Aust. N. Z. J. Stat. **42**(4), 2000, 463–477

RESIDUAL DIAGNOSTIC PLOTS FOR CHECKING FOR MODEL MIS-SPECIFICATION IN TIME SERIES REGRESSION

RICHARD FRACCARO[1,2]*, ROB J. HYNDMAN[1] AND ALAN VEEVERS[2]

Monash University and CSIRO Mathematical and Information Sciences

Summary

This paper considers residuals for time series regression. Despite much literature on visual diagnostics for uncorrelated data, there is little on the autocorrelated case. To examine various aspects of the fitted time series regression model, three residuals are considered. The fitted regression model can be checked using orthogonal residuals; the time series error model can be analysed using marginal residuals; and the white noise error component can be tested using conditional residuals. When used together, these residuals allow identification of outliers, model mis-specification and mean shifts. Due to the sensitivity of conditional residuals to model mis-specification, it is suggested that the orthogonal and marginal residuals be examined first.

Key words: autocorrelation; conditional residuals; generalized least squares; marginal residuals; mean shifts; model mis-specification; model transformation; orthogonal residuals; residual diagnostics; residual plots; time series regression.

1. Introduction

Regression models with autocorrelated errors have received much attention in recent years. Tsay (1984) presents an overview of time series regression. Puterman (1988) and Hossain (1990) discuss influence diagnostics, and Tsay (1986) and Ledolter (1988) consider the detection of outliers. Haslett & Hayes (1998) and Martin (1992) establish generalized versions of residuals and diagnostics that are commonly used when performing ordinary least squares (OLS) regression. However, little attention has been given to residual diagnostic plots for time series regression.

We consider a linear regression model with an autoregressive (AR) error,

$$Y_t = f(\boldsymbol{X}_t) + e_t\,, \qquad \Phi_p(B)e_t = z_t\,, \tag{1}$$

where $\boldsymbol{X}_t$ is a vector of explanatory variables assumed to be known, the regression model is $f(\boldsymbol{X}_t) = \boldsymbol{X}_t\boldsymbol{\beta}$ where $\boldsymbol{\beta}$ is a vector of coefficients, $\Phi_p(B) = (1 - \phi_1 B - \cdots - \phi_p B^p)$ is a polynomial of order p in the backshift operator B, and z_t is a zero mean Gaussian white noise series with variance σ^2.

Received September 1998; revised December 1999; accepted December 1999.
* Author to whom correspondence should be addressed.
[1] Dept Econometrics and Business Statistics, Monash University, Clayton VIC 3168, Australia.
[2] CSIRO Mathematical and Information Sciences, Private Bag 10, Clayton VIC 3168, Australia.
e-mail: Richard.Fraccaro@buseco.monash.edu.au
Acknowledgments. Richard Fraccaro was supported by a studentship from CSIRO Mathematical and Information Sciences. Rob Hyndman was supported by a grant from the Australian Research Council. The authors thank Richard Morton for useful discussions and an Associate Editor for helpful suggestions for improving the presentation.

Model (1) can also be written as

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{e}, \qquad \boldsymbol{e} \overset{\mathrm{d}}{=} \mathrm{N}(0, \boldsymbol{\Sigma}), \tag{2}$$

where the correlated error structure is represented in the matrix $\boldsymbol{\Sigma}$ which has (i, j)th element $\gamma(|i - j|)$, and where γ is the autocovariance function of the time series model represented by e_t. This representation of the time series regression model allows generalized least squares (GLS) to be used to estimate the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$. Judge *et al.* (1988 p.392) outline the use of GLS estimation in an iterative procedure; we use this method for obtaining parameter estimates for the examples here.

The limitation to AR models in (1) is not particularly restrictive because any ARMA model can be approximated by a high order AR model (see Brockwell & Davis, 1991 p.91). Hence, the results presented here can be extended to regression models with ARMA time series errors.

It is common practice with ordinary regression, when the errors are uncorrelated ($p = 0$), to graph the residuals against each of the explanatory variables. Patterns in residual plots indicate the fitted model is mis-specified. The pattern indicates the form of the mis-specification (e.g. a quadratic shape indicates that a quadratic term should be included in the model).

Our goal is to produce similar residual plots for models with autocorrelated errors. The resulting residual plots should allow assessment of other aspects of the fitted model, such as a check of the assumed properties of z_t. In Section 2, we examine a type of residual which, whilst being an intuitive diagnostic to use, is sometimes misleading when assessing the fitted regression model. Section 3 derives a more suitable type of residual and Section 4 demonstrates its use. A third type of residual examined in Section 5 provides further checking of other elements of the fitted model. When used together, these three types of residuals can reveal various aspects of the time series regression model.

2. Marginal residuals

The marginal expectation for model (1) is $\mathrm{E}(Y_t \mid \boldsymbol{X}) = f(\boldsymbol{X}_t)$. Departures from the best estimate of this expectation are called *marginal residuals*, $\hat{e}_t = Y_t - \hat{f}(\boldsymbol{X}_t)$. It would seem natural to produce diagnostics plots based on the marginal residuals. This approach is used when performing OLS estimation on uncorrelated data, but its use with autocorrelated data is problematic.

Asymptotically, $\mathrm{var}(\hat{\boldsymbol{e}}) = \boldsymbol{\Sigma}$ (following Fuller, 1996 p.519 Proposition 9.7.1) and so these residuals must be expected to exhibit autocorrelation which can lead to 'patterns' in a residual plot. These autocorrelation-induced patterns often interfere with other patterns that indicate mis-specification. Consequently, it is difficult to visually identify when mis-specification has occurred and what form of mis-specification is present.

Figure 1 is a graph of marginal residuals. This graph is based on the mean shift data example presented in Section 4.2. The autocorrelation in the data is evident in the residual plot and masks the existence of other patterns.

Under the hypothesis that the regression model has been correctly specified, the marginal residuals $\hat{\boldsymbol{e}}$ estimate the unobservable time series error process. Therefore it is suggested that marginal residuals be plotted in time order. Other types of residuals presented here could also be plotted against time, or against the fitted values as is the norm.

