ELSEVIER

Journal of Econometrics 80 (1997) 387–422

JOURNAL OF Econometrics

Detecting shocks: Outliers and breaks in time series

A.C. Atkinson[a], S.J. Koopman[c], N. Shephard*[,b]

[a] *Department of Statistics, London School of Economics, Houghton St., London WC2A 2AE, UK*
[b] *Nuffield College, Oxford, OX1 1NF, UK*
[c] *CentER, Tilburg University, 5000 LE Tilburg, The Netherlands*

Abstract

A single outlier in a regression model can be detected by the effect of its deletion on the residual s im of squares. An equivalent procedure is the simple intervention in which an extra parameter is added for the mean of the observation in question. Similarly, for unobserved components or structural time-series models, the effect of elaborations of the model on inferences can be investigated by the use of interventions involving a single parameter, such as trend or level changes. Because such time-series models contain more than one variance, the effect of the intervention is measured by the change in individual variances.

We examine the effect on the estimated parameters of moving various kinds of intervention along the series. The horrendous computational problems involved are overcome by the use of score statistics combined with recent developments in filtering and smoothing. Interpretation of the resulting time-series plots of diagnostics is aided by simulation envelopes.

Our procedures, illustrated with four example, permit keen insights into the fragility of inferences to specific shocks, such as outliers and level breaks. Although the emphasis is mostly on parameter estimation, forecast are also considered. Possible extensions include seasonal adjustment and detrending of series. © 1997 Elsevier Science S.A.

Keywords: Added variable; Deletion methods; Diagnostics; Dynamic linear mode; EM-algorithm; Fragility; Kalman filter; Monte-Carlo test; Outlier; Score test; Shocks; Simulation envelope; Structural change; Structural time-series model; Unobserved components model
JEL classification: C13; C22; C52; C53

* Corresponding author.

PII S0304-4076(97)00050-X

1. Purpose

The purpose of this paper is to develop powerful methods for detecting specific inadequacies in time-series models. We use the unobserved component or structural approach to time-series modelling, that is dynamic linear methods, to provide procedures which have much in common with the well-known diagnostics for multiple-regression models. A single outlier in a regression model can be detected by the effect of its deletion on the residual sum of squares. An equivalent procedure is the simple intervention in which an extra parameter is added for the mean of the observation in question. The effect on the residual sum of squares is identical in the two cases. Similarly, for unobserved-component time-series models, the effect of changes in the data on the model can be investigated by the use of interventions corresponding to physically identifiable features involving a single parameter, such as trend or level changes. Because these time-series models include more than one variance, the effect of the intervention is measured by the effect on the individual variances, which we call the parameters of the model. The paper describes the working out of this simple idea and illustrates it with examples.

There is a large and growing literature on regression diagnostics, including the book length treatments of Cook and Weisberg (1982), Atkinson (1985) and Chatterjee and Hadi (1988). Recent review articles for the extension to generalized linear models are Davison and Tsai (1992) and O'Hara Hines and Carter (1993). Central to this literature is the distinction between outlier detection and the influence of observations on specific inferences. There is likewise a large literature on the detection of outliers in time series, which is often taken to start with Fox (1972). The distinction made there between additive and innovation outliers is extended to multiple outliers by Bruce and Martin (1989) in a paper which, with its discussion, contains references to much of the literature on outliers in time series, for example, Muirhead (1986) and Tsay (1986). A more recent paper is Ljung (1993). Chen and Liu (1993) study the joint estimation of parameters and of outliers, while Balke (1993) demonstrates the difficulty outlier detection methods can face in the presence of a shift in level. However, there has been appreciably less work on influential observations than on outlier detection.

In regression models, an overall measure of the influence of an observation is Cook's distance (Cook, 1977). Peña (1990) obtains a similar measure for ARIMA models extended in Peña (1991) to models including regression. For state-space models, Kohn and Ansley (1989) obtain studentized residuals and leverage measures analogous to those for regression models when the variance parameters are known. Harrison and West (1991) and Harrison and Veerapen (1993) extend these results to Bayesian and non-Bayesian deletion diagnostics and obtain a version of Cook's distance. These results do not assume that all the variances are known, although it is crucial that the signal-to-noise ratios are assumed given. Atkinson and Shephard (1996) use the results to calculate deletion t-statistics for testing the influence of individual observations on regression coefficients. Their examples concentrate on constructed variables for power transformation of the time series.

Our approach is based on inspecting plausibly interesting and interpretable departures, some of which are related to those examined by the more conventional diagnostics. These departures are generated by a generalized form of intervention analysis. In standard intervention analysis interest is in the effect of a known exogenous change on the model. For example, Box and Tiao (1975) use an ARMA model to analyse the effect of the opening of a freeway in 1960, and of changes in the design of car engines in 1966, on the level of ozone in Los Angeles. A second example is given by Harvey and Durbin (1986) in which the interest is in the effect on the number of car drivers killed of legislation which became effective at the end of January 1983.

Although Harvey and Durbin, unlike Box and Tiao, use unobserved-component time-series models, the principle is the same in that these methods assume that the specific times of intervention or of shock to the system are known. A more recent development among time-series econometricians has been the attempt to detect the time of specific shocks in series, for example, the appendix to McCulloch and Tsay (1993) lists events that might have effected the world market in oil. More generally, following Perron (1989) and Rappoport and Reichlin (1989), there has been a significant effort to identify the time of slowdown in the growth rate of GDP in the US. Most of the work has focussed on univariate autoregressive models and the use of non-standard asymptotic theory to look at the maximum of breakpoint statistics over the whole sample. An example of recent work is Christiano (1992), who references much of the literature. Our purpose is the more general diagnostic one of determining whether inferences from the model, specifically parameter estimates and forecasts, are fragile, i.e. is they respond significantly to changes in the fitted model. To achieve this we use an intervention for each unobserved component and examine the effect as the intervention is moved along the series. Thus, for the irregular component of the model we look at the effect of modelling a single additive outlier at each time point.

Plausible departures from the model are given in Section 2. Since unobserved components models have ARIMA representations, our methods have an ARIMA interpretation. We argue that an advantage of our methods is that single identifiable causes for departures are more easily identified than by the procedures for ARIMA models which tend to smear departures in one component into diagnostics for several aspects of the model. In Section 3 we discuss how to assess the effect of the shock. Possible measures include the change in log-likelihood or the change in the estimated parameter. For inferential purposes we work with a marginal-likelihood function. Since we look at up to four forms of intervention for each time point, it is clearly not possible to re-estimate the parameters for each intervention. We therefore use a score statistic which avoids such re-estimation. From this we derive a one-step estimate of the effect of the intervention on the parameter. The statistical significance of these estimates is assessed by plots incorporating simulation envelopes. The details of the derivation of the score function and one-step estimates are given in Section 4. Also, in Section 4 we present a brief comparison of our scaled one-step estimator with the t-statistic found by running each intervention along the series. The t-statistic is improved by use of the one-step estimate to adjust the estimated variance for the effect of the intervention.

In standard regression models, the effect of the addition of an extra variable can be assessed by the use of an added-variable plot in which the residuals of the response are plotted against the residuals of the extra variable. Both sets of residuals are from regression on all variables already in the model. Regression of one set of residuals on the other yields the coefficient for the new variable in the multiple-regression model containing all variables, together with its associated t-test. Related, although more complicated, results for unobserved component models are given in Section 4.1. The combination of these results for added variables with the results on score tests provides elegant and speedy methods for the calculation of the simulation envelopes and stresses the relationship with diagnostics for regression models. These results are illustrated with an analysis of data on US exports to Latin America. Properties of the simulation envelope are investigated in Section 5. Further examples are in Section 6. In Section 7 we consider the effect of shocks on forecasts and the paper concludes in Section 8 with a few comments on unsolved problems.

2. Interesting directions

The unobserved component or Gaussian state-space model has proved to be a useful tool for handling linear and many non-linear time-series models. Book length treatments of the subject include West and Harrison (1989), who prefer the name dynamic linear models (Harvey, 1989 especially Ch. 3) and, from an engineering perspective, Anderson and Moore (1979). General expressions for the measurement and transition equations are given in Section 4. The model for the observations y_t may contain regression variables and, indeed, this is how the shocks are introduced in Section 4. In all our examples, y_t contains an irregular component. In addition, there may be level, slope, seasonal or other components such as cycles. To test how fragile the estimated model is, we modify one of the components, or introduce a new one, and see whether there are changes in important features of the fitted model. In particular, tests for stability to shocks in those components which are included, are determined from the effect of the addition of regression variables of the form shown in Fig. 1. We calculate the effect of moving each regression variables along the series, so adding one extra

