among different models of the same dimension, and hence recommend the critical values in Table 3 for practical use. In application, if the sample size is different from those used in the simulation, then we recommend interpolating the values of Table 3. To better understand the relationship between critical values and the number of components and sample size, we fitted three linear regressions in which each column of critical values in Table 3 at the 95th percentile serving as the dependent variables, with the number of components, k, and the number of observations, n, as the regressors. The fitted regressions are $\Lambda_i = 3.24 + .1561k + .0014n$, $\Lambda_L = 2.86 + .0780k + .0008n$, and $D_L = 1.21 + .0240k + .0005n$, for $i = I, A, T$. All estimated parameters are statistically significant, and the R^2 of the regressions are .97, .94, and .82. We used these three regressions to obtain critical values for various samples sizes and numbers of components and found that a small increase in the sample size produced a small increase in the critical values. The three regressions work well in general.

Finally, we also obtained critical values for the multivariate test statistics in (5) through the same simulation method for sample sizes $n = 100$ and 200. Table 4 shows the recommended critical values for practical use. These critical values are used in power comparisons of the next section.

8.2 Power and Robustness Comparison of the Statistics for Detecting Level Shifts

Next, we compare the performance of the multivariate LRT statistics, the test statistics based on projections, and the cusum test for detecting level shifts. We use sample sizes $n = 100$ and 200 and three different outlier sizes, $\mathbf{w}_L = 3 \times \mathbf{1}_k$, $4 \times \mathbf{1}_k$, and a random $\mathbf{w}_L$. The direction of the random $\mathbf{w}_L$ is generated by drawing a uniform [0, 1] random variable u for each component of $\mathbf{w}_L$ and defining $w_{L,i} = 0$ or 3 if u is in the interval $(0, 1/2)$ or $(1/2, 1)$. If $w_{L,i} = 0$ for all i, then we discard the outlier.

For a given sample size and level shift, we generated 1,000 series and computed the test statistic $J_{\max}$ in (5) for a level shift, the maximum projection statistic Λ_L in (13), and the maximum cusum statistic in (10) based on the proposed procedure. We compare each statistic with its critical values in Tables 3 and 4 at the 5% significance level and tabulate the number of times a level shift is detected. The results are given in the first part of Table 5 (see columns $J_{\max}$, Λ_L, and D_L). For all of the models considered, the cusum test outperforms the other two, but all three tests seem to have good power when the sample size is 200.

We also study the power of these three statistics in the presence of other outliers. Specifically, for each model, we generated 1,000 series of size $n = 100$. Each series is contaminated by an MIO at $h_I = 20$ with size $\mathbf{w}_I = w \times \mathbf{1}_k$, an MAO at $h_A = 40$ with size $\mathbf{w}_A = -w \times \mathbf{1}_k$, an MTC at $h_T = 80$ with size $\mathbf{w}_T = -w \times \mathbf{1}_k$, and an MLS at $h_L = 60$ with size $\mathbf{w}_L = w \times \mathbf{1}_k$, where $w = 3$ or 4. A random vector $\mathbf{w}$, generated by the same method as before, is also used as the size for all outliers. We compute and compare the three test statistics of level shift with their respective critical values in Tables 3 and 4 at the 5% significance level. The power of these three statistics are given in the second part of Table 5 (see columns $J_{\max}$, Λ_L, and D_L). All three tests are affected by the presence of other outliers, but similar to the case of a single level shift, the cusum test continues to outperform the other two test statistics. Furthermore, we measured the power loss of each test by

$$\operatorname{loss}(i) = 1 - \frac{\text{power with outliers in model } i}{\text{power with no outliers in model } i},$$

and obtained the mean power loss of the three test statistics for the eight models used with $w = 3$. The average losses for the multivariate statistic, the projection statistic, and the cusum test are 24.1%, 13.3%, and 5.4%. Therefore, the multivariate and projection test statistics for level shift seem to be more susceptible to masking effects than the cusum test statistic.

Finally, we study the robustness properties of these statistics in the presence of other outliers. With this objective, we obtain the empirical type I error and compare it with the nominal level of the test. We use a generating procedure similar to that of the power study to conduct the simulation. However, for each generated series, the outliers consist of an MIO at $h_I = 25$ with size