The nature of marginal residuals allows them to be treated as an observed time series; therefore current time series diagnostic methods can be used. For example, parameter changes

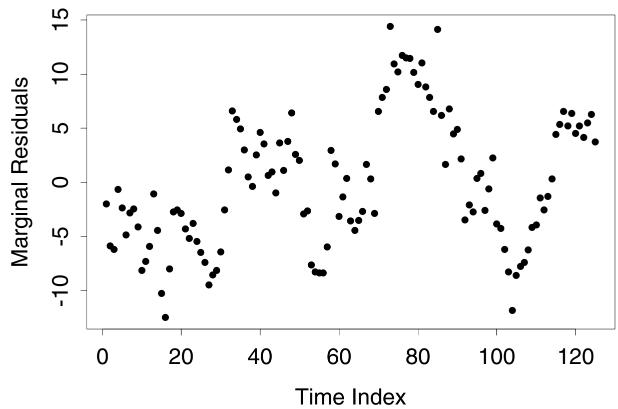

Figure 1. Marginal residuals when the error model is AR(1).
The pattern dominating the plot is due to autocorrelation.

can be detected with techniques discussed in Bagshaw & Johnson (1977), and outliers in error models can be identified by methods proposed by Ljung (1993) and Ledolter (1988).

3. Residual orthogonality

Let $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}$ denote the hat matrix from a linear model fitted using GLS regression. For ordinary least squares, $p = 0$ and $\boldsymbol{\Sigma}^{-1} = \sigma^{-2}\boldsymbol{I}$. In this case,

$$\hat{\boldsymbol{e}}^{\mathsf{T}}\boldsymbol{X} = \big((\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}\big)^{\mathsf{T}}\boldsymbol{X} = \boldsymbol{Y}^{\mathsf{T}}(\boldsymbol{I} - \boldsymbol{H})^{\mathsf{T}}\boldsymbol{X} = \boldsymbol{Y}^{\mathsf{T}}\boldsymbol{X} - \boldsymbol{Y}^{\mathsf{T}}\boldsymbol{H}^{\mathsf{T}}\boldsymbol{X} = \boldsymbol{0}.$$

Similarly, $\hat{\boldsymbol{e}}^{\mathsf{T}}\hat{\boldsymbol{Y}} = \hat{\boldsymbol{e}}^{\mathsf{T}}\boldsymbol{X}\hat{\beta} = \boldsymbol{0}$. Thus, the marginal residuals are orthogonal to $\hat{\boldsymbol{Y}}$ and to $\boldsymbol{X}$. We believe that this orthogonality is the essential reason why, for uncorrelated observations, it 'makes sense' to plot the residuals $\hat{\boldsymbol{e}}$ against $\boldsymbol{X}$ and $\hat{\boldsymbol{Y}}$.

However, for time series regression when $p > 0$, $\boldsymbol{H}^{\mathsf{T}}\boldsymbol{X} \neq \boldsymbol{X}$ and so the above orthogonality does not hold. Thus, the vector of residuals, $\hat{\boldsymbol{e}}$, is correlated with $\hat{\boldsymbol{Y}}$ and $\boldsymbol{X}$. As a result, patterns can appear in residual plots when, in fact, the residuals do not vary systematically.

A solution to finding a suitable type of residual for time series regression lies in the above orthogonality principle. The following section presents a residual orthogonal to $\hat{\boldsymbol{Y}}$ and $\boldsymbol{X}$.

3.1. The orthogonal residual

In GLS regression, the normal equations are $\boldsymbol{X}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} = \boldsymbol{X}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\hat{\beta}$. Therefore

$$\boldsymbol{X}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta}) = \boldsymbol{X}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{e}} = \boldsymbol{0}$$

and so $\hat{\boldsymbol{v}} = \boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{e}}$ is orthogonal to $\boldsymbol{X}$ (and can also be shown to be orthogonal to $\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\beta}$).

The orthogonal errors, $\boldsymbol{v} = \boldsymbol{\Sigma}^{-1}\boldsymbol{e}$, have mean $\boldsymbol{0}$ and their covariance matrix is $\boldsymbol{\Sigma}^{-1}$, which is not diagonal so they are correlated. However, the covariance has an interesting property that arises from the duality between autoregressive and moving average (MA) processes. Specifically, the inverse of the autocovariance matrix from an MA(p) process is approximately equal to the autocovariance matrix from an AR(p) process (Anderson, 1976). Murthy (1974) shows that for an AR(p) autocovariance matrix $\boldsymbol{\Sigma}$, the inverse can be represented as

$$\boldsymbol{\Sigma}^{-1} = \check{\boldsymbol{\Sigma}}^{-1} + \tilde{\boldsymbol{\Sigma}}^{-1}, \qquad (3)$$

where $\check{\boldsymbol{\Sigma}}^{-1}$ is an MA(p) autocovariance matrix, and $\tilde{\boldsymbol{\Sigma}}^{-1}$ is a matrix of zeros except for the leading and trailing $p \times p$ submatrices.