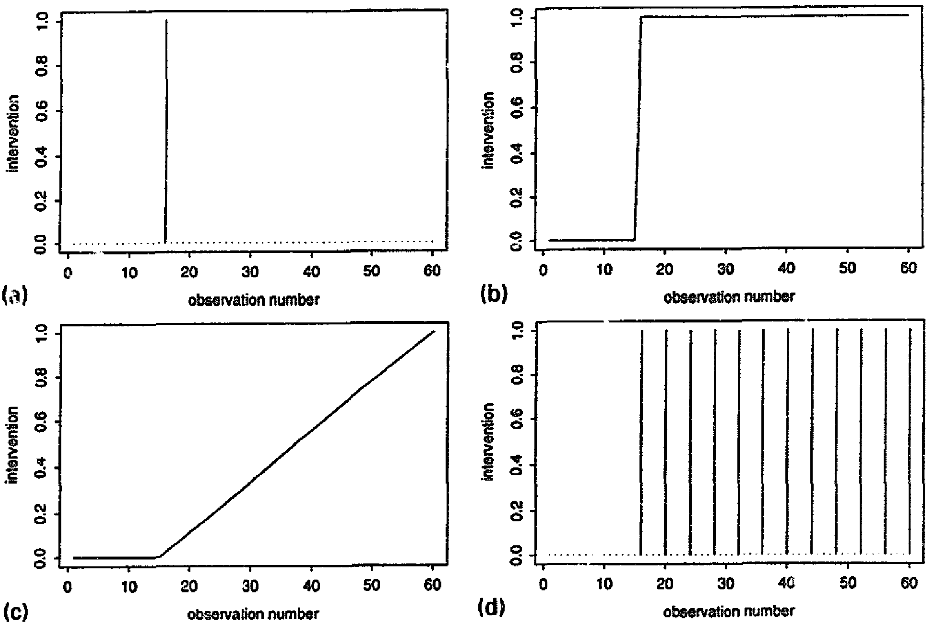

Fig. 1. Added variables for interventions modelling: (a) an additive outlier; (b) a level break; (c) a slope change and (d) seasonal change.

parameter to the model at each time point in turn, and see how the fit of the model changes as a result of this extra parameter. A large decrease in the estimate of one of the variances associated with the measurement or transition equations indicates sensitivity of the model to the intervention. Suppose, for example, there were an unusually large level break in the series at $t = 16$. The fitted model would then have an inflated value of the variance for the stochastic level. The addition of the regression variable of Fig. 1b at $t = 16$ would dramatically reduce the variance, which would also be somewhat reduced by adding the variable at $t = 15$ or 17. However, introducing the variable at appreciably larger or smaller values of t would have no, or a negligible, effect.

For regression models with independent errors the irregular intervention clearly corresponds to deletion of an observation. But with the more complicated error structure of time-series models, outliers affect neighbouring observations. Consider, for example, the effect of an additive outlier on an ARIMA model (Fox, 1972). Suppose that the model is a random walk. Then differencing will be necessary to obtain white noise. The result is that the additive outlier will be smeared over two adjacent time points. Similarly, differencing twice to remove a stochastic linear trend will spread the outlier over three time points. This effect is noticeable in the plots of the 'leave-k-out' diagnostics of Bruce and

Martin. The argument of the preceding paragraph on the effect of the variable for the change in level indicates that some of our diagnostics may also be subject to some smearing. We show in Section 4 that this effect is reduced by using the one-step estimates of the changes in the parameters due to the shocks.

The result of our analysis for any particular shock is a time-series plot accompanied by a simulation envelope. Plots of a similar type, but without envelopes, are given by Harvey and Koopman (1992) who estimate the residuals for the measurement and transition equations. The relationship between the two procedures is explored in Section 4. Harvey and Koopman rely on aggregate statistics, such as the skewness and kurtosis tests of normality and their combination in the Bowman–Shenton statistic (Bowman and Shenton, 1975), to provide tests of statistical significance. In line with the approach of regression diagnostics, we focus instead of statistics associated with individual observations and specific features of the data or model.

3. Measuring the effect of shocks

Let the parameter of interest be θ, which is a vector function of the variances in the measurement and transition equations, and let $\hat{\theta}$ be its maximum-likelihood estimator. It is the value of θ which controls the rate of discounting of past observations. We want to measure the effect on the estimation of θ of the intervention modelled by inclusion of the regression variable with coefficient β. One possibility, in line with the approach to added and constructed variables in regression, is to calculate the t-statistic for β. It is however not clear that this is the most appropriate quantity. The t-statistic may be expected to be more informative about how the estimate of β changes near the intervention than it is about changes in the estimate of θ. For the remainder of this section we focus on estimation of θ. The close relationship between the t-statistic and changes in the estimate of θ due to the intervention is described in Section 4.1, with some numerical assessment of the relationship in Section 4.3.

There is, however, an inferential complication. We require to make inferences about the time-series parameter θ in the presence of the nuisance parameter β. It is clearly possible to work with the profile-log-likelihood found by maximizing over β for given θ. But unless β and θ are approximately orthogonal (Cox and Reid, 1987) a biased estimate of θ will result. This is certainly the case for unobserved components models (Shephard, 1993b).

Instead we could

(1) employ a Bayesian argument to place a diffuse prior on the coefficient β and integrate out leading to a new likelihood function;
(2) construct a marginal-likelihood function (Tunnicliffe Wilson, 1989).
(3) derive the modified profile-likelihood function.

For the class of models we are considering (Gaussian unobserved components models with explanatory variables) the three approaches mercifully yield an identical solution.

Let the variable applied at time t be of the jth type, perhaps one of the four of Fig. 1. To stress the connection with deletion diagnostics in regression we call the estimate of the parameter from the marginal-likelihood $\hat{\theta}_{(t,j)}$. If the maximized log-likelihood for the unperturbed model is $L(\hat{\theta})$, the difference in the log-likelihood due to addition of the explanatory variable is

$$L_{(t,j)}(\hat{\theta}_{(t,j)}) - L(\hat{\theta}). \tag{1}$$

Because $L_{(t,j)}(.)$ is a marginal-log-likelihood this difference can be positive or negative: the standard asymptotic chi-squared distribution does not apply. Calculation of (1) requires estimation of the parameter $\hat{\theta}_{(t,j)}$. Unfortunately, the estimation of the parameters in an unobserved components model is the computationally most intensive step, involving many passes of the Kalman filter. We therefore work with a score version of the difference in log-likelihoods which avoids re-estimation of the parameters for each combination of t and j.

Let the score for the marginal-log-likelihood be

$$s_{(t,j)}(\theta) = \frac{\partial L_{(t,j)}(\theta)}{\partial \theta}$$

with the corresponding definition for $s(\theta)$ as a derivative of $L(\theta)$. Then the change or difference in scores due to the intervention evaluated at some θ is

$$c_{(t,j)}(\theta) = s_{(t,j)}(\theta) - s(\theta), \tag{2}$$

The second term on the right-hand side of (2) is identically zero when $\theta = \hat{\theta}$. Although these scores are of interest, they are relatively highly correlated across the various elements of θ. A more nearly orthogonal set of variables is obtained from the one-step estimates of the parameters

$$\tilde{\theta}_{(t,j)} = \hat{\theta} + \hat{J}^{-1}(\hat{\theta})s_{(t,j)}(\hat{\theta}) = \hat{\theta} + \hat{J}^{-1}(\hat{\theta})c_{(t,j)}(\hat{\theta}). \tag{3}$$

In (3) $\hat{J}(\hat{\theta})$ is the observed information found from the numerical matrix of second derivatives from the likelihood fitting. Explicit expressions for the score and one-step estimates for vector θ are given in the next section. For plotting we standardize the one-step estimates to give Wald-like t-statistics for the elements of $\tilde{\theta}_{(t,j)}$ (for example, Cox and Hinkley, 1974, p. 314).

The empirical distribution of $\tilde{\theta}_{(t,j)}$ can be found by simulation using $\hat{\theta}$ as the true parameter value to generate new series $y^{(k)}$ for k from 1 to K. The maximum-likelihood estimate of θ for this simulation is $\hat{\theta}^{(k)}$ and we can find the simulated one-step estimate $\tilde{\theta}^{(k)}_{(t,j)}$ from $c^{(k)}_{(t,j)}\{\hat{\theta}^{(k)}\}$ given by (2) when the data are $y^{(k)}$ and $\theta = \hat{\theta}^{(k)}$. However, we avoid calculation of $\hat{\theta}^{(k)}$ by noting that $c_{(t,j)}(\theta)$ is the difference of two scores. If the model is correctly specified and the sample size is large, its log-likelihood will be approximately quadratic, with the same curvature as that of its simulated versions. Thus, the two scores will be approximately linear in θ with the same slope so that, on taking differences, the precise value of θ used to evaluate the scores is not important. So, from (3), we take the empirical distribution of

$$\breve{\theta}^{(k)}_{(t,j)} = \hat{\theta} + \hat{J}^{-1}(\hat{\theta})c^{(k)}_{(t,j)}(\hat{\theta}), \tag{4}$$

where

$$c^{(k)}_{(t,j)}(\hat{\theta}) = s^{(k)}_{(t,j)}(\hat{\theta}) - s^{(k)}(\hat{\theta}), \tag{5}$$

but now $s^{(k)}(\hat{\theta})$ is, in general, not equal to zero. We stress the importance of this simple approximation which makes feasible the calculation of simulation envelopes for the computationally intensive methods of this paper.