Table 2. Empirical Critical Values of the Test Statistics Considered

			95th percentiles					99th percentiles				
n	k	M	Λ_I	Λ_A	Λ_L	Λ_T	D_L	Λ_I	Λ_A	Λ_L	Λ_T	D_L
50	2	1	3.53	3.48	2.91	3.47	1.26	3.81	3.75	3.24	3.76	1.32
		2	3.47	3.47	2.87	3.34	1.26	3.90	3.76	3.17	3.70	1.33
		3	3.57	3.41	2.97	3.44	1.25	3.92	3.84	3.26	3.72	1.32
	3	4	3.90	3.77	3.08	3.70	1.28	4.10	3.94	3.41	3.91	1.35
		5	3.85	3.71	2.94	3.65	1.29	4.13	3.94	3.49	3.94	1.36
		6	3.89	3.66	3.07	3.70	1.28	4.11	3.91	3.45	3.93	1.35
	5	7	4.20	3.92	3.34	3.77	1.31	4.45	4.24	3.69	4.02	1.36
	10	8	4.92	4.67	3.75	4.30	1.34	5.13	4.89	4.21	4.56	1.39
100	2	1	3.75	3.71	3.08	3.64	1.34	4.08	4.03	3.40	4.06	1.44
		2	3.75	3.61	3.06	3.59	1.33	4.15	3.98	3.45	3.98	1.43
		3	3.76	3.62	3.08	3.63	1.34	4.12	4.02	3.50	4.06	1.43
	3	4	4.03	3.80	3.19	3.78	1.37	4.39	4.10	3.50	4.15	1.45
		5	4.08	3.91	3.16	3.77	1.36	4.45	4.09	3.49	4.14	1.45
		6	4.00	3.78	3.16	3.75	1.36	4.36	4.06	3.56	4.10	1.45
	5	7	4.49	4.14	3.47	4.13	1.44	4.77	4.51	3.84	4.49	1.51
	10	8	5.30	4.94	3.82	4.79	1.49	5.59	5.28	4.25	5.10	1.53
200	2	1	4.01	3.92	3.15	3.95	1.41	4.28	4.24	3.54	4.29	1.50
		2	3.97	3.88	3.18	3.84	1.40	4.20	4.19	3.50	4.28	1.49
		3	3.95	3.84	3.21	3.80	1.40	4.25	4.11	3.49	4.10	1.50
	3	4	4.27	4.05	3.25	4.07	1.43	4.77	4.39	3.69	4.46	1.55
		5	4.34	4.09	3.25	4.06	1.42	4.77	4.44	3.58	4.50	1.54
		6	4.32	3.98	3.29	4.04	1.42	4.69	4.34	3.71	4.37	1.55
	5	7	4.69	4.32	3.56	4.33	1.54	5.04	4.70	3.94	4.66	1.67
	10	8	5.53	5.13	3.84	5.07	1.62	5.93	5.57	4.27	5.42	1.74
300	2	1	4.05	4.03	3.25	4.00	1.43	4.32	4.31	3.65	4.30	1.56
		2	4.07	3.93	3.22	4.01	1.42	4.30	4.27	3.66	4.30	1.55
		3	4.04	3.92	3.26	3.98	1.43	4.34	4.34	3.72	4.25	1.56
	3	4	4.38	4.18	3.38	4.11	1.47	4.73	4.50	3.79	4.53	1.61
		5	4.38	4.17	3.32	4.17	1.46	4.80	4.61	3.64	4.52	1.60
		6	4.36	4.13	3.38	4.17	1.47	4.72	4.50	3.77	4.60	1.60
	5	7	4.82	4.45	3.63	4.48	1.54	5.31	4.79	3.99	4.83	1.70
	10	8	5.62	5.22	3.89	5.11	1.67	5.99	5.60	4.30	5.46	1.81
400	2	1	4.13	4.08	3.41	4.09	1.44	4.50	4.47	3.83	4.47	1.58
		2	4.17	4.05	3.42	4.05	1.45	4.62	4.37	3.77	4.50	1.57
		3	4.19	4.05	3.38	4.13	1.44	4.64	4.39	3.78	4.60	1.58
	3	4	4.36	4.24	3.45	4.26	1.47	4.89	4.61	3.93	4.70	1.62
		5	4.44	4.34	3.54	4.24	1.48	4.79	4.71	3.85	4.65	1.61
		6	4.44	4.21	3.48	4.20	1.48	4.86	4.65	3.89	4.52	1.62
	5	7	4.90	4.47	3.60	4.51	1.57	5.38	4.89	4.05	4.88	1.73
	10	8	5.69	5.24	3.92	5.15	1.68	6.03	5.68	4.32	5.53	1.82
500	2	1	4.18	4.19	3.46	4.21	1.45	4.62	4.54	3.84	4.45	1.60
		2	4.18	4.13	3.42	4.11	1.46	4.60	4.52	3.82	4.53	1.59
		3	4.18	4.07	3.46	4.14	1.45	4.48	4.48	3.82	4.50	1.59
	3	4	4.46	4.28	3.50	4.26	1.49	4.91	4.71	3.96	4.72	1.62
		5	4.51	4.32	3.53	4.32	1.48	4.98	4.73	3.91	4.75	1.63
		6	4.44	4.31	3.56	4.25	1.49	4.93	4.67	3.91	4.67	1.63
	5	7	4.95	4.53	3.65	4.57	1.59	5.40	4.92	4.10	5.01	1.78
	10	8	5.77	5.25	3.95	5.19	1.69	6.06	5.70	4.35	5.58	1.84

NOTE: These values are based on sample size n and 10,000 realizations. M denotes the models in Table 1.

$\mathbf{w}_I = w \times \mathbf{1}_k$, an MAO at $h_A = 50$ with size $\mathbf{w}_A = -w \times \mathbf{1}_k$, and an MTC at $h_T = 75$ with size $\mathbf{w}_T = w \times \mathbf{1}_k$, where $w = 3$ or 4. Again, we also used a random vector $\mathbf{w}$ generated as before for the size of all outliers. The last eight rows of Table 5 give the frequencies that the test statistic is greater than its empirical 95th percentile of Tables 3 and 4. These frequencies denote chances of a false detection of a level shift by the three statistics. Once again, the cusum statistic outperforms the other two in maintaining the size of a test and being robust to the other types of outliers. The multivariate and projection statistics seem not robust to the presence of other outliers.

8.3 Power Comparison of the Multivariate and Univariate Statistics for Other Outliers

In this section we investigate the power of the test statistics for detecting other types of outlier. The outliers considered are MAOs, MIOs, and MTCs. Again, we used the eight models in Table 1 and sample sizes $n = 100$ and 200. The outlier occurs

Table 3. Recommended Critical Values of the Test Statistics Considered for Sample Size n

		95th percentiles			99th percentiles		
n	k	$\Lambda_I, \Lambda_A, \Lambda_T$	Λ_L	D_L	$\Lambda_I, \Lambda_A, \Lambda_T$	Λ_L	D_L
50	2	3.5	2.9	1.26	3.8	3.2	1.33
	3	3.8	3.0	1.28	4.0	3.4	1.35
	5	4.0	3.3	1.31	4.2	3.7	1.36
	10	4.6	3.7	1.34	4.9	4.2	1.39
100	2	3.7	3.1	1.34	4.1	3.4	1.43
	3	3.9	3.2	1.36	4.2	3.5	1.45
	5	4.2	3.4	1.44	4.6	3.9	1.51
	10	5.0	3.8	1.49	5.3	4.2	1.53
200	2	3.9	3.2	1.40	4.2	3.6	1.50
	3	4.1	3.3	1.42	4.5	3.7	1.55
	5	4.4	3.6	1.54	4.8	3.9	1.67
	10	5.2	3.8	1.62	5.6	4.3	1.74
300	2	4.0	3.3	1.43	4.3	3.7	1.56
	3	4.2	3.4	1.47	4.6	3.8	1.60
	5	4.6	3.6	1.54	5.0	4.1	1.70
	10	5.3	3.9	1.67	5.7	4.3	1.81
400	2	4.1	3.4	1.44	4.5	3.8	1.58
	3	4.3	3.5	1.48	4.7	3.9	1.62
	5	4.6	3.6	1.57	5.1	4.2	1.73
	10	5.4	3.9	1.68	5.8	4.3	1.82
500	2	4.2	3.4	1.45	4.5	3.8	1.59
	3	4.3	3.5	1.49	4.8	3.9	1.63
	5	4.7	3.7	1.59	5.1	4.1	1.78
	10	5.5	4.0	1.69	5.9	4.4	1.84

at $t = n/2$ and assumes three possible sizes as before. For each combination of model, sample size, and outlier, we generated 1,000 series to compute the proposed test statistics. We then compared the statistics with their empirical 95th percentiles of Tables 3 and 4, and tabulated the frequencies of detecting a significant outlier. Table 6 summarizes the power of various test statistics. From this table, it seems that projection test statistics outperform their corresponding multivariate counterparts. Overall, our limited simulation study supports the use of projections and cusum statistics in detecting outliers in a vector time series.