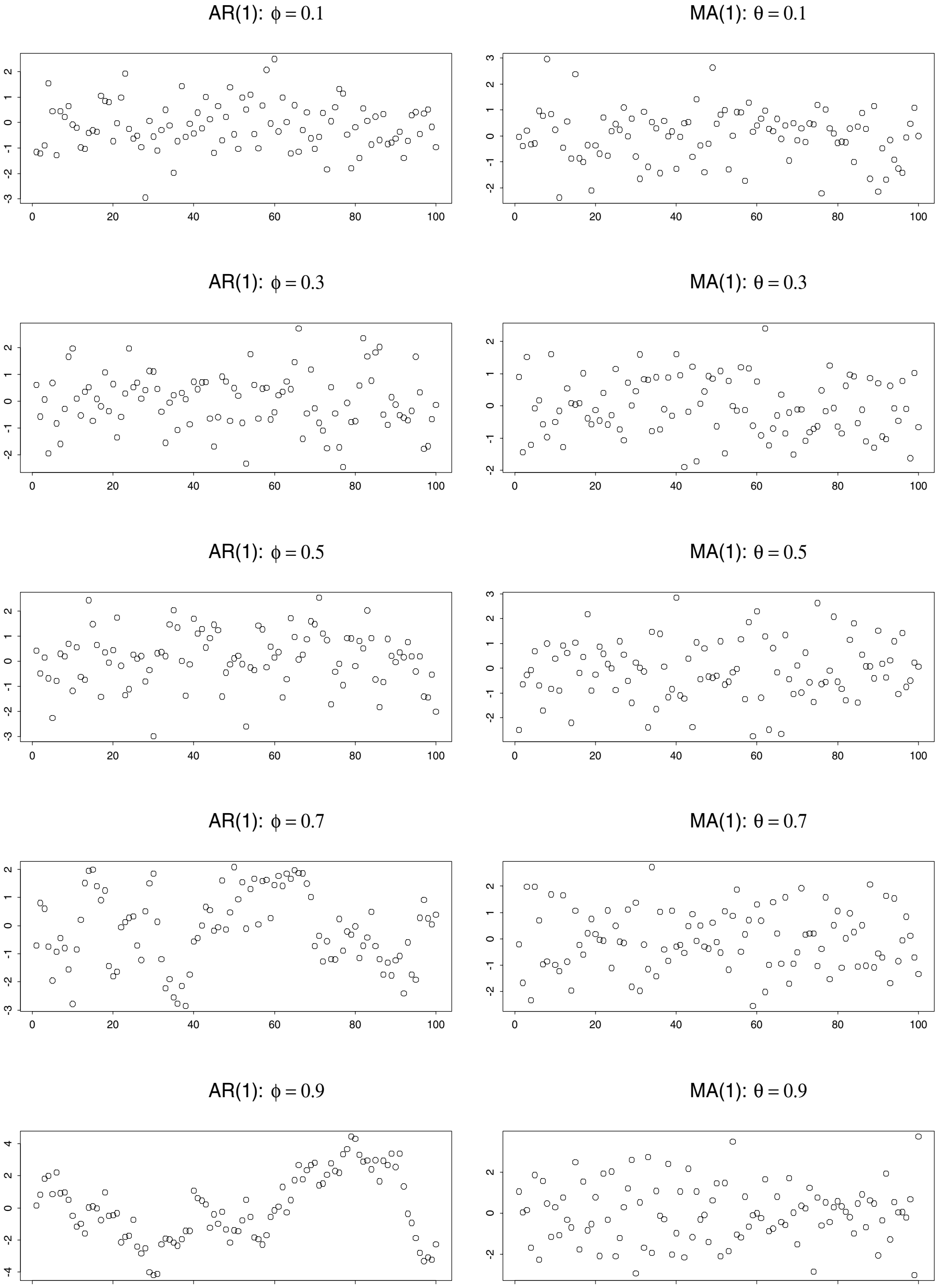

Figure 2. Left: simulated AR(1) series showing that the autocorrelation can be confused with mis-specification, especially with large ϕ. Right: simulated MA(1) series showing that the lower order autocorrelation does not lead to patterns likely to be confused with mis-specification regardless of the value of θ.

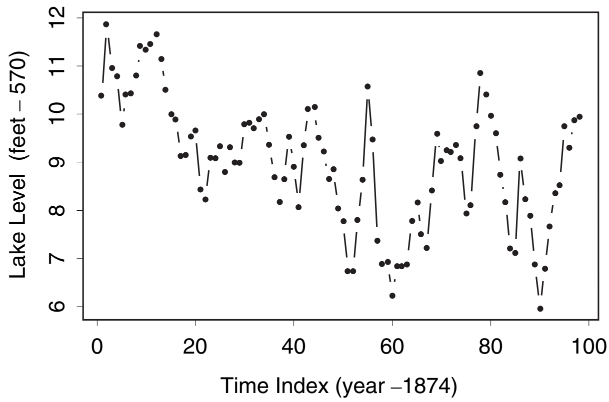

Figure 3. Level of Lake Huron in feet from 1875 to 1972

Now, for an MA(p) process, the autocovariance function satisfies $\gamma(k) = 0$ for $k > p$. Therefore, the matrix $\check{\boldsymbol{\Sigma}}^{-1}$ consists of zeros except for the main diagonal and up to p off-diagonals either side of the main diagonal. Adding the matrix $\tilde{\boldsymbol{\Sigma}}^{-1}$ only changes some of the non-zero values in the matrix $\check{\boldsymbol{\Sigma}}^{-1}$ and so $\boldsymbol{\Sigma}^{-1}$ has the same pattern as $\check{\boldsymbol{\Sigma}}^{-1}$. Therefore, the ith orthogonal error, $\boldsymbol{v}_i$, can only be correlated with those p orthogonal errors that occur immediately before and after it. For a low order AR process (and with sufficiently large n), $\boldsymbol{\Sigma}^{-1}$ is nearly diagonal, and so the orthogonal errors have low order autocorrelation.

The orthogonal residuals are defined as $\hat{\boldsymbol{v}} = \hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{e}}$. The duality property described above has an important consequence for the use of orthogonal residuals in a residual plot. As illustrated in Figure 2, low-order autocorrelation is not obvious in a scatterplot. Therefore an observer is not distracted from other patterns that may indicate mis-specification or the presence of outlying observations. Similarly for orthogonal residuals, low-order autocorrelation will not detract from the presence of other patterns or unusual residuals that exist in the plot.

The orthogonal residuals have estimated covariance matrix

$$\begin{aligned}
\operatorname{cov}(\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{e}}) &= \hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{I} - \hat{\boldsymbol{H}})\hat{\boldsymbol{\Sigma}}(\boldsymbol{I} - \hat{\boldsymbol{H}})^{\mathsf{T}}\hat{\boldsymbol{\Sigma}}^{-1} \\
&= \hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{I} - \hat{\boldsymbol{H}})\hat{\boldsymbol{\Sigma}}\big(\boldsymbol{I} - \hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X}(\boldsymbol{X}^{\mathsf{T}}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})\boldsymbol{X}^{\mathsf{T}}\big)\hat{\boldsymbol{\Sigma}}^{-1} \\
&= \hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{I} - \hat{\boldsymbol{H}})\big(\hat{\boldsymbol{\Sigma}} - \boldsymbol{X}(\boldsymbol{X}^{\mathsf{T}}\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X}^{\mathsf{T}}\big)\boldsymbol{X}^{\mathsf{T}})\hat{\boldsymbol{\Sigma}}^{-1} \\
&= \hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{I} - \hat{\boldsymbol{H}})(\boldsymbol{I} - \hat{\boldsymbol{H}}) \\
&= \hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{I} - \hat{\boldsymbol{H}}).
\end{aligned}$$

Note that the estimated hat matrix, $\hat{\boldsymbol{H}}$, is idempotent. The estimated standard deviation of $\hat{v}_i$ is therefore

$$\hat{\sigma}\sqrt{\big(\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{I} - \hat{\boldsymbol{H}})\big)_{ii}}\,. \qquad (4)$$

4. Examples using orthogonal residuals

4.1. Lake Huron data

Figure 3 shows the level of Lake Huron in feet, reduced by 570, as recorded over the years from 1875 to 1972. The data are listed in Brockwell & Davis (1991 p.555).