4. Evaluating the score

The models we analyse in this paper can be placed into the Gaussian state-space form

$$\begin{aligned} &y_t = Z_t\alpha_t + \varepsilon_t,\ \alpha_t = T_t\alpha_{t-1} + \eta_t, \quad t = 1, \dots, n, \\ &\alpha_0 | Y_0 \sim \mathrm{N}(a_0, P_0), \qquad \varepsilon_t \sim \mathrm{N}(0, H_t), \qquad \eta_t \sim \mathrm{N}(0, Q_t), \end{aligned} \tag{6}$$

where Y_t denotes the information available at time t. Here $\alpha_0 | Y_0$, (ε_t) and (η_s) are independent of one another for all t and s and the initial conditions a_0 and P_0 are known. We will assume, for simplicity of exposition, that all the parameters which index this model are in the H_t and Q_t matrices. Then the score can be derived in two stages, following Koopman and Shephard (1992), who exploit an argument similar to that leading to the EM algorithm. First a Kalman filter is run through the data

$$\begin{aligned} &a_{t+1|t} = T_{t+1}a_{t|t-1} + K_t v_t, && P_{t+1|t} = T_{t+1}P_{t|t-1}L_t' + Q_{t+1}, \\ &v_t = y_t - Z_t a_{t|t-1}, && F_t = Z_t P_{t|t-1} Z_t' + H_t, \\ &K_t = T_{t+1}P_{t|t-1}Z_t'F_t^{-1}, && L_t = T_{t+1} - K_t Z_t, \quad t = 1, \dots, n. \end{aligned} \tag{7}$$

storing only v_t, F_t^{-1} and K_t. In our implementation this is followed by the disturbance smoother of Koopman (1993) which is passed through the data, starting with $r_n = 0$ and $N_n = 0$ and letting

$$\begin{aligned} &e_t = F_t^{-1}v_t - K_t'r_t, && r_{t-1} = Z_t'F_t^{-1}v_t + L_t'r_t, \\ &D_t = F_t^{-1} + K_t'N_tK_t, && N_{t-1} = Z_t'F_t^{-1}Z_t + L_t'N_tL_t, \quad t = n, \dots, 1. \end{aligned} \tag{8}$$

In (8) e_t, r_{t-1}, D_t and N_{t-1} have a simple intuitive interpretation: H_te_t is the smoothed estimate of ε_t and Q_tr_{t-1} the corresponding estimate for η_t. The mean squared errors are, respectively, $H_t - H_t D_t H_t'$ and $Q_t - Q_t N_{t-1} Q_t'$. These two recursive calculations yield the score function which has the form

$$\frac{\partial \log L(\theta)}{\partial \theta_i} = \frac{1}{2}\sum_{t=1}^{n} \operatorname{tr}\left((e_t e_t' - D_t)\frac{\partial H_t}{\partial \theta_i}\right) + \frac{1}{2}\sum_{t=1}^{n} \operatorname{tr}\left((r_{t-1} r_{t-1}' - N_{t-1})\frac{\partial Q_t}{\partial \theta_i}\right). \tag{9}$$

which is one of the Koopman and Shephard results.

As the disturbance smoother only involves the multiplication of typically sparse vectors and matrices it tends to be much more rapid than the conventional smoothing algorithms. Usually, it takes about the same time as does the Kalman filter for a forward pass.

4.1. Algebra of added variables

Suppose we now change the model to

$$\begin{aligned} &y_t = Z_t\alpha_t + X_t\beta + \varepsilon_t, \qquad \alpha_t = T_t\alpha_{t-1} + \eta_t, \quad t = 1, \ldots, n, \\ &\begin{pmatrix}\beta \\ \alpha_0\end{pmatrix} \Big| Y_0 \sim N\left(\begin{pmatrix} S_0^{-1}s_0 \\ a_0\end{pmatrix}, \begin{pmatrix} S_0^{-1} & 0 \\ 0 & P_0\end{pmatrix}\right), \\ &\varepsilon_t \sim \mathrm{N}(0, H_t), \qquad \eta_t \sim \mathrm{N}(0, Q_t), \end{aligned} \tag{10}$$

where the prior is now independent of ε_t and η_s for all t and s. Usually, S_0 and s_0 may be taken to be a matrix and vector of zeros, respectively. Here X_t allows the introduction of explanatory variables, and hence interventions, into the model.

At first sight it seems that if we wish to manipulate the filtered and smoothed quantities, or the corresponding score, for this new model we have to re-run both (7) and (8), followed by recalculation of (9). However, this is not the case. Use of the auxiliary recursions of de Jong (1989) circumvents this necessity. His recursions have a number of parts. The first part, the auxiliary filter, corrects the old run of the Kalman filter for the presence of the new explanatory variables. It does this by first computing

$$E\beta | Y_t = b_t = S_t^{-1} s_t \quad \text{and} \quad E(\beta - b_t)(\beta - b_t)' | Y_t = S_t^{-1}, \tag{11}$$

where S_t and s_t are found recursively through

$$\begin{aligned} &s_t = s_{t-1} + V_t' F_t^{-1} v_t, \qquad S_t = S_{t-1} + V_t' F_t^{-1} V_t, \\ &V_t = -X_t - Z_t A_{t|t-1}, \qquad A_{t+1|t} = T_t A_{t|t-1} + K_t V_t, \quad t = 1, \ldots, n. \end{aligned}$$

starting with $A_{1|0} = T_1 a_0$.

The auxiliary filter is the generalization to time-series models of the calculations leading to the added variable plot for regression models mentioned in Section 1. There the effect of a new regression variable X_t was found by first calculating its residual after regression on the other variables. Regression of the residuals of y_t on these new residuals provided quantities for the adjustment of the fitted model for inclusion of the extra variable. The same intuition applies to the auxiliary filter. We re-run the Kalman filter not on y_t, but instead on $-X_t$. The model is now

$$-X_{tj} = Z_t\alpha_t^* + \varepsilon_t, \qquad \alpha_t^* = T_t\alpha_{t-1}^* + \eta_t, \quad t = 1, \dots, n, \quad j = 1, \dots, p, \tag{12}$$

that is each vector of the matrix X_t is fitted by the old model. This generates the sequence of innovations which are stacked as V_t, each with mean square error F_t and Kalman gain K_t. They are unchanged from the old Kalman filter because the dynamics of the model are not altered and so do not have to be recomputed. Having run both y_t and $-X_t$ through the filter, the estimate of β is then b_t. This follows from the standard theory of generalized least squares as v_t and V_t are the recursive residuals from the two fits. It is useful to note that if y_t and β are univariate then all of these recursions, except for $A_{t+1|t}$, will be scalar. As $A_{t+1|t}$ is a vector and is formed by multiplying sparse matrices and vectors it will not take long to run: running through this filter is computationally trivial. Here only V_t will need to be stored for later use by the auxiliary smoother.

On its own, this result allows the efficient computation of the t-statistic for β, for testing whether the intervention is significant. This was mentioned in Section 3 and will be discussed at some length next.

The estimates from the old Kalman filter $a_{t+1|t}$ and $P_{t+1|t}$ can now be corrected for the effect of introducing the regression variables. The new values will be indicated by *. They take on the form

$$a_{t+1|t}^* = a_{t+1|t} + A_{t+1|t}b_t, \qquad P_{t+1|t}^* = P_{t+1|t} - A_{t+1|t}S_tA_{t+1|t}',$$
$$t = 1, \dots, n.$$

These quantities will be useful later when we discuss the fragility of forecasts, but for now we turn our attention to correcting the disturbance smoother.

The second part of de Jong's recursions, the auxiliary smoother, has the form

$$E_t = F_t^{-1}V_t - K_t'R_t, \quad R_{t-1} = Z_t'E_t + T_t'R_t, \quad t = n, \dots, 1$$

starting with $R_n = 0$. Again E_t and R_{t-1} have a simple intuitive interpretation. H_tE_t and Q_tR_{t-1} are the smoothed estimates of ε_t and η_{t-1} using $-X_t$ as the observations.

If, in an obvious notation, we write the new smoothing quantities for the model which includes the explanatory variables as e_t^*, R_t^*, D_t^* and N_t^* then

$$e_t^* = e_t + E_tb_n, \qquad r_{t-1}^* = r_{t-1} + R_{t-1}b_n,$$
$$D_t^* = D_t - E_tS_n^{-1}E_t', \quad N_{t-1}^* = N_{t-1} - R_{t-1}S_n^{-1}R_{t-1}'.$$

These simple expressions mean that the score for the marginal-log-likelihood when X_t is the jth intervention at time t is

$$\frac{\partial \log L_{(t,j)}(\theta)}{\partial\theta_i} = \frac{1}{2}\sum_{t=1}^{n} \operatorname{tr}\left((e_t^* e_t^{*\prime} - D_t^*)\frac{\partial H_t}{\partial\theta_i}\right) + \frac{1}{2}\sum_{t=1}^{n} \operatorname{tr}\left((r_{t-1}^* r_{t-1}^{*\prime} - N_{t-1}^*)\frac{\partial Q_t}{\partial\theta_i}\right),$$

which implies

$$\begin{aligned}\frac{\partial \log L_{(t,j)}(\theta)}{\partial\theta_i} - \frac{\partial \log L(\theta)}{\partial\theta_i} &= \frac{1}{2}\sum_{t=1}^{n} \operatorname{tr}\left(E_t(b_n b_n' - S_n^{-1})E_t'\frac{\partial H_t}{\partial\theta_i}\right)\\ &+ \frac{1}{2}\sum_{t=1}^{n} \operatorname{tr}\left(R_{t-1}(b_n b_n' - S_n^{-1})R_{t-1}'\frac{\partial Q_t}{\partial\theta_i}\right)\\ &+ \sum_{t=1}^{n} \operatorname{tr}\left(e_t b_n' E_t'\frac{\partial H_t}{\partial\theta_i}\right)\\ &+ \sum_{t=1}^{n} \operatorname{tr}\left(r_{t-1} b_n' R_{t-1}'\frac{\partial Q_t}{\partial\theta_i}\right). \qquad (13)\end{aligned}$$

To gain some insight into the meaning of this expression it is useful to return to y_t being univariate and β being scalar. Then $t_\beta = b_n/S_n^{-1/2}$ is the t-statistic for the significance of the intervention which is being assessed. The change in the score (13) becomes

$$\frac{b_n^2 - S_n^{-1}}{2}\sum_{t=1}^{n}\left(E_t^2\frac{\partial H_t}{\partial\theta_i} + \operatorname{tr}\left(R_{t-1}R_{t-1}'\frac{\partial Q_t}{\partial\theta_i}\right)\right) + b_n\sum_{t=1}^{n}\left(e_t E_t\frac{\partial H_t}{\partial\theta_i} + \operatorname{tr}\left(r_{t-1}R_{t-1}'\frac{\partial Q_t}{\partial\theta_i}\right)\right).$$

There are a number of general points to be made about this expression. First,

$$\sum_{t=1}^{n}\left(E_t^2\frac{\partial H_t}{\partial\theta_i} + \operatorname{tr}\left(R_{t-1}R_{t-1}'\frac{\partial Q_t}{\partial\theta_i}\right)\right)$$

does not depend on the data, only on the form or 'design' of the intervention. Hence, the first part of the expression is a scaled and relocated version of the squared t-statistic.