9. AN ILLUSTRATIVE EXAMPLE

We illustrate the performance of the proposed procedures by analyzing a real example. The data are the logarithms of the annual gross national product (GNP) of Spain, Italy, and France, denoted by S, I, and F, from 1947 to 2003. The series have 57 observations and are depicted by solid lines in Figure 2. Because the GNP are clearly nonstationary, we take the first difference of each series. We then compute the projection directions using the proposed procedure of Section 6 and apply the level shift detection algorithm to detect ramp shifts in the original series. The critical value is chosen by means of the regressions in Section 8.1 and turns out to be 1.32. The algorithm detects a ramp shift at time $h_1^L = 1975$. The value of the test statistic (10) for the time index is 1.39. To estimate the effect of the ramp shift, we first check whether the series are cointegrated using Johansen's test (Johansen 1991). We find a cointegration vector $\boldsymbol{\beta}$, and we use the AIC to select the error correction model with a cointegrating vector given by $\nabla \mathbf{Y}_t = \mathbf{D}_1 \nabla \mathbf{Y}_{t-1} - \boldsymbol{\alpha}\boldsymbol{\beta}' \mathbf{Y}_{t-1} + \mathbf{A}_t$, where the estimated parameters are

$$\widehat{\mathbf{D}}_1 = \begin{pmatrix} .299 & .095 & .510 \\ .069 & .344 & .524 \\ .100 & .221 & .728 \end{pmatrix},$$

$$\widehat{\boldsymbol{\alpha}} = \begin{pmatrix} .007 \\ -.001 \\ .003 \end{pmatrix},$$

and

$$\widehat{\boldsymbol{\beta}} = \begin{pmatrix} 10.762 \\ -22.355 \\ 11.975 \end{pmatrix}.$$

Note that this model is equivalent to the VAR(2) model $\mathbf{Y}_t = \boldsymbol{\Pi}_1 \mathbf{Y}_{t-1} + \boldsymbol{\Pi}_2 \mathbf{Y}_{t-2} + \mathbf{A}_t$ with $\widehat{\boldsymbol{\Pi}}_1 = \mathbf{I} + \widehat{\mathbf{D}}_1 - \widehat{\boldsymbol{\alpha}}\widehat{\boldsymbol{\beta}}'$ and $\widehat{\boldsymbol{\Pi}}_2 = -\widehat{\mathbf{D}}_1$. Then, using this model, we remove the effect of the ramp shift by estimating the regression model $\mathbf{A}_t = (\mathbf{I} - \widehat{\boldsymbol{\Pi}}_1 B - \widehat{\boldsymbol{\Pi}}_2 B^2)\mathbf{w} R_t^{(1975)} + \mathbf{E}_t$, and obtain the series free from the effect of the ramp shift by $\mathbf{Y}_t^* = \mathbf{Y}_t - \widehat{\mathbf{w}} R_t^{(1975)}$.

Next, we consider the detection of other types of outlier. The critical value is chosen using the fitted regression of Section 8.1 and turns out to be 3.8. The first part of Table 7 summarizes the results of the detection procedure. It identifies an MLS in 1966, and its effect on the series is removed as in the case of the MRS. The procedure then detects an MAO in 1975, which is estimated and cleaned from the series. The procedure fails to detect any other outliers and is terminated. The outlier-adjusted series are shown by dashed lines in Figure 2.

After identifying the outliers for the series, we estimate jointly the outlier effects and the model parameters using a first-order vector error correction model with a cointegration relationship. The estimated effects of the three detected outliers, along with the t-ratios of the estimates, are given in Table 8. The ramp shift detected by the algorithm in 1975 implies a decrease in the rate of GNP growth of about 2% in the three countries with the largest in Italy (−2.2%) and smallest in Spain

Table 4. Empirical Critical Values of the Multivariate Statistics Considered

		95th percentiles				99th percentiles			
n	k	$J_{max}(I)$	$J_{max}(A)$	$J_{max}(L)$	$J_{max}(T)$	$J_{max}(I)$	$J_{max}(A)$	$J_{max}(L)$	$J_{max}(T)$
100	2	14.0	13.5	9.5	13.6	16.1	15.7	11.3	16.3
	3	16.0	15.8	11.7	15.4	18.3	18.1	13.6	18.4
	5	18.9	19.2	14.4	18.7	21.5	22.3	17.3	21.4
	10	25.4	25.8	19.2	26.7	27.6	28.7	21.8	26.7
200	2	15.5	15.9	10.8	15.7	18.5	19.4	13.4	18.9
	3	17.8	18.3	12.9	18.0	20.7	20.8	15.9	20.3
	5	21.8	22.3	16.6	21.7	25.4	25.6	20.0	24.2
	10	29.8	30.3	21.7	29.6	33.3	34.5	25.7	33.1

Table 5. Empirical Power of Multivariate, Projection, and Cusum Test Statistics for Detecting a Level Shift

							$\mathbf{w} = 3 \times \mathbf{1}_k$			$\mathbf{w} = 4 \times \mathbf{1}_k$			$\mathbf{w}$ = random		
n	k	M	h_I	h_A	h_L	h_T	J_{max}	Λ_L	D_L	J_{max}	Λ_L	D_L	J_{max}	Λ_L	D_L
100	2	1			50		70.0	83.0	100	96.6	98.2	100	53.0	82.9	99.2
	2	2			50		58.3	82.6	100	89.2	96.9	100	52.6	81.0	99.3
	2	3			50		46.5	73.3	87.6	91.5	93.2	100	41.4	70.0	86.3
	3	4			50		93.6	92.7	99.6	100	98.8	100	60.4	71.0	98.4
	3	5			50		68.1	98.8	100	94.7	99.7	100	27.5	95.3	100
	3	6			50		86.3	86.6	88.7	98.9	99.0	99.2	28.3	83.6	88.1
	5	7			50		43.3	98.1	98.6	83.7	99.8	99.8	33.3	75.0	85.6
	10	8			50		28.6	93.8	96.4	72.7	96.9	98.3	15.0	80.2	80.9
200	2	1			100		80.7	95.4	100	98.1	99.6	100	74.3	83.9	100
	2	2			100		92.2	95.6	100	97.0	99.3	100	86.1	93.4	100
	2	3			100		78.1	90.6	99.5	97.1	98.4	100	73.1	87.2	97.2
	3	4			100		98.7	98.9	100	100	100	100	92.2	74.7	100
	3	5			100		85.8	99.8	100	99.0	100	100	83.7	96.6	100
	3	6			100		97.5	97.8	100	99.6	99.7	100	80.8	95.8	100
	5	7			100		70.9	99.8	99.9	94.7	99.9	100	73.5	75.7	97.6
	10	8			100		69.4	100	100	97.3	100	100	53.2	86.5	96.1
100	2	1	20	40	60	80	45.6	76.0	92.6	73.0	91.8	99.3	45.0	77.8	92.9
	2	2	20	40	60	80	52.0	87.6	100	74.3	96.7	100	63.7	87.4	100
	2	3	20	40	60	80	17.0	53.6	66.6	40.3	84.3	86.6	35.6	70.4	100
	3	4	20	40	60	80	73.6	74.0	90.3	92.0	85.0	100	47.3	64.5	90.6
	3	5	20	40	60	80	61.3	81.6	100	71.3	92.0	100	78.0	91.5	100
	3	6	20	40	60	80	63.6	55.3	74.3	87.0	80.0	95.0	25.6	79.4	99.8
	5	7	20	40	60	80	14.3	86.0	87.4	17.3	90.4	93.9	25.4	59.3	84.7
	10	8	20	40	60	80	3.2	58.4	74.6	2.7	61.9	79.8	6.0	58.1	71.6
100	2	1	25	50		75	4.0	14.0	3.3	2.6	26.3	3.3	3.5	15.0	4.5
	2	2	25	50		75	.3	9.0	3.0	1.3	11.0	2.0	.7	8.0	1.7
	2	3	25	50		75	4.0	21.3	4.3	2.3	31.3	4.6	.9	10.5	2.7
	3	4	25	50		75	6.3	15.6	3.0	7.0	28.0	6.3	4.3	19.3	4.1
	3	5	25	50		75	3.3	8.0	4.0	2.3	12.6	4.3	.3	9.0	2.7
	3	6	25	50		75	16.3	22.3	4.3	16.6	27.3	4.6	1.9	12.3	3.9
	5	7	25	50		75	3.4	36.1	4.5	4.4	49.5	5.5	5.8	31.7	4.8
	10	8	25	50		75	4.6	44.8	4.3	3.9	60.4	4.6	4.5	38.9	3.3