TABLE 1

Summary of parameter estimation for the Lake Huron data

Parameter	Estimate	Standard error
Intercept	10.099	0.4601
Time Index coefficient	–0.022	0.0080
AR(2) model		
Lag 1 coefficient	0.977	
Lag 2 coefficient	–0.278	
σ	0.705	

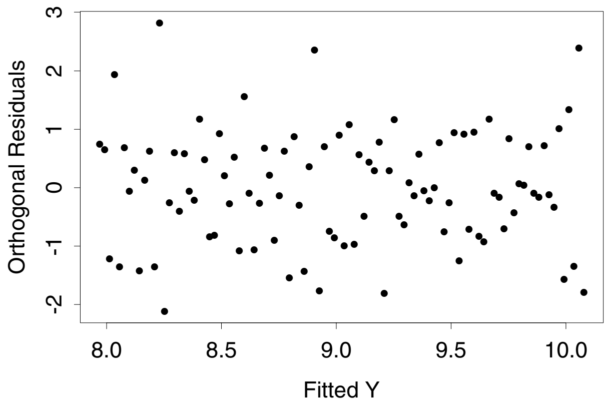

Figure 4. Studentized orthogonal residual plot for the Lake Huron data

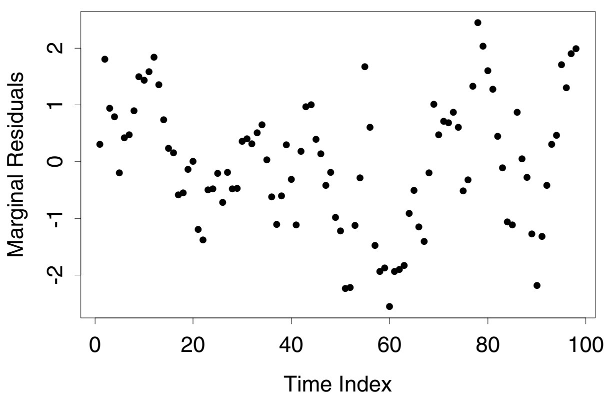

Figure 5. Marginal residual plot for the Lake Huron data

A time series regression model was fitted to the data and the result is summarized in Table 1. A linear relationship is shown to exist between lake level and time, with the errors following an AR(2) process. There is a slight downward trend in the lake level during the time in which observations were made.

In Figure 4 the studentized orthogonal residuals (obtained by dividing $\hat{\upsilon}_i$ by (4)) are plotted against the fitted observations, $\hat{\boldsymbol{Y}}$. The marginal residuals, plotted in time order, are shown in Figure 5.

The marginal residual plot reveals systematic variation in the residuals which could be mistaken as an indication that the fitted model is inadequate. Instead, the pattern in this plot is

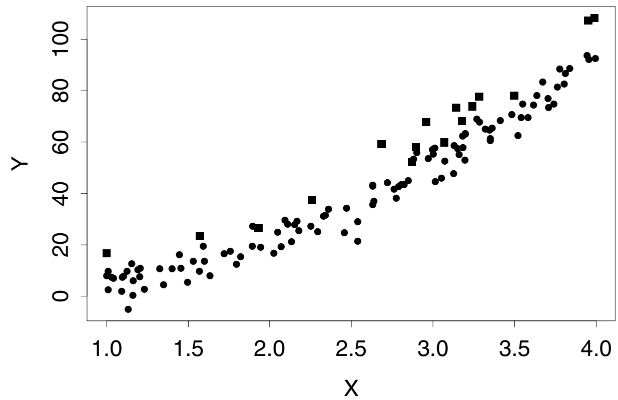

Figure 6. Graph of y_i vs x_i for the mean shift simulation data. The mean-shifted observations are indicated with squares.

a result of the autocorrelation in the residuals. In contrast, the plot of studentized orthogonal residuals does not indicate any such systematic variation. Apart from a few possible outliers, the orthogonal residuals indicate that the fitted time series regression model is satisfactory in explaining the level of Lake Huron over time.

4.2. Mean shifts

A mean shift is a type of effect that one would like to be able to detect when it occurs within a time series. In a mean shift, the mean of a time series process changes by a fixed quantity for several consecutive observations. The following example illustrates how orthogonal residuals and marginal residuals can be used together to identify a mean shift.

A dataset of 125 observations was simulated using the formula

$$y_i = 2 - 5x_i + 7x_i^2 + e_i\,,$$

where e_i was an AR(1) process with coefficient $\phi = 0.85$ and with $\sigma^2 = 2.5^2$. The variable x_i was randomly generated from the continuous uniform distribution $\mathrm{U}(1,4)$. To simulate a mean shift, the value for e_i of observations 70 to 85 was increased by 10 units. Figure 6 illustrates the relationship between y_i and x_i. The process y_i and the mean-shifted values can be seen in Figure 7. Table 2 shows the result of fitting a quadratic relationship between y_i and x_i.

Figure 8 shows that observations 69, 70 and 85 have orthogonal residuals remarkably different from the other values. From this, it could be concluded that the only noticeable feature of the data is the presence of a few outliers. In fact, the mean shift is responsible for these large residual values.

Consider the low-order autocorrelation of orthogonal residuals discussed in Section 3.1, and the covariance properties discussed regarding (3). The residual for observation i depends on the $(i-1)$th and $(i+1)$th observations. Observation 85 has a time series process mean that is the same as that for observation 84 but different from that for observation 86. Consequently, observation 85 has a large orthogonal residual value. Similarly observations 69 and 70 have large residual values. The residual value for observation 86 would also be expected to be large, but the orthogonal residual plot reveals that it is not remarkable. Observations 71 to 84 all have the same time series process means as their neighbours, and therefore do not have large orthogonal residual values.