The second term is rather different. Although it is a multiple of the t-statistic,

$$\sum_{t=1}^{n}\left(e_t E_t\frac{\partial H_t}{\partial\theta_i} + \operatorname{tr}\left(r_{t-1}R_{t-1}'\frac{\partial H_t}{\partial\theta_i}\right)\right)$$

depends on the data and the form of the intervention. For time-invariant models, this is a function of $\sum e_t E_t$ and $\sum r_{t-1} R_{t-1}$.

To understand these expressions in more detail we focus on time invariant models. Further, the scalar H will depend only on a single parameter which we take as θ_1. All the other elements of θ will be related to Q. We will assume Q is diagonal and that θ_{k+1} relates only to the (k)th element of the diagonal. This setup will be true in all our applied work. If, for a moment, we focus on elements of θ which correspond to H then

$$\frac{\partial \log L_{(t,j)}(\theta)}{\partial \theta_1} - \frac{\partial \log L(\theta)}{\partial \theta_1} = \frac{\partial H}{\partial \theta_1}\left(\frac{t_\beta^2 - 1 + 2t_\beta t_\beta^*}{2}\right) S_n^{-1} S_E,$$

where $t_\beta^* = (\sum e_t E_t / \sum E_t^2)/S_n^{-1/2}$ and $S_E = \sum_{t=1}^n E_t^2$. Of course t_β^* is a scaled version of the regression coefficient of e_t on E_t, that is of the old smoothed estimate of ε_t on the new one when the intervention is included in the model. If the intervention has little effect the change in the score will be just $(t_\beta - 1)^2$ as t_β^* should be close to one. When the effect is more pronounced, the score statistic may be far from this value.

In the same manner

$$\frac{\partial \log L_{(t,j)}(\theta)}{\partial \theta_{k+1}} - \frac{\partial \log L(\theta)}{\partial \theta_{k+1}} = \frac{\partial Q_{(k)}}{\partial \theta_{k+1}}\left(\frac{t_\beta^2 - 1 + 2t_\beta t_{\beta(k)}^+}{2}\right) S_n^{-1} S_{R(k)}$$

gives the corresponding score for the Q parameters. Here we have defined

$$S_{R(k)} = (\textstyle\sum R_{t-1} R'_{t-1})_{(k,k)},\ t_{\beta(k)}^+ = \dfrac{(\sum r_{t-1} R'_{t-1})_{(k,k)}}{S_{R(k)}} / S_n^{-1/2}$$

In these expressions, r_{t-1} and R_{t-1} represent scaled smoothed estimates of η_t before and after the intervention. The same logic applies to them, and their effect, as applied above for e_t and E_t.

4.2. *One-step re-estimation*

The central theme of this paper is the assessment of the fragility, or lack of robustness, of the inferences from the data analysis to small changes in the specification of the model. In time-series modelling these will be mainly determined by the value of the estimated parameters H_t and Q_t, for they control the rate of discounting and so influence forecasting, seasonal adjustment, detrending, etc. To focus attention on these parameters we argued in Section 3 that we transform from the scores into one-step scoring estimates of θ. If we write $\jmath(\theta)$ to denote the observed information matrix at θ, then we will be interested in the distance

$$\tilde{\theta}_{(t,j)} - \hat{\theta} = \jmath^{-1}(\hat{\theta}) \frac{\partial \log L_{(t,j)}(\hat{\theta})}{\partial \theta}. \tag{14}$$

or more precisely its scaled analogue

$$d^i_{(t,j)} = \frac{\tilde{\theta}^i_{(t,j)} - \hat{\theta}^i}{\hat{J}^{-(1/2)ii}(\hat{\theta})}, \tag{15}$$

where the superscripts denote elements of vectors and matrices. Recall that $\hat{J}^{-1}(\hat{\theta})$ will be a byproduct from an optimizing algorithm for finding $\hat{\theta}$ and so its presence does not add much to the complexity of our diagnostic statistics. Our intention is to graph the elements of $d_{(t,j)}$ against t, for each value of j which indexes the type of intervention, and discuss their usefulness in assessing fragility.

4.2.1. Example. Latin-American exports

To see the form these diagnostic quantities take, we re-examine the data on US exports to Latin-America discussed in Burman (1985), Bruce and Martin (1989) and Harvey and Koopman (1992). To make the problem harder we follow the last of these authors and work at the quarterly, rather than monthly, level of aggregation – leaving the data ranging from 66Q1 to 83Q4. The logarithms of the raw series are plotted in Fig. 2a.

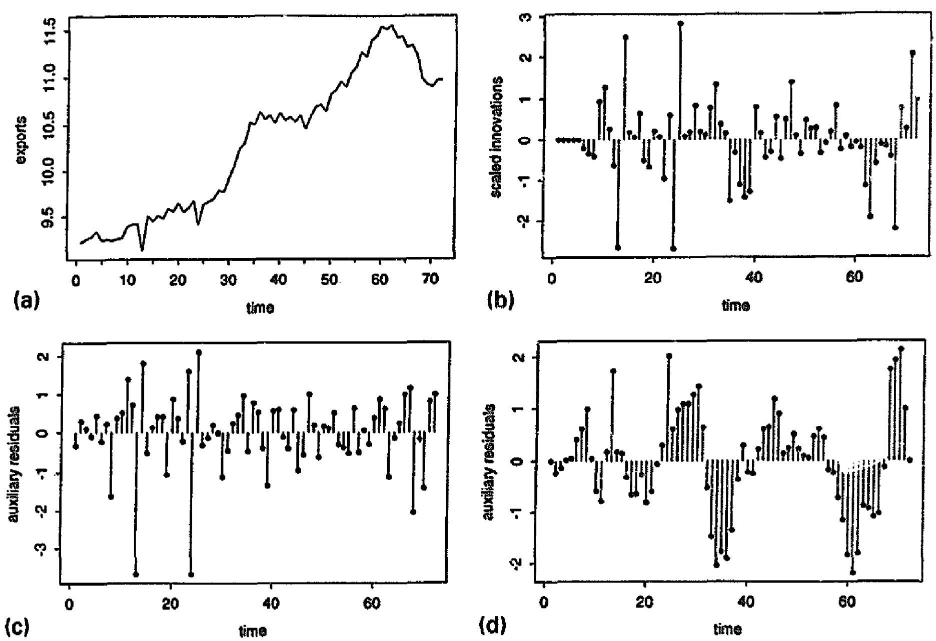

Fig. 2. Latin-American exports: (a) data; (b) innovations from the fitted model; (c) auxiliary residuals for the irregular component; (d) auxiliary residuals for a slope change.

A natural model to apply to these quarterly data is the local linear trend with additive evolving seasonal component. However, when this model is estimated, the seasonal component turns out to be fixed and the variance of the level component is estimated as zero. The local-linear trend model thus reduces to a smooth trend model

$$y_t = \mu_t + \gamma_t + \varepsilon_t, \varepsilon_t \sim NID(0, \sigma_\varepsilon^2),$$

$$\mu_t = \mu_{t-1} + \beta_{t-1},$$

$$\beta_t = \beta_{t-1} + \zeta_t, \zeta_t \sim NID(0, \sigma_\zeta^2),$$

where γ_t is understood to mean a fixed seasonal pattern.

This smoothed trend model is used in a variety of contexts by Theil and Wage (1964), Kitagawa (1981), Gersch and Kitagawa (1983) and Ng and Young (1990) and is discussed in Harvey (1989 p. 286). Wecker and Ansley (1983) show that it corresponds to fitting a cubic spline.

The estimated parameters and standard diagnostics are

$$\widehat{\log} \sigma_\varepsilon = -2.88, \qquad \widehat{\log} \sigma_\zeta = -3.49,$$

$$\hat{\jmath}^{-1} = \begin{pmatrix} 0.012 & -0.0038 \\ -0.0038 & 0.036 \end{pmatrix},$$

$$S = -0.22, K = 4.75, \qquad N(\chi_2^2) = 9.50, \quad Q_{10}(\chi_8^2) = 6.9. \tag{16}$$

Notice that we have chosen to take the logarithms of the standard deviations to form the parameterization $\theta_1 = \log \sigma_\varepsilon$ and $\theta_2 = \log \sigma_\zeta$. This is to improve the sampling behaviour of $\hat{\theta}$ and to avoid negative variances in the one-step procedure (14). We use the notation $\widehat{\log} \sigma$ in preference to $\log \hat{\sigma}$ to emphasize that, in (16), $\hat{\jmath}$ is the information for θ rather than for σ_ε and σ_ζ. Also in (16) the aggregate statistics S and K are the estimated third and fourth moments measuring the skewness and kurtosis of the innovations and N is the normality test due to Bowmans and Shenton (1975). These statistics, often used to assess if there are any outliers in the model, are based on the scaled innovations displayed in Fig. 2b. It is this graph which most time-series modellers use to assess whether there are any problems with their fitted model. For these data the Box–Ljung Q-statistic (Ljung and Box, 1978) takes on a satisfactory value. There does not seem to be any systematic departure of the observed series from the model, such as would be indicated by residual correlation.