NOTE: n is the sample size, M denotes the model in Table 1, h_i denotes time point at which a type i outlier occurs, and $\mathbf{w}$ is the outlier.

Table 6. Empirical Power of Multivariate and Projection Test Statistics for Detecting an Outlier in a Vector Time Series

		$n = 100, h = 50$						$n = 200, h = 100$					
		$\mathbf{w} = 3 \times \mathbf{1}_k$		$\mathbf{w} = 4 \times \mathbf{1}_k$		$\mathbf{w}$ = random		$\mathbf{w} = 3 \times \mathbf{1}_k$		$\mathbf{w} = 4 \times \mathbf{1}_k$		$\mathbf{w}$ = random	
k	M	J_{max}	Λ_I	J_{max}	Λ_I	J_{max}	Λ_I	J_{max}	Λ_I	J_{max}	Λ_I	J_{max}	Λ_I
MIO													
2	1	59.9	77.8	95.0	98.0	44.2	65.9	58.8	71.8	92.6	96.8	42.6	56.1
2	2	53.9	71.5	89.8	95.3	36.4	58.1	58.5	68.4	92.5	95.5	38.7	52.8
2	3	51.1	68.6	88.0	95.2	40.5	61.2	57.0	68.3	92.7	95.6	42.4	54.0
3	4	81.9	91.5	99.6	99.2	46.3	74.5	81.6	87.5	100	100	48.3	66.7
3	5	62.0	76.1	95.1	97.4	33.2	70.0	67.4	76.2	97.7	98.5	34.2	64.0
3	6	61.8	78.1	92.3	97.6	38.5	70.2	67.1	75.8	98.2	98.8	42.8	63.8
5	7	95.4	99.7	100	100	58.7	91.4	96.7	99.3	100	100	57.5	82.5
10	8	99.6	100	100	100	69.7	99.7	98.1	100	100	100	80.4	98.4
MAO													
2	1	86.6	93.6	99.3	99.3	58.3	86.3	87.3	93.6	99.3	100	55.2	74.3
2	2	67.0	96.0	96.0	100	38.9	87.6	67.0	91.0	95.6	99.3	35.6	75.7
2	3	91.0	99.0	99.6	100	44.4	87.7	98.0	98.0	100	100	36.2	69.1
3	4	98.3	99.3	99.6	100	68.6	84.9	99.3	99.6	100	100	66.1	69.3
3	5	78.6	95.6	98.6	100	37.5	86.3	82.6	89.6	99.0	99.0	37.7	75.3
3	6	97.0	99.0	99.6	100	39.7	91.2	98.8	97.8	100	100	46.3	80.8
5	7	91.1	99.4	100	100	74.9	93.5	97.5	97.5	100	100	72.6	84.8
10	8	99.2	100	100	100	81.6	97.1	100	100	100	100	90.1	95.3
MTC													
2	1	61.3	88.6	93.3	98.6	46.1	82.6	61.0	88.6	98.0	98.0	41.8	78.3
2	2	64.5	97.0	94.0	100	58.8	93.7	66.0	92.0	94.6	99.3	62.1	87.1
2	3	71.3	93.6	92.6	99.3	57.7	86.7	73.6	90.3	97.7	98.6	60.8	81.8
3	4	90.0	98.0	100	100	59.4	84.9	92.0	92.8	99.5	99.5	54.9	68.1
3	5	71.0	97.6	95.3	99.6	67.0	91.4	75.3	93.3	99.3	100	61.7	77.3
3	6	82.6	95.6	97.0	98.6	52.4	88.2	95.9	95.7	100	99.2	53.9	81.5
5	7	97.2	99.6	100	100	69.4	88.4	99.2	100	100	100	71.1	80.2
10	8	97.2	100	100	100	76.3	97.4	99.4	100	100	100	81.3	94.7

NOTE: n is the sample size, M denotes the model in Table 1, h is the time index of outlier, and $\mathbf{w}$ is the size of the outlier.

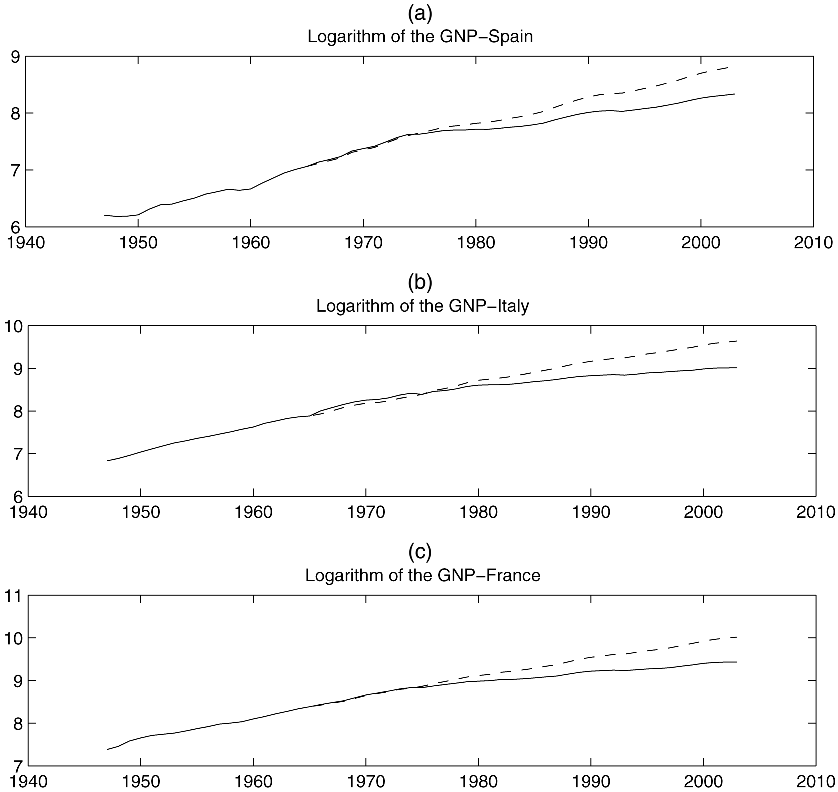

Figure 2. Original (——) and Modified (- - - -) Logarithms of the GNP of (a) Spain, (b) Italy, and (c) France.