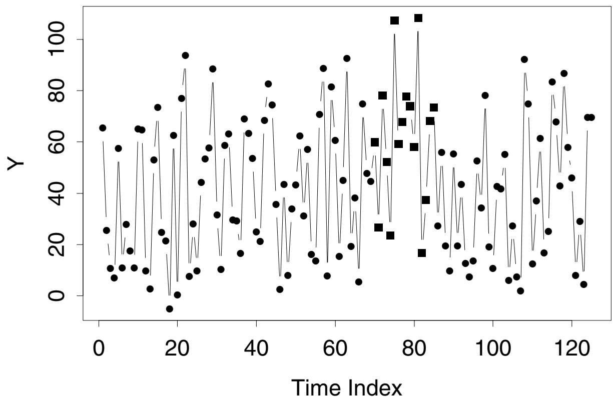

Figure 7. Graph of y_i in time order for the mean shift simulation data. The mean-shifted observations are indicated with squares.

TABLE 2

Summary of parameter estimation for the mean shift simulation data

Parameter	Estimate	Standard error
Intercept	6.30	2.49
X coefficient	–7.23	1.37
X^2 coefficient	7.49	0.28
AR(1) model		
Lag 1 coefficient	0.875	
σ	2.921	

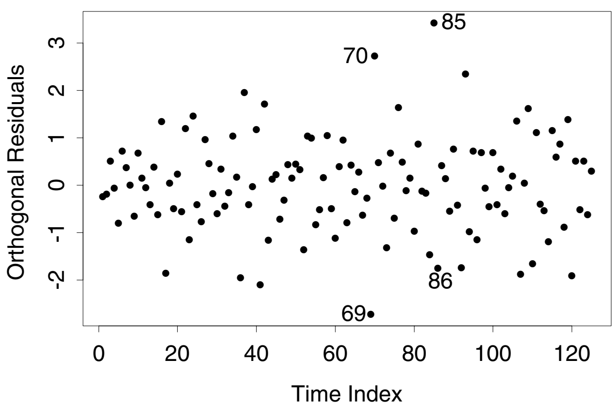

Figure 8. Studentized orthogonal residuals for the mean shift simulation data

As part of model checking, one examines the marginal residuals next. If there may be mean shifts, it can be useful to view the plot of marginal residuals. Because of their autocorrelated nature, marginal residuals can provide insight into the underlying time series process. Figure 9 suggests that observations 70 to 85 do not follow the trend for marginal residuals established by the other observations. The conclusion to be reached here is that it is not the presence of three outlying observations (Figure 8) but the mean shift apparent in Figure 9 that is responsible for the indication of outliers in the plot of orthogonal residuals. This example

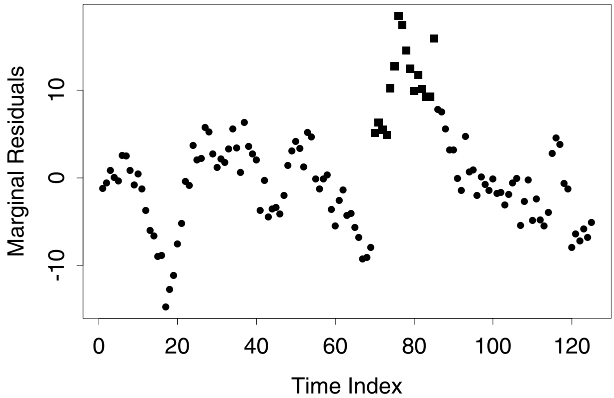

Figure 9. Marginal residuals for the mean shift simulation data. The mean-shifted observations are indicated with squares.

TABLE 3

Summary of parameter estimation for the mean shift simulation data when the quadratic term has been omitted from the model

Parameter	Estimate	Standard error
Intercept	–32.03	2.04
X coefficient	29.37	0.61
AR(2) model		
Lag 1 coefficient	0.354	
Lag 2 coefficient	0.193	
σ	6.761	

highlights how useful it can be to consider the orthogonal residual plot and marginal residual plot together.

4.3. Mis-specification

To demonstrate how orthogonal residuals can be used to identify model mis-specification, consider the mean shift simulation data presented above; there we used a quadratic relationship to model the relationship between y_i and x_i, based on the pattern suggested in Figure 6. Suppose we fitted a straight line relationship instead. Table 3 details the resulting parameter estimates, and reveals that a higher order time series error model was fitted.

The studentized orthogonal residuals (Figure 10) display a quadratic pattern, suggesting that the fitted regression model has been mis-specified. Note that the pattern induced by mis-specification overwhelms any other features in the plot, such as the presence of possible outliers indicated in the orthogonal residual plot for the correctly specified model, Figure 8. When the marginal residuals are plotted against the fitted values (not shown), they follow a quadratic pattern, similar to that displayed in Figure 10.

5. Conditional residuals

For the time series regression model, it is possible to calculate the expectation of Y_t conditional on previous values of the observations. Let $\boldsymbol{Y}_t^{(p)}$ denote the partitioned vector

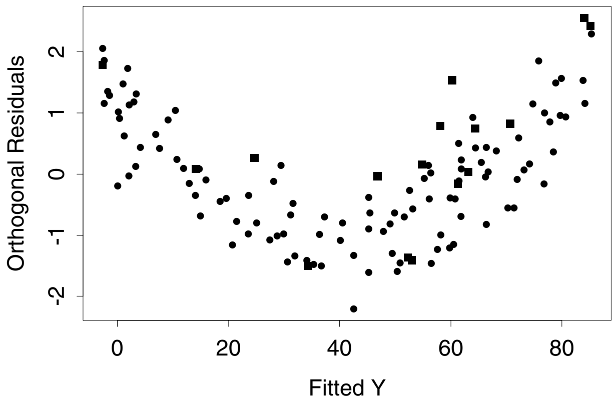

Figure 10. Studentized orthogonal residuals for fitting a mis-specified model to the mean shift simulation data. The mean-shifted observations are indicated with squares.