An interesting feature of the fitted model is that the observed information matrix suggests that $\hat{\sigma}_\zeta^2$ is quite poorly determined (note that the usual t-statistic to test if $\sigma_\zeta^2 = 0$ is not appropriate here – see Harvey (1989, Section 5.1.2) and Shephard (1993a). This is not surprising for there is a very small sample size. Another interesting feature of $\hat{\jmath}^{-1}(\hat{\theta})$ is that it is quite near to being diagonal, which means the transformation from the scores to the one-step re-estimated parameters is not very important here. In the next section we will see examples where this is not true: use of the parameters, rather than the scores, leads to appreciably sharper inferences.

The analysis of Fig. 2b is not straightforward for anyone but a very experienced time-series modeller. The negative outliers at $t = 13$ and 24, which are quite clearly additive outliers from Fig. 2a, appear as negative shocks. The adjustments of the fitted model to these shocks in both cases causes equally large positive shocks at the next time period – indeed, the largest absolute-scaled innovation is at time 25! This confusing type of shape is even more obscure when there is a slope change. Candidate times for a slope shock are around 35, 45, 60 and 70. These are shown by clusters of negative innovations from 35 to 39 and 61 to 67 and positive ones from 69 to the end of the sample. The possible slope change at time 45 is not at all clear from Fig. 2b.

Although it is possible to glean a great deal of information from Fig. 2b, it takes a good deal of experience to be able to exploit it to the full. The problems of looking at innovation plots become much more severe when more involved models are fitted, such as some of those looked at in Section 6.

Harvey and Koopman (1992) suggest a plot which does aid the modeller, indicating ways of removing some of the difficulties of interpretation of plots such as Fig. 2b. Their idea is to plot the smoothed estimates of ε_t and ζ_t. They call these the auxiliary residuals. They are shown for the Latin-American export series in Figs. 2c and d. It is certainly the case that Fig. 2c shows the additive outliers in a clearer way. This would allow the modeller to include simple dummy variables to take out these problem observations. Interpretation of Fig. 2d to yield information about shocks to the slope is more difficult. The problem is that the smoothed series for ζ_t is much more serially correlated than that for ε_t: Harvey and Koopman (1992) prove that these smoothed values follow an ARIMA(2,0,0) process if the data actually follow the model we have fitted. It is not clear from Fig. 2d where the most effective site for a slope intervention would be – the series is so smooth that precise location of the shock is difficult. A further problem is that it is not immediately clear how these smoothed values related to the way θ is estimated. Hence, we may remove a point with high values of auxiliary residuals to find it has little effect on the estimated parameters.

In contrast, our analysis focuses on the fragility of the estimated results to changes in the fitted model and we measure this by looking at $d_{(t,j)}$ given by (15). The departures, or interventions, will be irregular shocks and slope breaks. These will be fitted for every possible time period, that is $2n$ different explanatory variables will be used. For each of these we record the one-step movements in the parameter estimates. These are displayed for the export data in Fig. 3. The figures give results for both types of intervention and both elements of $d_{(t,j)}$. We call the plots of the parameter movements for $\log \sigma_\varepsilon$ against the slope changes and $\log \sigma_\eta$ against shocks, 'cross plots'.

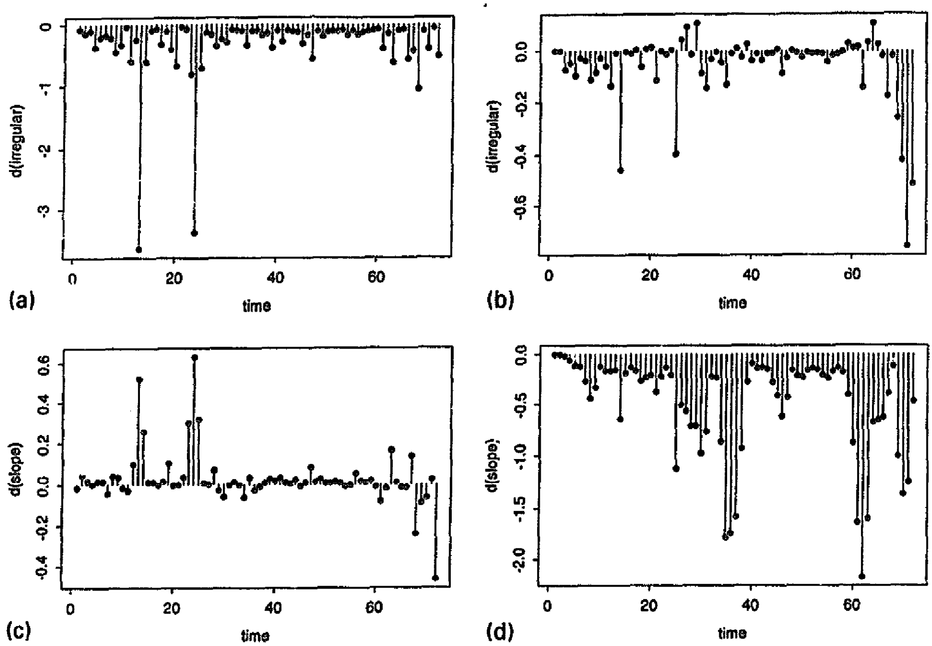

Fig. 3. Latin-American exports: (a) $d(1, t)$, the scaled estimated change in the irregular parameter, for the series of outlier interventions; (b) $d(1, t)$ for slope interventions – a 'cross' plot; (c) $d(3, t)$, the scaled estimated change in the slope parameter for the outlier intervention – another 'cross' plot; (d) $d(3, t)$ for the slope intervention.

For the moment focus on Figs. 3a and d. Fig. 3a assesses the one-step movement in the parameter associated with σ_ε^2, the measurement error variance, as a result of fitting an additive outlier intervention. It suggests strongly that the 13th and 24th observations are indeed the important additive outliers.

Likewise, Fig. 3d assesses the one-step movement associated with σ_ζ^2 and indicates that the most important slope breaks probably take place near observations 35–37 and 61–63. It is difficult to time these changes more precisely – external information would have to be used for this purpose. The other possible slope break takes place right at the end of the sample and will thus be very important when assessing the forecasts from the fitted model.

The unscaled differences (14) behind the relocated and scaled quantities shown in Fig. 3a, the values of which are not given here, imply that if an additive outlier regressor is added to the model at observation 13 then the one-step re-estimated parameter will move from $\widehat{\log\sigma_\varepsilon} = -2.88$ to $\log\widetilde{\sigma_{\varepsilon(13,1)}} = -3.28$, where the notation indicates both observation number, 13, and intervention type: 1 denotes the additive outlier. In contrast, the cross-plot Fig. 3c gives $\log\widetilde{\sigma_{\zeta(13,1)}} = -3.49$, suggesting no move at all. When we re-estimate the model using this intervention we actually achieve

$$\begin{pmatrix}\log\widetilde{\sigma_{\varepsilon_{(13.1)}}}\\ \log\widetilde{\sigma_{\zeta_{(13.1)}}}\end{pmatrix}=\begin{pmatrix}-3.28\\ -3.49\end{pmatrix},\qquad \begin{pmatrix}\widehat{\log\sigma_{\varepsilon_{(13.1)}}}\\ \widehat{\log\sigma_{\zeta_{(13.1)}}}\end{pmatrix}=\begin{pmatrix}-3.07\\ -3.46\end{pmatrix}$$

which shows the one-step procedure has overestimated the change slightly but has picked up the most important features of this outlier – that it affects the estimate of σ_ε^2, but not that of σ_ζ^2.

The other important additive outlier seems to be at observation 24. The corresponding numbers are

$$\begin{pmatrix}\log\widetilde{\sigma_{\varepsilon_{(24.1)}}}\\ \log\widetilde{\sigma_{\zeta_{(24.1)}}}\end{pmatrix}=\begin{pmatrix}-3.25\\ -3.49\end{pmatrix},\qquad \begin{pmatrix}\widehat{\log\sigma_{\varepsilon_{(24.1)}}}\\ \widehat{\log\sigma_{\zeta_{(24.1)}}}\end{pmatrix}=\begin{pmatrix}-3.09\\ -3.42\end{pmatrix}.$$

The one-step procedure predicts a slightly smaller decrease in the estimate of σ_ε^2 than is associated with the 13th observation: however, the values of $\hat{\sigma}_\varepsilon$ are virtually identical in the two cases. Thus, the procedure is able to pick up the major features of the effect of the shock. Although this may not be entirely satisfactory for some purposes it is probably satisfactory for a diagnostic statistic. Indeed, over-estimation of the effect of deletion of observations is a standard feature of diagnostics for non-linear models. Examples for the shifted power transformation, given by Atkinson (1986), show that interesting observations are correctly identified, even if the precise values of deletion statistics are not given.

If both additive outliers are removed then

$$\widehat{\log\sigma_\varepsilon} = -3.40, \qquad \widehat{\log\sigma_\zeta} = -3.43,$$

$$\hat{J}^{-1} = \begin{pmatrix}0.0105 & -0.00348\\ -0.00348 & 0.0403\end{pmatrix},$$

$$S = -0.34, \qquad K = 4.16, \qquad N(\chi_2^2) = 5.32, \quad Q_{10}(\chi_8^2) = 8.06. \tag{17}$$

Comparison of (17) with (16) shows that the model is now more tightly specified, in that the signal-to-noise ratio has risen considerably and that there is no longer significant evidence against normality.