(-1.67%). The detected shift can be associated with the first oil crisis. In fact, the ramp shift is visible from the plot of the series. The proposed algorithm also identifies an MAO in 1975 indicating that the first year of the oil crisis also had a transitory effect that was most important in Italy (-6.7%). As mentioned in Section 6, the proposed procedure allows for multiple outlier detections at a time point. In this case, one MRS affecting all of the components and one MAO affecting primarily the GNP of Italy are found. The final fitted vector error correction model is

$$\nabla \mathbf{Y}_t = \begin{pmatrix} .2856 & .1839 & .3461 \\ .0341 & .6710 & .2721 \\ -.0912 & .3765 & .5778 \end{pmatrix} \nabla \mathbf{Y}_{t-1} - \begin{pmatrix} .007 \\ -.000 \\ .002 \end{pmatrix} (\,14.412 \quad -21.651 \quad 8.132\,)\,\mathbf{Y}_{t-1} + \mathbf{A}_t.$$

There are marked changes in the parameter estimates of the model with and without outlier detection. For instance, substantial changes in the diagonal elements of the $\mathbf{D}_1$ matrix are observed before and after the outlier detection for the Italian and French GNP. The estimates of the cointegration vector also change. The estimated long-run equilibrium relationship between the variables before outlier detection is roughly $(.5S + .5F) - I$. After outlier modeling, the cointegration vector becomes roughly $(.64S + .36F) - I$, which gives heavier weight to the Spanish GNP.

Finally, we compare the results with those obtained by applying the procedure of Tsay et al. (2000). We obtained critical values for the multivariate test statistics in a simulation not reported here using the models in Table 1. The 5% critical values for the multivariate statistics are 17.3 for MIO, MAO, and MTC and 14.8 for MLS. The critical values for the component statistics are 3.9 for MIO, MAO, and MTC and 3.6 for MLS. The

Table 7. Outliers Found by the Proposed Algorithm and the Tsay, Peña, and Pankratz Procedure

Proposed procedure						
Iterations	(Λ_I, h_I)	(Λ_A, h_A)	(Λ_L, h_L)	(Λ_T, h_T)	*Time*	*Type*
1	(4.11, 1966)	(4.05, 1965)	(4.78, 1966)	(4.22, 1966)	1966	MLS
2	(3.37, 1976)	(4.77, 1975)	(4.15, 1975)	(4.45, 1975)	1975	MAO
3	(3.14, 1960)	(3.74, 1960)	(3.49, 1960)	(3.68, 1960)		
Procedure of Tsay, Peña, and Pankratz						
Iterations	(J_I, h_I)	(J_A, h_A)	(J_L, h_L)	(J_T, h_T)	*Time*	*Type*
1	(15.08, 1966)	(15.54, 1965)	(11.39, 1975)	(14.11, 1966)		
Iterations	(C_I, h_I)	(C_A, h_A)	(C_L, h_L)	(C_T, h_T)	*Time*	*Type*
2	(3.78, 1966)	(3.84, 1965)	(3.16, 1975)	(3.42, 1966)		

Table 8. Estimation of the Sizes of the Outliers Detected by the Algorithm

Time	Type	S (t-ratio)	I (t-ratio)	F (t-ratio)
1966	MLS	.0165	.0473	.0152
		(1.7046)	(7.0546)	(2.0114)
1975	MRS	−.0167	−.0224	−.0196
		(−1.9723)	(−2.4668)	(−2.2817)
1975	MAO	−.0434	−.0672	−.0312
		(−1.8392)	(−4.1121)	(−1.6917)

second part of Table 7 summarizes the results using the same first-order vector error correction model. The procedure fails to detect any outliers at the 5% level, even though some of the test statistics are only slightly smaller than the critical values.

APPENDIX: PROOFS

Proof of Lemma 1

Taking into account that $y_t = x_t + r_t$, x_t, and r_t are independent, that $E[x_t] = E[x_t^3] = 0$, and recalling that $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i$,

$$E[(y_t - \bar{y})^4] = E[(x_t + \widetilde{r}_t)^4] = E[x_t^4] + 6E[x_t^2]\widetilde{r}_t^2 + \widetilde{r}_t^4.$$

As $\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{Y}}\mathbf{v} = 1$, the kurtosis coefficient of y, is given by

$$\gamma_y(\mathbf{v}) = \frac{1}{n}\sum_{t=1}^n \big(E[x_t^4] + 6E[x_t^2]\widetilde{r}_t^2 + \widetilde{r}_t^4\big).$$

Finally, as $E[x_t^2] = \mathbf{v}'\mathbf{v}$, $E[x_t^4] = 3E[x_t^2]^2 = 3(\mathbf{v}'\mathbf{v})^2$, $\frac{1}{n}\sum_{t=1}^n \widetilde{r}_t^2 = \mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{R}}\mathbf{v}$, and $\mathbf{v}'\mathbf{v} = \mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{Y}}\mathbf{v} - \mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{R}}\mathbf{v}$, we obtain $\gamma_y(\mathbf{v}) = 3(\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{Y}}\mathbf{v})^2 - 3(\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{R}}\mathbf{v})^2 + \omega_r(\mathbf{v})$.

Proof of Lemma 2

The Lagrangian for the extreme points of $\gamma_y(\mathbf{v})$ is $\pounds(\mathbf{v}) = 3 - 3(\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{R}}\mathbf{v})^2 + \omega_r(\mathbf{v}) - \lambda(\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{Y}}\mathbf{v} - 1)$, with gradient

$$\nabla\pounds(\mathbf{v}) = -12(\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{R}}\mathbf{v})\boldsymbol{\Sigma}_{\mathbf{R}}\mathbf{v} + \left(\frac{4}{n}\sum_{t=1}^n \widetilde{r}_t^2\widetilde{\mathbf{R}}_t\widetilde{\mathbf{R}}_t'\right)\mathbf{v} - 2\lambda\boldsymbol{\Sigma}_{\mathbf{Y}}\mathbf{v}.$$

Letting $\nabla\pounds(\mathbf{v})$ equal 0, multiplying by $\mathbf{v}'$ in the equality and taking into account the constraint $\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{Y}}\mathbf{v} = 1$, we have $\lambda = -6(\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{R}}\mathbf{v})^2 + 2\omega_r(\mathbf{v})$. Because $\boldsymbol{\Sigma}_{\mathbf{R}} = \frac{1}{n}\sum_{t=1}^n \widetilde{\mathbf{R}}_t\widetilde{\mathbf{R}}_t'$, we have

$$-12(\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{R}}\mathbf{v})\boldsymbol{\Sigma}_{\mathbf{R}}\mathbf{v} + 4\left(\frac{1}{n}\sum_{t=1}^n \widetilde{r}_t^2\widetilde{\mathbf{R}}_t\widetilde{\mathbf{R}}_t'\right)\mathbf{v} = \left(-12(\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{R}}\mathbf{v})^2 + \frac{4}{n}\sum_{t=1}^n \widetilde{r}_t^2\right)(\mathbf{I} + \boldsymbol{\Sigma}_{\mathbf{R}})\mathbf{v}.$$