$[\boldsymbol{Y}^*_{t-p} \mid Y_t]^\top$ where $\boldsymbol{Y}^*_{t-p} = [Y_{t-p} \cdots Y_{t-1}]$. Then $\boldsymbol{Y}_t^{(p)}$ has a multivariate normal distribution with mean

$$[f^*_{t-p} \mid f(\boldsymbol{X}_t)]^\top, \quad \text{where} \quad f^*_{t-p} = [f(\boldsymbol{X}_{t-p})\, f(\boldsymbol{X}_{t-p+1}) \cdots f(\boldsymbol{X}_{t-1})],$$

and covariance matrix

$$\begin{bmatrix} \boldsymbol{\Sigma}_p & \boldsymbol{\gamma}_p \\ \boldsymbol{\gamma}_p^\top & \gamma(0) \end{bmatrix},$$

where $\boldsymbol{\Sigma}_p$ has (i, j)th element $\gamma(|i-j|)$ $(1 \le i, j \le p)$, and $\boldsymbol{\gamma}_p$ has ith element $\gamma(p-i)$. Then, applying equation (8a.2.11) from Rao (1973), we obtain

$$\begin{aligned} \mathrm{E}[Y_t \mid Y_{t-1}, \dots, Y_{t-p}, \boldsymbol{X}] &= f(\boldsymbol{X}_t) + \boldsymbol{\gamma}_p \boldsymbol{\Sigma}_p^{-1} (Y^*_{t-p} - f^*_{t-p})^\top \\ &= \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + f(\boldsymbol{X}_t) - \phi_1 f(\boldsymbol{X}_{t-1}) - \cdots - \phi_p f(\boldsymbol{X}_{t-p}), \end{aligned}$$

because $\boldsymbol{\gamma}_p \boldsymbol{\Sigma}_p^{-1} = [\phi_p\ \phi_{p-1}\ \cdots\ \phi_1]$ by the Yule–Walker equations (see e.g. Brockwell & Davis, 1991 p.239).

We call the difference between Y_t and this expectation the 'conditional residual'

$$\begin{aligned} \hat{z}_t &= Y_t - \hat{\phi}_1 Y_{t-1} - \hat{\phi}_2 Y_{t-2} - \cdots - \hat{\phi}_p Y_{t-p} - \hat{f}(\boldsymbol{X}_t) + \hat{\phi}_1 \hat{f}(\boldsymbol{X}_{t-1}) + \cdots + \hat{\phi}_p \hat{f}(\boldsymbol{X}_{t-p}) \\ &= \hat{\Phi}_p(B)\hat{e}_t \qquad (\text{for } t > p). \end{aligned}$$

Assuming that the regression model has not been mis-specified, the conditional residuals are estimates for the unobservable z_t, and can be graphed to assess whether $\hat{z}_t$ satisfies model assumptions.

An alternative approach for deriving the conditional residuals is to transform the terms in model (2) so that the transformed errors are uncorrelated. Let $\boldsymbol{P}$ be a lower triangular matrix such that $\boldsymbol{\Sigma}^{-1} = \boldsymbol{P}^\top \boldsymbol{P}$. Then multiplying (2) by $\boldsymbol{P}$ we obtain $\boldsymbol{PY} = \boldsymbol{PX\beta} + \boldsymbol{Pe}$. The covariance matrix for this transformed model is $\mathrm{var}(\boldsymbol{Pe}) = \boldsymbol{P}^\top \mathrm{var}(\boldsymbol{e})\boldsymbol{P} = \boldsymbol{I}$, so the error terms are independent with unit variance.

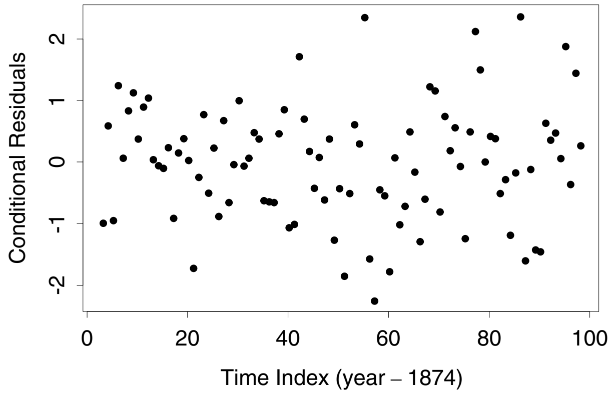

Figure 11. Standardized conditional residuals for the Lake Huron data

The effect of the transformation is easy to understand, in the AR(1) case, for example, where

$$\boldsymbol{P} = \frac{1}{\sigma}\begin{bmatrix} \sqrt{1-\phi^2} & 0 & 0 & \cdots & 0 & 0 \\ -\phi & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\phi & 1 & \ddots & & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & & \ddots & \ddots & 1 & 0 \\ 0 & 0 & \cdots & 0 & -\phi & 1 \end{bmatrix}.$$

Then, for $t > p$, the tth term of $\boldsymbol{Pe}$ is $(e_t - \phi e_{t-1})/\sigma = z_t/\sigma$. However, this transformed model also allows the calculation of conditional errors for $1 \le t \le p-1$.

Generally, for $t > p$, the (t,i)th element of $\boldsymbol{P}$ is

$$P_{t,i} = \begin{cases} 1/\sigma & i = t, \\ -\phi_k/\sigma & i = t-k, \quad k = 1, \dots, p, \\ 0 & \text{otherwise} \end{cases}$$

(see Knottnerus, 1991 p.15). Consequently, $\boldsymbol{P}Y_t = \Phi_p(B)Y_t/\sigma$, $\boldsymbol{PX}_t = \Phi_p(B)\boldsymbol{X}_t/\sigma$ and $\boldsymbol{P}e_t = \Phi_p(B)e_t/\sigma = z_t/\sigma$, for $t > p$. Therefore, the conditional residuals are $\hat{z}_t = \hat{\sigma}\hat{\boldsymbol{P}}\hat{e}_t$, where $\hat{\sigma}$ and $\hat{\boldsymbol{P}}$ are estimates of σ and $\boldsymbol{P}$. We call the quantities $\hat{\boldsymbol{P}}\hat{e}_t = \hat{z}_t/\hat{\sigma}$ 'standardized' conditional residuals.

A relationship exists between conditional residuals and orthogonal residuals. This is demonstrated by premultiplying the standardized conditional residual by $\hat{\boldsymbol{P}}^\top$, yielding

$$\hat{\boldsymbol{P}}^\top\hat{\boldsymbol{z}}/\hat{\sigma} = \hat{\boldsymbol{P}}^\top\hat{\boldsymbol{P}}\hat{\boldsymbol{e}} = \hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{e}} = \hat{\boldsymbol{v}}.$$

For an AR(1) error model, $\hat{\sigma}\hat{v}_t = \hat{z}_t - \hat{\phi}\hat{z}_{t+1}$ for $t \ge 2$, which is a non-invertible MA(1) process. This result is as expected from the results in Section 3.1. In general, $\hat{\sigma}\hat{v}_t = \hat{\Phi}_p(B^{-1})\hat{z}_t$ (for $t > p$) is a non-invertible MA(p) process, where B^{-1} denotes the forward shift operator.