As we have mentioned, the location and importance of a slope intervention is much less clear cut than that of the additive outlier. Fig. 3d indicates three candidate locations. The first is between observations 35 and 37, the second between 61 and 63 and the third at 70. To investigate the properties of our method, we proceed by fitting all three in turn – putting the slope investigations at 36, 62 and 70.

We find that

$$\begin{pmatrix}\widetilde{\log\sigma}_{\varepsilon(36,3)}\\ \widetilde{\log\sigma}_{\zeta(36,3)}\end{pmatrix}=\begin{pmatrix}-2.87\\ -3.83\end{pmatrix},\qquad \begin{pmatrix}\widehat{\log\sigma}_{\varepsilon(36,3)}\\ \widehat{\log\sigma}_{\zeta(36,3)}\end{pmatrix}=\begin{pmatrix}-2.88\\ -3.46\end{pmatrix},$$

$$\begin{pmatrix}\widetilde{\log\sigma}_{\varepsilon(62,3)}\\ \widetilde{\log\sigma}_{\zeta(62,3)}\end{pmatrix}=\begin{pmatrix}-2.88\\ -3.90\end{pmatrix},\qquad \begin{pmatrix}\widehat{\log\sigma}_{\varepsilon(62,3)}\\ \widehat{\log\sigma}_{\zeta(62,3)}\end{pmatrix}=\begin{pmatrix}-2.86\\ -3.61\end{pmatrix},$$

$$\begin{pmatrix}\widetilde{\log\sigma}_{\varepsilon(70,3)}\\ \widetilde{\log\sigma}_{\zeta(70,3)}\end{pmatrix}=\begin{pmatrix}-2.89\\ -3.75\end{pmatrix},\qquad \begin{pmatrix}\widehat{\log\sigma}_{\varepsilon(70,3)}\\ \widehat{\log\sigma}_{\zeta(70,3)}\end{pmatrix}=\begin{pmatrix}-2.89\\ -3.54\end{pmatrix},$$

where the estimates are now for the third kind of intervention.

These numbers suggest that the problem of overestimation which we encountered for the additive outliers is more of a problem here: in particular, the effect of observation 36 is indicated as appreciable by the one-step method but is shown to be negligible when the parameter is re-estimated. However, the two other cases are more suggestive. Importantly, we again find that the σ_ε^2 parameter is not predicted to move significantly and that this prediction turns out to be accurate. This suggests our procedure may give a clearer picture of the difference between additive outliers and breaks than do previous methods.

The analysis of the graphs for the slope change suggests that there is only tentative evidence to suggest that the estimator of $\log\sigma_\zeta$ is fragile. We have attempted to find places where slope breaks are likely to change this estimator, but it would seem that the changes which do occur are only small.

Perhaps this is not surprising. The series has an appreciable number of periods where the slope seems to change, so a stochastic slope model is appropriate: modifying the model to remove one of these periods should not change the parameter estimates too dramatically. Of course, things would be rather different if the series had one or two periods where sharp changes in slope dominated changes at all the other times in the sample.

4.3. Relationship with t-statistics

In Section 4.1 we derived a relationship between the change in score due to the intervention and the t-statistic t_β for that intervention. This t-statistic can be calculated at each time point for each intervention, as is the scaled distance given by (15), and so can provide an alternative diagnostic procedure. In this section we give some numerical results of a comparison between these alternatives, using the Latin-American export data.

The t-statistic was written in Section 4.1 as $t_\beta = b_n/S_n^{1/2}$. There are several versions which might be of interest depending on how S_n is estimated. We consider two. The first uses the global estimate of S_n, not adjusted in any way for the effect of the intervention. In the second the variances comprising S_n are adjusted using the one-step correction given in (14). In Section 4.1 we showed that, when the intervention has little effect, the change in the score will be approximately $-(t_\beta - 1)^2$. The t-statistic here could be written more fully as $t_\beta(t,j)$ to stress dependence on both the time and nature of the intervention. Then the comparisons take the form of index plots of the scaled distances $d^i_{(t,j)}$ (15) and of $-\{t_\beta(t,j) - 1\}^2$. A slight complication is that the scales of the two quantities are different. We scale the plots of the distances by the standard deviation of the distances estimated from the n values in the plot. Both sets of values of $-(t_\beta - 1)^2$ are scaled by the estimated standard deviation for the statistic using adjusted estimates of the variances. This choice exhibits the changes due to the effect of the intervention on the estimation of S_n, the numerator of the statistic b_n being the same in both plots.

Fig. 4a reproduces Fig. 3a with the addition of the index plot of $-(t_\beta - 1)^2$ for the unadjusted statistic. Although the pattern is similar, the values of our distances for the two outliers are greater than those of the function of the unadjusted t-statistic. However, as Fig. 4b shows, use of our one-step approximation to the change in variance leads to a function of the adjusted t-statistic

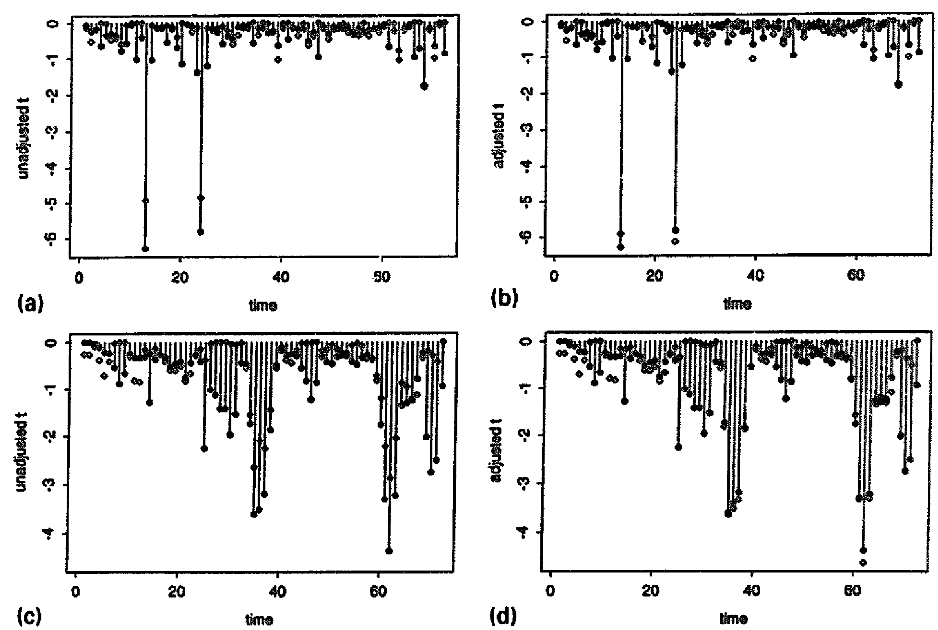

Fig. 4. Latin-American exports: (a)(●) $d(1,t)$, (◇) $-(t_\beta - 1)^2$ from the unadjusted t-statistic for the series of outlier interventions; (b) as (a) but with one-step adjustment of the parameters in the t-statistic; (c)(●), $d(3,t)$, (◇) $-(t_\beta - 1)^2$ from the unadjusted t-statistic for the series of slope interventions; (d) as (c) but with one-step adjustment of the parameters in the t-statistic.

which is similar to our distance. A similar pattern is shown for the slope intervention in Figs. 4c and d. Again, for the larger values, adjustment leads to a function of the t-statistic which behaves very like our distance. There are however some interesting differences, both around observation 30 and at the end of the series, where our distance seems to pick up the change in slope visible in Fig. 2a which is not detected by the t-statistic.

Figures such as these cannot provide a complete comparison of the power of the two procedures. It is however clear that, because of the scaling of the n values, a procedure with low power would tend to have fewer extreme values, whether large or small, than a procedure with higher power, an effect which is visible in the plots of the adjusted and unadjusted statistics. The comparison raises questions about the t-statistic in which the elements of S_n are fully re-estimated, rather than through an approximation as here. But these results do show both that the distances we have derived are comparable in performance with the t-statistic and that our derivation leads to a useful one-step procedure for updating estimates of variance. Finally, the notation $t_\beta(t,j)$, with the absence of a superscript i, stresses that the t-statistic solely measures the effect of the jth intervention. There is no source of information like that in the cross-plots of Figs. 3b and c.

5. Simulation of intervention measures

Although the plots of $\tilde{\theta}_{(t,j)} - \hat{\theta}$ are scaled by an estimator of the standard error of $\hat{\theta}^i$, and hence the scales in Fig. 3 are suggestive, the distribution of $d_{(t,j)}$ is unknown. In regression diagnostics a similar type of problem exists when Cook's statistic is used (Atkinson, 1985, p. 25). There, simulation envelopes are often employed to give some impression of likely scatter. Here we follow a similar path.

As outlined in Section 3, we simulate a new time series $y^{(k)} = (y_1^{(k)}, \ldots, y_n^{(k)})'$ using $\hat{\theta}$ as the true value and record the value of $s(\hat{\theta}; y^{(k)})$. If n is large then the log-likelihood should be approximately quadratic. In turn, this means that the distance this new score moves when we change the model is

$$\frac{\partial \log L_{(t,j)}(\hat{\theta}; y^{(k)})}{\partial \theta} - s(\hat{\theta}; y^{(k)})$$

will be an asymptotically valid (as $n \to \infty$) replication from the distribution we require. Hence, we can produce a Monte-Carlo sample of

$$d_{t,j}^{i(k)} = \hat{J}^{(1/2)ii}(\hat{\theta})\hat{J}^{-1}(\hat{\theta})\left(\frac{\partial \log L_{(t,j)}(\hat{\theta}; y^{(k)})}{\partial \theta} - s(\hat{\theta}; y^{(k)})\right),$$

$$\text{where } k = 1, \ldots, K. \tag{18}$$

and so asymptotically appropriate pointwise bounds on the plotted distance measures.