Therefore,

$$-3(\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{R}}\mathbf{v})\boldsymbol{\Sigma}_{\mathbf{R}}\mathbf{v} + 3(\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{R}}\mathbf{v})^2\boldsymbol{\Sigma}_{\mathbf{R}}\mathbf{v} + \left(\frac{1}{n}\sum_{t=1}^n \widetilde{r}_t^2\widetilde{\mathbf{R}}_t\widetilde{\mathbf{R}}_t'\right)\mathbf{v} - \frac{1}{n}\sum_{t=1}^n \widetilde{r}_t^2\boldsymbol{\Sigma}_{\mathbf{R}}\mathbf{v} = -3(\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{R}}\mathbf{v})^2\mathbf{v} + \frac{1}{n}\sum_{t=1}^n \widetilde{r}_t^2\mathbf{v}$$

and, finally,

$$\sum_{t=1}^n \left[\widetilde{r}_t^2 - 3(\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{R}}\mathbf{v}) - \frac{\mu(\mathbf{v})}{n}\right]\widetilde{\mathbf{R}}_t\widetilde{\mathbf{R}}_t'\mathbf{v} = n(\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{R}}\mathbf{v})^2(\gamma_r(\mathbf{v}) - 3)\mathbf{v}.$$

Thus the extreme directions of $\pounds(\mathbf{v})$ under $\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{Y}}\mathbf{v} = 1$ are the eigenvectors of the matrix $\sum_{t=1}^n \beta_t(\mathbf{v})\widetilde{\mathbf{R}}_t\widetilde{\mathbf{R}}_t'$ with eigenvalues $\mu(\mathbf{v}) = n(\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{R}}\mathbf{v})^2(\gamma_r(\mathbf{v}) - 3)$, where $\beta_t(\mathbf{v}) = [(\mathbf{v}'\widetilde{\mathbf{R}}_t)^2 - 3(\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{R}}\mathbf{v}) - \mu(\mathbf{v})/n]$. From (7), we get that $\gamma_y(\mathbf{v}) = 3 - \sigma_r^4(3 - \gamma_r(\mathbf{v})) = 3 + \mu(\mathbf{v})/n$. Therefore, the maximum or the minimum of $\gamma_y(\mathbf{v})$ will be given when $\mu(\mathbf{v})$ is as large or as small as possible, and the maximum and the minimum of the kurtosis will be given by the maximum and the minimum of the eigenvalues of the matrix $\sum_{t=1}^n \beta_t(\mathbf{v})\widetilde{\mathbf{R}}_t\widetilde{\mathbf{R}}_t'$.

Proof of Theorem 1

We use the equalities $\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{R}}\mathbf{v} = \frac{1}{n}\sum_{t=1}^n \widetilde{r}_t^2$ and $(\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{R}}\mathbf{v})^2\gamma_r(\mathbf{v}) = \frac{1}{n}\sum_{t=1}^n \widetilde{r}_t^4$.

a. For an MAO, $r_h = \mathbf{v}'\mathbf{w}$, $r_t = 0$, $\forall t \neq h$, and $\bar{r} = r_h/n$. First, $n(\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{R}}\mathbf{v})^2\gamma_r(\mathbf{v}) = c_1 r_h^4$ and $\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{R}}\mathbf{v} = c_2 r_h^2$, where

$$c_1 = \left(1 - \frac{1}{n}\right)\left[\left(1 - \frac{1}{n}\right)^3 + \frac{1}{n^3}\right], \qquad c_2 = \frac{1}{n}\left(1 - \frac{1}{n}\right),$$

and the eigenvalues are given by $\mu(\mathbf{v}) = c_0 r_h^4$, where $c_0 = c_1 - 3nc_2^2$. In contrast, after some algebra, it can be shown that

$$\left[\sum_{t=1}^n \beta_t(\mathbf{v})\widetilde{\mathbf{R}}_t\widetilde{\mathbf{R}}_t'\right]\mathbf{v} = [m_1 r_h^3 + m_2 r_h^5]\mathbf{R}_h, \tag{A.1}$$

where

$$m_1 = \left(1 - \frac{1}{n}\right)\left[\frac{1}{n^3} + \left(1 - \frac{1}{n}\right)^3 - 3c_2\right], \qquad m_2 = -c_0\frac{1}{n}\left(1 - \frac{1}{n}\right).$$

Because $\mathbf{R}_h = \mathbf{w}$,

$$\mathbf{v} = \frac{m_1 r_h^3 + m_2 r_h^5}{c_0 r_h^4}\mathbf{w},$$

and the other eigenvectors are orthogonal to $\mathbf{w}$. Moreover, because the eigenvalues are given by $c_0 r_h^4$ and $c_0 > 0$ for $n > 5$, we get that the maximum of the kurtosis coefficient is given in the direction of $\mathbf{w}$, whereas the minimum is attained in the orthogonal directions to $\mathbf{w}$.

b. For an MTC, $r_t = 0$ if $t < h$, $r_t = \delta^{t-h} r_h$ for $t \geq h$ and $\bar{r} = m r_h$, where $m = (1 - \delta^{n-h+1})/(n(1 - \delta))$. First, $n(\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{R}}\mathbf{v})^2\gamma_r(\mathbf{v}) = c_1 r_h^4$ and $\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{R}}\mathbf{v} = c_2 r_h^2$, where

$$c_1 = (h-1)m^4 + \sum_{t=h}^n (\delta^{t-h} - m)^4,$$

$$c_2 = \frac{1}{n}\left[(h-1)m^2 + \sum_{t=h}^n (\delta^{t-h} - m)^2\right],$$

and the eigenvalues are given by $\mu(\mathbf{v}) = c_0 r_h^4$, where $c_0 = c_1 - 3nc_2^2$. In contrast, after some algebra, it can be shown that the relation (A.1) is valid in this case for the values

$$m_1 = (h-1)(m^4 - 3c_2 m^2) + \sum_{t=h}^n \big[(\delta^{t-h} - m)^4 - 3c_2(\delta^{t-h} - m)^2\big],$$

$$m_2 = -\frac{c_0}{n}\left[(h-1)m^2 + \sum_{t=h}^n (\delta^{t-h} - m)^2\right],$$

and one eigenvector is proportional to $\mathbf{w}$ and the others are orthogonal to it. Because the eigenvalues are given by $c_0 r_h^4$, the kurtosis coefficient of y_t is maximized or minimized when $\mathbf{v}$ is proportional to $\mathbf{w}$ depending on the sign of c_0, which in general depends on the values of n, h, and δ.