Seber (1977 p.172) presents a different method of transformation which results in Best Linear Unbiased Scaled residuals. This produces a set of $(n-p)$ transformed residuals, instead of the n residuals produced from the methods described above.

Figure 11 is a plot of standardized conditional residuals for the Lake Huron dataset examined above. The conditional residuals are uncorrelated and appear to indicate that model assumptions regarding z_t are satisfied.

5.1. Interpreting conditional residuals

As stated above, the conditional residual $\hat{z}_t$ is an estimate of the unobservable z_t when the model has not been mis-specified. However, mis-specification of the time series error model can greatly affect the conditional residuals. Consider an example for which the following AR(3) time series error model is appropriate

$$e_t = \phi_1 e_{t-1} + \phi_2 e_{t-2} + \phi_3 e_{t-3} + z_t ,$$

but for which an AR(1) error model is used in the time series regression model. The resulting conditional residuals are no longer a function of z_t alone. This error model mis-specification can result in conditional residual plots with patterns induced by autocorrelation.

Therefore, the procedure used to fit a time series regression model to data affects the way the conditional residuals should be interpreted. A procedure that allows the error model to be optimally chosen through an iterative procedure (such as in iterative GLS) usually results in conditional residuals demonstrating a white noise pattern. If instead the analyst specifies the error model, patterns in the conditional residuals may be attributable to mis-specification of the error model rather than, say, mean shifts or outliers.

Further complications can arise when the regression model is mis-specified. If the error model is fixed by the analyst, unexplained variation that exists because of the regression model mis-specification is not accounted for in the time series error model. This unexplained variation is present in the conditional residuals, resulting in patterns in residual plots.

The situations outlined above indicate the sensitivity of the conditional residual to model mis-specification. Because of this, we suggest that conditional residuals only be examined once the orthogonal and marginal residuals have been analysed and any apparent model mis-specification has been corrected.

6. Unified use of residuals

The marginal, orthogonal and conditional residuals can be used together for checking and analysing the model, as shown by the following example.

In a metal production facility a response y_i depends on another variable x_i and measurements of both are recorded over time. Thirty-six pairs of observations are shown in Figure 12, where a linear relationship between y_i and x_i appears appropriate. (The data presented are a linear transformation of observations recorded directly from the production plant. For reasons of confidentiality, the names of the variables and their origin cannot be disclosed.) Observations 23 and 24 are numbered because they are prominent in the plots to follow. Note also the points depicted as squares in the top-right-hand corner of the plot; these correspond to observations 12 to 14.

Figure 13 is a plot of y_i in time order and shows evidence of autocorrelation. Note that observations 12 to 14 appear to be inconsistent with the trend set by the other observations. These observations are not necessarily outliers because they also have large x_i values as shown in Figure 12.

Table 4 summarizes the result of fitting a straight line relationship between the two variables. To assess the fit of the model in Table 4, we examine the orthogonal residuals presented in Figure 14. The graph does not reveal any mis-specification or any other problems in the fitted regression model. Observations 23 and 24 are again labelled.

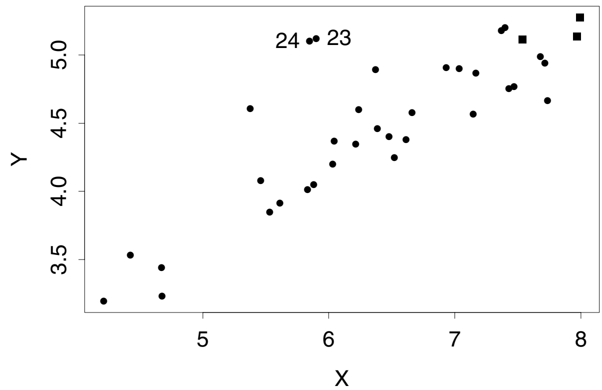

Figure 12. Graph of y_i vs x_i for the metal production data. Observations 12, 13 and 14 are denoted with squares. Observations 23 and 24 are numbered.

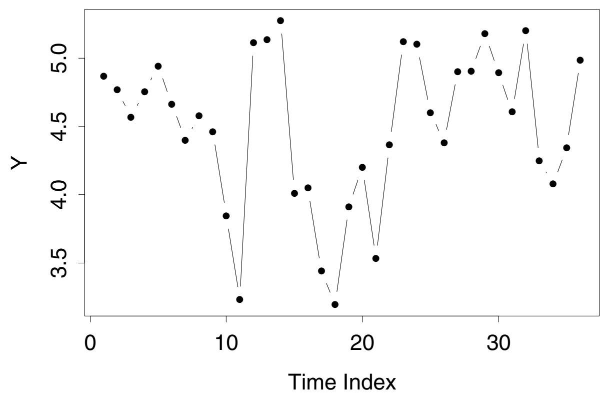

Figure 13. Graph of y_i in time order for the metal production data

TABLE 4

Summary of parameter estimation for the metal production data

Parameter	Estimate	Standard error
Intercept	1.415	0.3259
X coefficient	0.479	0.0478
AR(1) model		
Lag 1 coefficient	0.583	
σ	0.255	

Figure 15, showing the marginal residuals, reveals the underlying time series error process. Here, observations 23 and 24 have residual values that are inconsistent with the trend established by the other values. These two observations do not follow the time series error model assumed to be responsible for autocorrelation in the observations.

Finally, Figure 16, showing the conditional residuals, confirms that observation 23 is discordant. Since observations 23 and 24 are both large and have similar y_i values, the resulting conditional residual for observation 24 is not discordant.