At first sight the computations involved in producing these replications look horrendous. However, as the parameters are always fixed at $\hat{\theta}$, most of the Kalman filter and smoothing algorithms have only to be run once for $P_{t+1|t}$, F_t, K_t, L_t, D_t and N_t do not depend on the data. As these are the most cumbersome parts of the computations these replications can be computed very quickly.

We use the K replications to construct pointwise one-sided confidence intervals. The probability of inclusion in the resulting confidence envelope of all, or all but a specified number, of the vector of measures for the whole sample could be found, if required, by simulation method similar to those of Flack and Flores (1989). One-sided envelopes are appropriate as we introduce new features into the model in order to reduce the variation in, for example, the irregular or slope components. Typically, we will want to display 95 and 99% intervals. Table 1 gives some values of K and of M, the order-statistic corresponding to these two levels. Increasing K (and so M) can be expected to increase the precision of the bounds. The main difficulty with M being large is the requirement to store and sort a large number of replications for each value of t as the simulation goes through 1 to n. If K is small the simulation envelopes for $d^i_{(t,j)}$ will tend to be quite ragged as we run through the index t. We can increase the smoothness of this line by increasing K, but a more sensible procedure is to exploit a smoothing algorithm to iron out some of the kinks.

The simulation envelopes for regression introduced by Atkinson (1981) took K as 19. In a study of mixed model analysis of variance Dempster et al. (1984)

Table 1
Monte-Carlo envelopes: Replication size K and order statistics M. The probabilities are $1 - M/(K+1)$. Numerical results are averages of 100 medians of samples of 60 from the exponential distribution

K	M	Estimated bound
99%		
99	1	4.977
399	4	4.713
999	10	4.636
1999	20	4.628
Population bound		4.605
95%		
99	5	3.041
399	20	3.027
999	50	3.001
1999	100	3.004
Population bound		2.996

took $K = 119$ and compared the results with those for $K = 8000$. Hall and Titterington (1989) give results on properties of Monte-Carlo tests, including dependence on K, for statistics which are asymptotically pivotal. In our case the effect of increasing K and exploiting a smoothing algorithm can be gleaned from Figs. 5 and 6 which display 95% and 99% envelopes for $d^1_{(t,1)}$, equation (18), the statistic which assesses how far the irregular variance moves when an additive outlier regressor is included in the model. Fig. 5 gives the raw envelopes for $K = 99$, 199 and 999. The corresponding envelopes for these values of K using a smoothing algorithm (the S-Plus function 'lowess') are displayed in Fig. 6.

A feature of these smoothed envelopes is that they become narrower as K increases. To investigate this phenomenon a small simulation experiment was

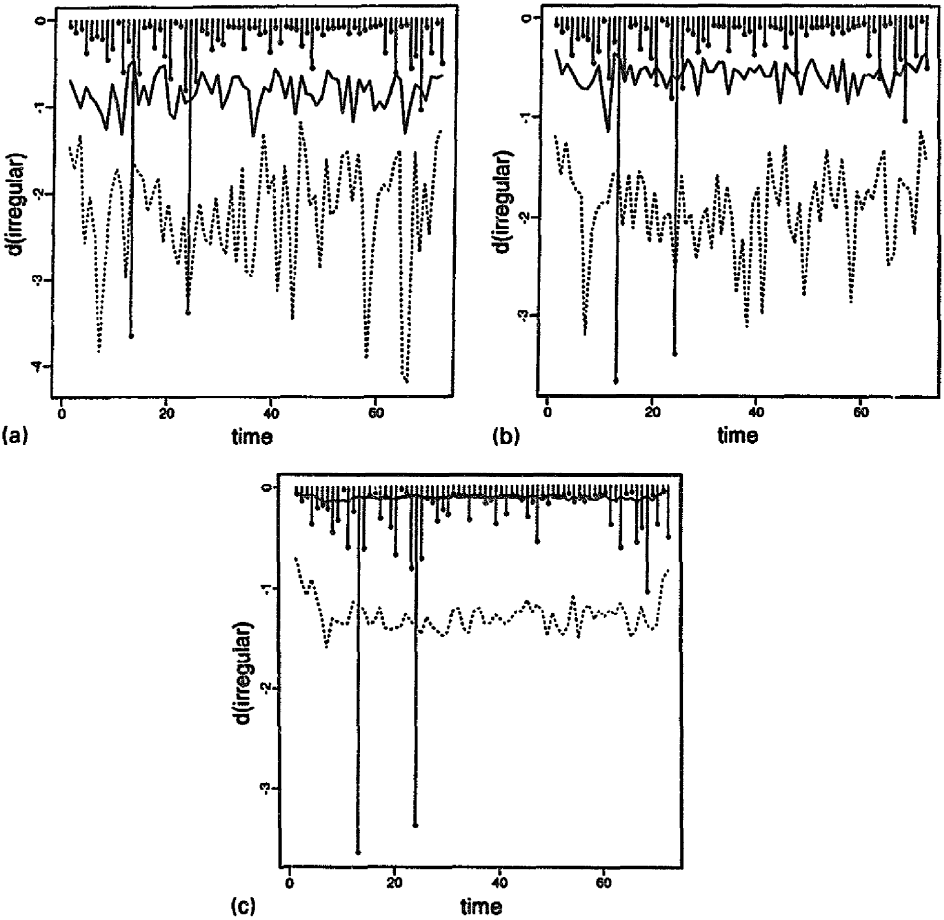

Fig. 5. Latin-American exports: simulation envelopes for the distance $d(1,t)$ with an outlier intervention: (a) $K = 99$; (b) $K = 199$; (c) $K = 999$.

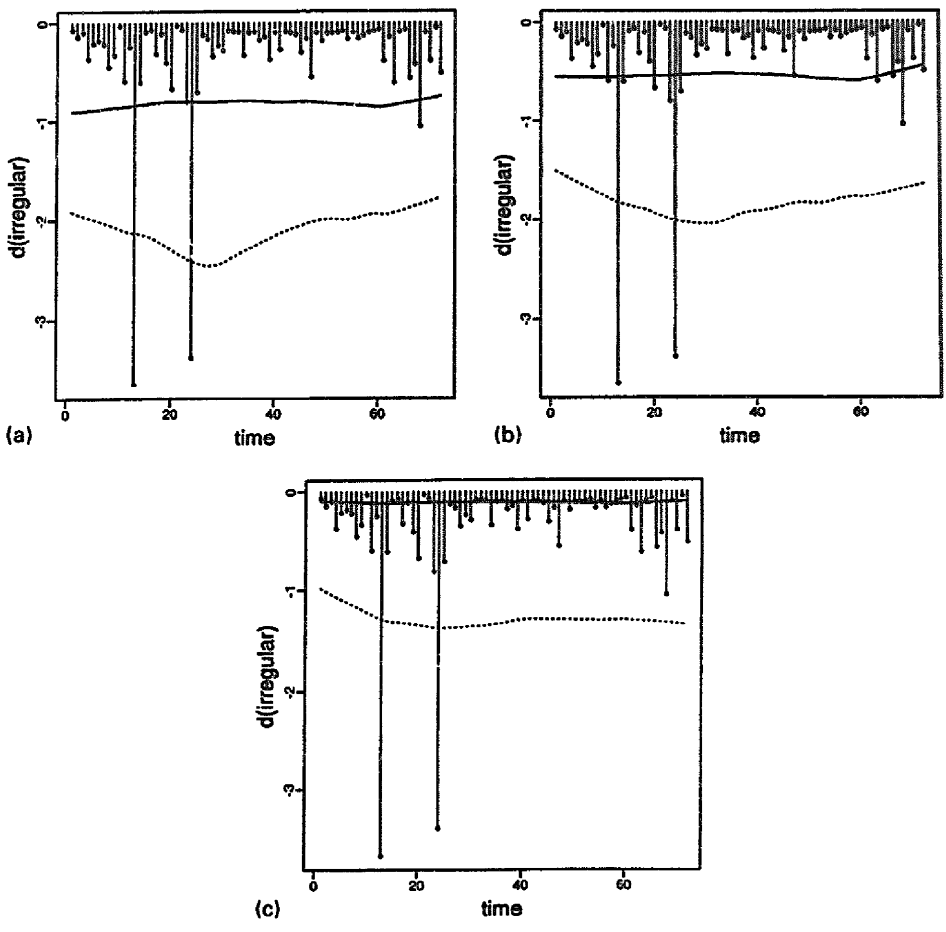

Fig. 6. Latin-American exports: smoothed versions of the envelopes of Fig. 4: (a) $K = 99$; (b) $K = 199$; (c) $K = 999$.

performed, the results of which are summarized in Table 1 and Fig. 7. This phenomenon is not apparent in the results of Dempster et al. (1984) who were sampling from the normal distribution. Since the distribution of the scores is skewed, we sampled from the unit exponential distribution. To mimic the effect of median-based smoothing on a series, 60 independent observations were taken and the median calculated. This was repeated 100 times for eight combinations of K and M. The boxplots in Fig. 7 and the means in Table 1 show that the envelopes can be expected to become less extreme as K increases. The limiting population values are the percentage points of the exponential distribution: log 20 for the 95% limit and log 100 for that at 99%. One other theoretical check on these results is for $K = 99$ and $M = 1$, when the median of the extreme order

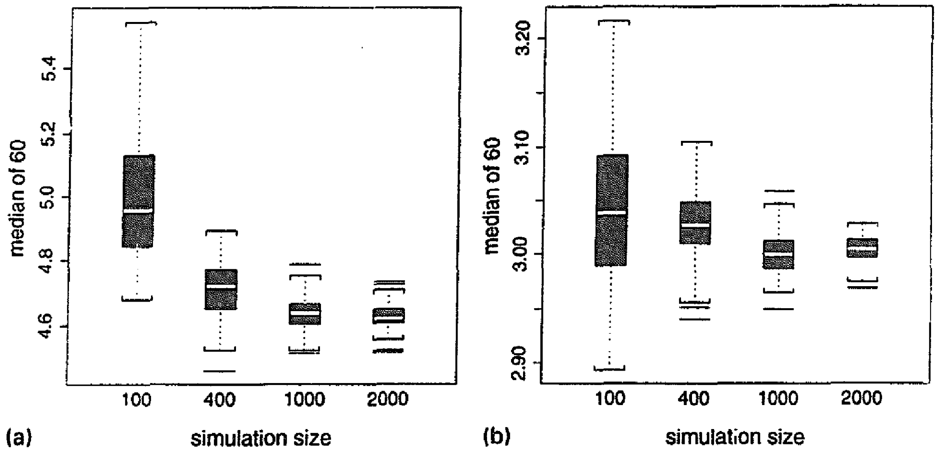

Fig. 7. Dependence of smoothed envelopes on simulation size K: medians of samples of 60 from the exponential distribution: (a) 1% envelope; (b) 5% envelope.