c. For an MLS, $r_t = 0$ if $t < h$, $r_t = r_h$ for $t \geq h$ and $\bar{r} = \frac{n-h+1}{n} r_h$. First, $n(\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{R}}\mathbf{v})^2\gamma_r(\mathbf{v}) = c_1 r_h^4$ and $\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{R}}\mathbf{v} = c_2 r_h^2$, where

$$c_1 = \frac{(h-1)(n-h+1)}{n^4}[(n-h+1)^3 + (h-1)^3],$$

$$c_2 = \frac{(h-1)(n-h+1)}{n^2},$$

and the eigenvalues are given by $\mu(\mathbf{v}) = c_0 r_h^4$, where $c_0 = c_1 - 3nc_2^2$. In contrast, after some algebra, it can be shown that the relation (A.1) is valid in this case for the values

$$m_1 = \frac{(h-1)(n-h+1)}{n^4}\left[(n-h+1)^3 + (h-1)^3 - 3c_2\right]$$

and

$$m_2 = -c_0 \frac{(h-1)(n-h+1)}{n^2},$$

showing that one eigenvector is proportional to $\mathbf{w}$ and the others are orthogonal to it. The eigenvalues are given by $c_0 r_h^4$, and it is not difficult to see that $c_0 < 0$ if and only if

$$h \in \left(1 + \frac{1}{2}\left(1 - \frac{1}{\sqrt{3}}\right)n, 1 + \frac{1}{2}\left(1 + \frac{1}{\sqrt{3}}\right)n\right)$$

and $c_0 > 0$ otherwise. Thus the maximum of the kurtosis coefficient is given in the direction of $\mathbf{w}$ if $c_0 > 0$, and the minimum of the kurtosis coefficient is given in the direction of w if $c_0 < 0$.

Proof of Corollary 1

The result follows immediately from Theorem 1, because the relation $\mathbf{A}_t = \mathbf{E}_t + \mathbf{w}I_t^{(h)}$ coincides with the MAO case in a white noise series.

Proof of Theorem 2

First, the results of Lemmas 1 and 2 continues to hold in both cases. Let $\beta_1 = \mathbf{v}'\mathbf{w}_1$ and $\beta_2 = \mathbf{v}'\mathbf{w}_2$.

a. Let $\mathbf{R}_t = \mathbf{w}_1 I_t^{(h_1)} + \mathbf{w}_2 I_t^{(h_2)}$. Following the steps of the proof of Theorem 1, and after some algebra, it can be shown that the eigenvectors of the matrix $\sum_{t=1}^n \beta_t(\mathbf{v})\widetilde{\mathbf{R}}_t\widetilde{\mathbf{R}}_t'$, ignoring terms $O(n^{-1})$, are given by

$$\mathbf{v} = \frac{\beta_1^3}{(\beta_1^4 + \beta_2^4)}\mathbf{w}_1 + \frac{\beta_2^3}{(\beta_1^4 + \beta_2^4)}\mathbf{w}_2, \tag{A.2}$$

and their orthogonal, with eigenvalues $\mu(\mathbf{v}) = \beta_1^4 + \beta_2^4$.

(1) If $\mathbf{w}_1$ and $\mathbf{w}_2$ are proportional, then $\mathbf{w}_2 = \varrho\mathbf{w}_1$, $\mathbf{v} = \mathbf{w}_1/\|\mathbf{w}_1\|$, and $\mu(\mathbf{v}) = (1+\varrho^4)\|\mathbf{w}_1\|^4$. Thus the kurtosis coefficient is maximized when $\mathbf{v}$ is proportional to $\mathbf{w}_1$ and is minimized when $\mathbf{v}$ is orthogonal to $\mathbf{w}_1$.

(2) If $\mathbf{w}_1$ and $\mathbf{w}_2$ are orthogonal, then $\mathbf{v} = \mathbf{w}_1/\|\mathbf{w}_1\|$ with $\mu(\mathbf{v}) = \|\mathbf{w}_1\|^4$ and $\mathbf{v} = \mathbf{w}_2/\|\mathbf{w}_2\|$ with $\mu(\mathbf{v}) = \|\mathbf{w}_2\|^4$ are extremes of the kurtosis coefficient. Multiplying (A.2) by $\mathbf{w}_1$ and $\mathbf{w}_2$, we obtain

$$\mathbf{v} = \pm\frac{\|\mathbf{w}_2\|^2}{\|\mathbf{w}_1\|(\|\mathbf{w}_1\|^4 + \|\mathbf{w}_2\|^4)^{3/2}}\mathbf{w}_1 \pm \frac{\|\mathbf{w}_1\|^2}{\|\mathbf{w}_2\|(\|\mathbf{w}_1\|^4 + \|\mathbf{w}_2\|^4)^{3/2}}\mathbf{w}_2, \tag{A.3}$$

with $\mu(\mathbf{v}) = \|\mathbf{w}_1\|^4\|\mathbf{w}_2\|^4/(\|\mathbf{w}_1\|^4 + \|\mathbf{w}_2\|^4)$. Thus the vector with larger norm between $\mathbf{w}_1$ and $\mathbf{w}_2$ gives the global maximum, the other one gives a local maximum, and the directions in (A.3) give the global minimum.

(3) From (A.2), we get

$$\beta_1^4 + \beta_2^4 = \beta_1^2\|\mathbf{w}_1\|^2 + \frac{\beta_2^3}{\beta_1}\mathbf{w}_1'\mathbf{w}_2 = \frac{\beta_1^3}{\beta_2}\mathbf{w}_1'\mathbf{w}_2 + \beta_2^2\|\mathbf{w}_2\|^2.$$

Let $\psi = \beta_1/\beta_2$, which verifies $\psi^4 - a_1\psi^3 + a_2\psi - 1 = 0$, where $a_1 = \|\mathbf{w}_1\|/\|\mathbf{w}_2\|\cos\varphi$ and $a_2 = \|\mathbf{w}_2\|/\|\mathbf{w}_1\|\cos\varphi$. Note that a_1 and a_2 are simultaneously positive or negative. Assume first that $a_1 > 0$. The polynomial $p(\psi) = \psi^4 - a_1\psi^3 + a_2\psi - 1$ tends to ∞ when $\psi \to \infty$. To show that the largest root of the polynomial tends to ∞ when $\psi \to \infty$, it is sufficient to show the following:

(c.1) If $r = a_1 - 1$, then $p(r)$ is negative and $p(\lambda)$ has a root larger than r.

(c.2) $p'(\psi)$ is positive in the interval (r, ∞), and thus $p(\psi)$ is strictly positive in this interval, so the root is the largest one.