The apparently inconsistent conclusions between the three residual plots illustrate that these plots should be interpreted differently. In the metal production data above, observations

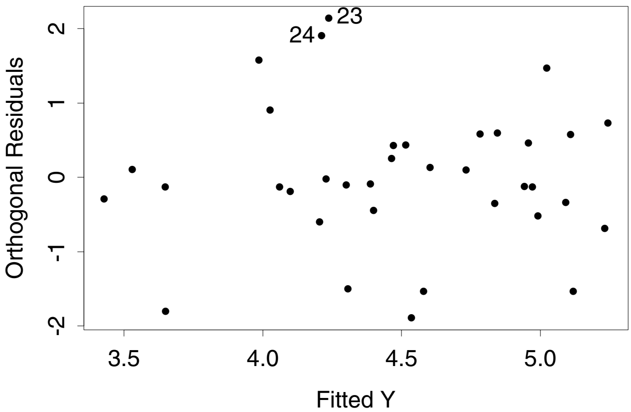

Figure 14. Studentized orthogonal residuals for the metal production data. Observations 23 and 24 are numbered.

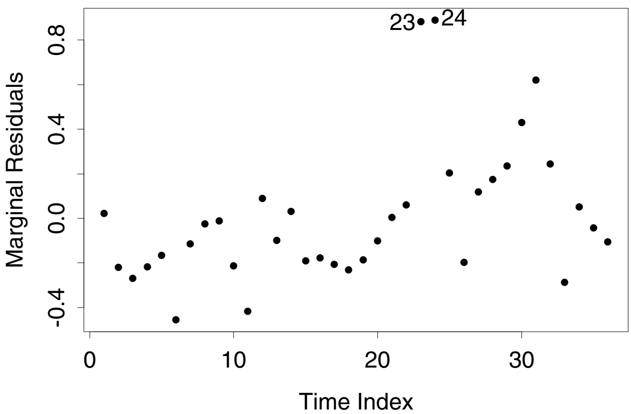

Figure 15. Marginal residuals for the metal production data. Observations 23 and 24 are numbered.

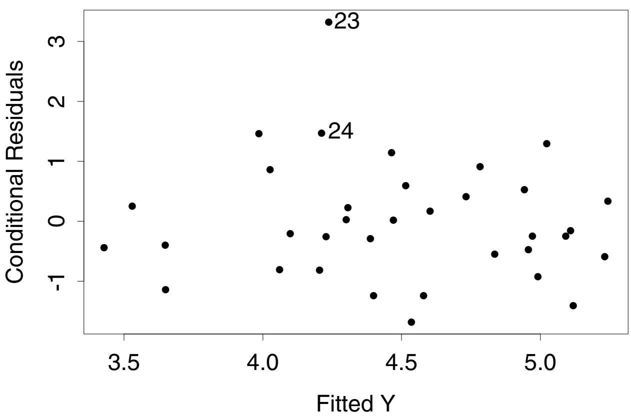

Figure 16. Conditional residual plot for the metal production data. Observations 23 and 24 are numbered.

23 and 24 have been highlighted as unusual observations. As far as the regression model is concerned, these observations are not outliers; their y_i and x_i values are consistent with those of other observations. However, the marginal residual values suggest these observations are discordant. Reconciling these two conclusions suggests that the process was producing y_i values (with associated x_i values) consistent with other observations, but that these values were not expected at time points 23 and 24; these observations were produced contrary to the underlying autocorrelation. One could surmise that some special cause was operating over the interval in which these two observations were recorded.

7. Conclusion

We have considered the need for suitable residual diagnostic plots for time series regression. Although the marginal residual may be intuitively appealing, we have shown that it is not suitable for identifying mis-specification in the regression model. However, it is useful for checking the unobserved time series error process if the regression model is correctly specified. For identifying model mis-specification, we have proposed the orthogonal residual which is orthogonal to both the fitted values and the covariates, and which possesses low-order autocorrelation. When used in conjunction with marginal residual plots, the orthogonal residual plots can help identify mean shifts and other patterns. Finally, we have shown that conditional residuals are useful in checking the white noise error component. These residuals are sensitive to the regression and error model fitted, and we suggest that they be analysed only after orthogonal and marginal residuals have been examined.

Together, these three residuals provide the means for examining various aspects of the fitted model and for identifying problems such as model mis-specification, mean shifts and outliers.

References

ANDERSON, O.D. (1976). On the inverse of the autocorrelation matrix for a general moving average. *Biometrika* **63**, 391–394.

BAGSHAW, M. & JOHNSON, R.A. (1977). Sequential procedures for detecting parameter changes in a time series model. *J. Amer. Statist. Assoc.* **72**, 593–597.

BROCKWELL, P.J. & DAVIS, R.A. (1991). *Time Series: Theory and Methods*, 2nd edn. New York: Springer-Verlag.

FULLER, W.A. (1996). *Introduction to Statistical Time Series*, 2nd edn. New York: John Wiley & Sons.

HASLETT, J. & HAYES, K. (1998). Residuals for the linear model with general covariance structure. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **60**, 201–215.

HOSSAIN, A. (1990). Detection of influential observations in regression model with autocorrelated errors. *Comm. Statist. Theory Methods* **19**, 1047–1060.

JUDGE, G.G., HILL, R.C., GRIFFITHS, W.E., LÜTKEPOHL, H. & LEE, T. (1988). *Introduction to the Theory and Practice of Econometrics*, 2nd edn. New York: John Wiley & Sons.

KNOTTNERUS, P. (1991). *Linear Models with Correlated Disturbances.* Berlin: Springer-Verlag.

LEDOLTER, J. (1988). Outlier diagnostics in time series analysis. *J. Time Ser.* **11**, 317–324.

LJUNG, G.M. (1993). On outlier detection in time series. *J. Amer. Statist. Assoc.* **55**, 559–567.

MARTIN, R.J. (1992). Leverage, influence and residuals in regression models when observations are correlated. *Comm. Statist. Theory Methods* **21**, 1183–1212.

MURTHY, D.N.P. (1974). On the inverse of the covariance matrix of a first order moving average. *Sankhyā Ser. A* **36**, 223–225.

PUTERMAN, M.L. (1988). Leverage and influence in autocorrelated regression models *Appl. Statist.* **37**, 76–86.

RAO, C.R. (1973). *Linear Statistical Inference and Its Applications*, 2nd edn. New York: John Wiley & Sons.

SEBER, G.A.F. (1977). *Linear Regression Analysis.* New York: John Wiley & Sons.

TSAY, R.S. (1984). Regression models with time series errors. *J. Amer. Statist. Assoc.* **79**, 118–124.

TSAY, R.S. (1986). Time series model specification in the presence of outliers. *J. Amer. Statist. Assoc.* **81**, 132–141.