To show (c.1), $p(r) = -a_1^3 + 3a_1^2 - 3a_1 + a_2a_1 - a_2 < 0$ if a_1 is large enough. To show (c.2), $p'(r) = a_1^3 - 6a_1^2 + 9a_1 - 4 + a_2 > 0$ if a_1 is large enough. This proves that the largest root of $p(\psi)$ tends to ∞ when $a_1 \to \infty$. Thus

$$\mathbf{v} = \frac{1}{\beta_1}\left[\frac{\lambda^4}{1+\lambda^4}\mathbf{w}_1 + \frac{1}{1+\lambda^4}\mathbf{w}_2\right] \to \mathbf{w}_1,$$

when a_1 is sufficiently large, giving the maximum of the kurtosis coefficient $\mu(\mathbf{v}) = \beta_1^4(1+1/\lambda^4) \to \|\mathbf{w}_1\|^4$ that increase with a_1. If $a_1 < 0$, then take $r = a_1 + 1$, and the preceding reasoning shows that the smallest root of the polynomial tends to $-\infty$ when $a_1 \to -\infty$, and the same conclusions hold.

b. Let $\mathbf{R}_t = \mathbf{w}_1 I_t^{(h_1)} + \mathbf{w}_2 S_t^{(h_2)}$. Following the steps of the proof of Theorem 1, and after some algebra, it can be shown that, ignoring terms $O(n^{-1})$,

$$\frac{1}{n}\left[\sum_{t=1}^n \beta_t(\mathbf{v})B_t\right]\mathbf{v} = m_0\mathbf{w}_2,$$

for some constant m_0 with corresponding eigenvalues $\mu(\mathbf{v})/n = (\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{R}}\mathbf{v})^2\gamma_r(\mathbf{v}) - 3(\mathbf{v}'\boldsymbol{\Sigma}_{\mathbf{R}}\mathbf{v})^2 \simeq c_0\beta_2^4$ for some constant c_0, which proves the stated result.

[Received August 2004. Revised August 2005.]

REFERENCES

Bai, J. (1994), "Least Squares Estimation of a Shift in Linear Processes," *Journal of Time Series Analysis*, 15, 453–472.

Balke, N. S. (1993), "Detecting Level Shifts in Time Series," *Journal of Business & Economic Statistics*, 11, 81–92.

Bianco, A. M., Garcia Ben, M., Martínez, E. J., and Yohai, V. J. (2001), "Outlier Detection in Regression Models With ARIMA Errors Using Robust Estimates," *Journal of Forecasting*, 20, 565–579.

Carnero, M. A., Peña, D., and Ruiz, E. (2003), "Detecting Level Shifts in the Presence of Conditional Heteroscedasticity," technical report, Universidad Carlos III de Madrid.

Chang, I., and Tiao, G. C. (1983), "Estimation of Time Series Parameters in the Presence of Outliers," Technical Report 8, University of Chicago, Statistics Research Center.

Chang, I., Tiao, G. C., and Chen, C. (1988), "Estimation of Time Series Parameters in the Presence of Outliers," *Technometrics*, 3, 193–204.

Chen, C., and Liu, L. (1993), "Joint Estimation of Model Parameters and Outlier Effects in Time Series," *Journal of the American Statistical Association*, 88, 284–297.

Engle, R., and Granger, C. W. J. (1987), "Co-Integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55, 251–276.

Franses, P. H., and Lucas, A. (1998), "Outlier Detection in Cointegration Analysis," *Journal of Business & Economic Statistics*, 16, 459–468.

Fox, A. J. (1972), "Outliers in Time Series," *Journal of the Royal Statistical Society*, Ser. B, 34, 350–363.

Huber, P. (1985), "Projection Pursuit" (with discussion), *The Annals of Statistics*, 13, 435–525.

Inclán, C., and Tiao, G. C. (1994), "Use of Cumulative Sums of Squares for Retrospective Detection of Changes of Variance," *Journal of the American Statistical Association*, 89, 913–923.

Johansen, S. (1991), "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, 59, 1551–1580.

Jones, M. C., and Sibson, R. (1987), "What Is Projection Pursuit?" (with discussion), *Journal of the Royal Statistical Society*, Ser. A, 150, 1–36.

Justel, A., Peña, D., and Tsay, R. S. (2000), "Detection of Outlier Patches in Autoregressive Time Series," *Statistica Sinica*, 11, 651–673.

Le, N. D., Martin, R. D., and Raftery, A. E. (1996), "Modeling Flat Stretches, Bursts, and Outliers in Time Series Using Mixture Transition Distribution Models," *Journal of the American Statistical Association*, 91, 1504–1515.

Luceño, A. (1998), "Detecting Possibly Non-Consecutive Outliers in Industrial Time Series," *Journal of the Royal Statistical Society*, Ser. B, 60, 295–310.

Lütkepohl, H. (1993), *Introduction to Multiple Time Series Analysis* (2nd ed.), New York: Springer-Verlag.

Lütkepohl, H., Saikkonen, P., and Trenkler, C. (2004), "Testing for the Cointegration Rank of a VAR Process With Level Shift at Unknown Time," *Econometrica*, 72, 647–662.

Maravall, A., and Mathis, A. (1994), "Encompassing Univariate Models in Multivariate Time Series," *Journal of Econometrics*, 61, 197–233.

McCulloch, R. E., and Tsay, R. S. (1993), "Bayesian Inference and Prediction for Mean and Variance Shifts in Autoregressive Time Series," *Journal of the American Statistical Association*, 88, 968–978.

——— (1994), "Bayesian Analysis of Autoregressive Time Series via the Gibbs Sampler," *Journal of Time Series Analysis*, 15, 235–250.

Pan, J., and Fang, K. (2002), *Growth Curve Models and Statistical Diagnostics*, New York: Springer-Verlag.

Pan, J., Fung, W., and Fang, K. (2000), "Multiple Outlier Detection in Multivariate Data Using Projection Pursuit Techniques," *Journal of Statistical Planning and Inference*, 83, 153–167.

Peña, D., and Prieto, F. J. (2001a), "Multivariate Outlier Detection and Robust Covariance Matrix Estimation" (with discussion), *Technometrics*, 43, 286–310.

——— (2001b), "Cluster Identification Using Projections," *Journal of the American Statistical Association*, 96, 1433–1445.

Posse, C. (1995), "Tools for Two-Dimensional Exploratory Projection Pursuit," *Journal of Computational and Graphics Statistics*, 4, 83–100.

Rao, C. R. (1973), *Linear Statistical Inference and Its Applications*, New York: Wiley.

Sánchez, M. J., and Peña, D. (2003), "The Identification of Multiple Outliers in ARIMA Models," *Communications in Statistics, Part A—Theory and Methods*, 32, 1265–1287.

Tsay, R. S. (1986), "Time Series Model Specification in the Presence of Outliers," *Journal of the American Statistical Association*, 81, 132–141.

——— (1988), "Outliers, Level Shifts and Variance Changes in Time Series," *Journal of Forecasting*, 7, 1–20.

Tsay, R. S., Peña, D., and Pankratz, A. E. (2000), "Outliers in Multivariate Time Series," *Biometrika*, 87, 789–804.