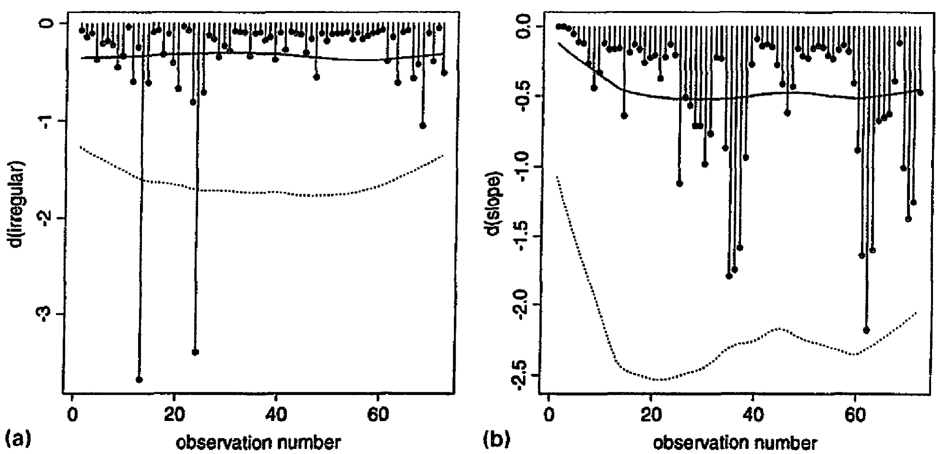

Fig. 8. Latin-American exports: (a) $d(1,t)$ for the outlier intervention; (b) $d(3,t)$ for the slope intervention: the envelopes are for $K = 399$.

statistic is given by $-\log\{1-(0.5)^{1/K}\} = 4.965$, agreeing with the simulated value. Of course, our statistics are not independent and do not have an exponential distribution, nor is smoothing identical with calculating a median. But these results provide some qualitative explanation of the phenomenon. Taken together with envelopes for values of K other than those of Fig. 6, they do suggest that envelopes constructed using 399 replications and smoothing tend to yield most of the important features of the shape of the curves. This is what we will use in the rest of this paper.

Now to consider the use of the envelopes in making inferences, rather than the properties of the envelopes themselves. Fig. 8 shows the raw statistics for both additive outliers and slope breaks along with their smoothed 95 and 99% envelopes. They suggest that the additive outlier points at 13 and 24 are the most important departures from our model. They are a considerable distance outside the 99% envelope. The slope breaks, on the other hand, are much less crucial. Fig. 8b indicates that all the candidate points are within the 99% interval and so it is not obvious that the model will be fragile to changes in slope. As we saw earlier, this was the conclusion we came to when we actually did put slope interventions into the model and re-estimated the parameters.

6. Further illustrations

6.1. Purse data

The monthly numbers of purse, in English handbag, snatchings in Hyde Park, Chicago, given by Reed (1987) are discussed by Harvey (1989, pp. 89–90), who fits a local-level model: the data are given on p. 516. The discrete observations range from 3 to 36, perhaps large enough for use of a continuous transformation to normality. Atkinson and Shephard (1996) model the series after first transforming by taking logarithms. We follow their approach.

The local-level model is $y_t = \mu_t + \varepsilon_t$, $\mu_t = \mu_{t-1} + \eta_t$, with estimated parameters

$$\widehat{\log}\,\sigma_\varepsilon = -0.931, \qquad \widehat{\log}\,\sigma_\eta = -2.11,$$

$$\hat{J}^{-1} = \begin{pmatrix} 0.011 & -0.010 \\ -0.010 & 0.11 \end{pmatrix},$$

while the diagnostics take on the values

$$S = -0.59, \quad K = 3.96, \quad N(\chi_2^2) = 6.80, \quad Q_8(\chi_7^2) = 4.74.$$

The logarithmic transformation is rejected by the normality statistics – the 5% point of χ_2^2 is 5.99. This rejection of normality is mainly caused by the existence of the two additive outliers at times 15 and 60. These are the two smallest readings, both equal to three, which either precede or follow a much larger value. The two outliers are shown distinctly by Fig. 9a which displays the $d_{(t,1)}$, corresponding to the fitting of an additive outlier. The implied one-step

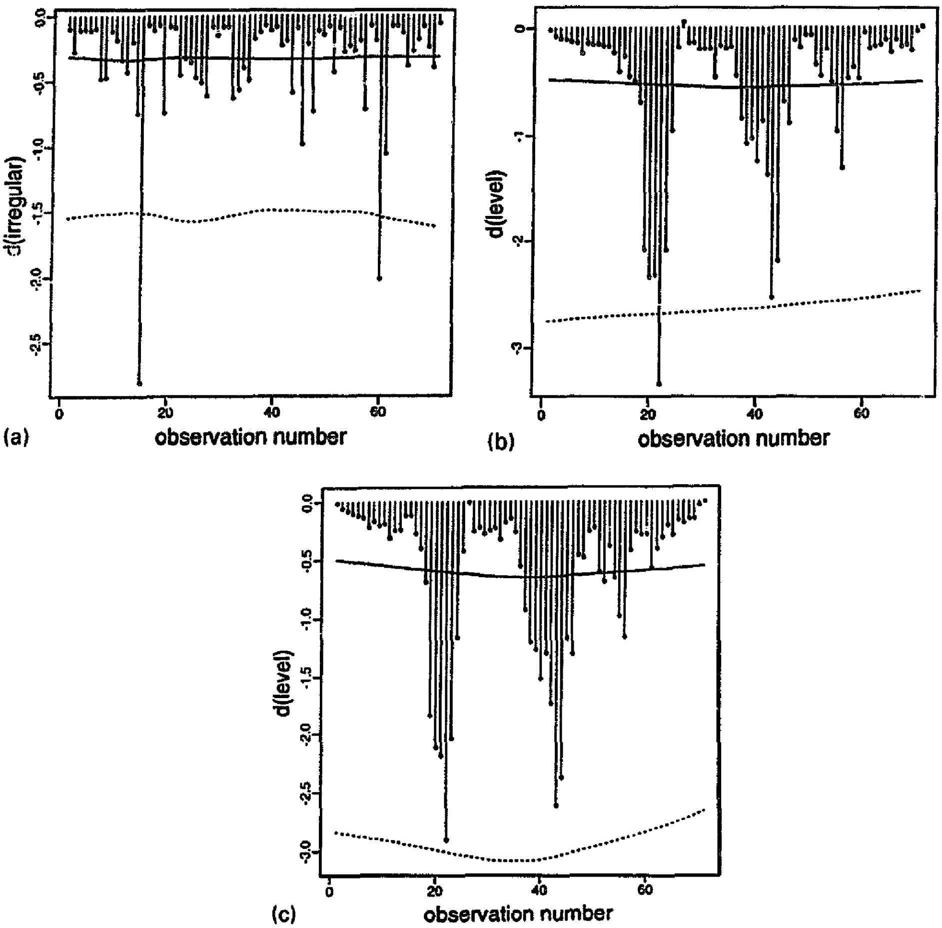

Fig. 9. Purse data: (a) effect of an additive outlier; (b) effect of the intervention for a break in level; (c) $d(2, t)$ for the level break when the two additive outliers are accommodated.

and fully iterated parameter estimates for these two cases are

$$\begin{pmatrix} \widetilde{\log \sigma_{\varepsilon_{(15,1)}}} \\ \widetilde{\log \sigma_{\eta_{(15,1)}}} \end{pmatrix} = \begin{pmatrix} -1.08 \\ -2.15 \end{pmatrix}, \quad \begin{pmatrix} \widehat{\log \sigma_{\varepsilon_{(15,1)}}} \\ \widehat{\log \sigma_{\eta_{(15,1)}}} \end{pmatrix} = \begin{pmatrix} -1.02 \\ -2.14 \end{pmatrix},$$

$$\begin{pmatrix} \widetilde{\log \sigma_{\varepsilon_{(60,1)}}} \\ \widetilde{\log \sigma_{\eta_{(60,1)}}} \end{pmatrix} = \begin{pmatrix} -1.04 \\ -2.12 \end{pmatrix}, \quad \begin{pmatrix} \widehat{\log \sigma_{\varepsilon_{(60,1)}}} \\ \widehat{\log \sigma_{\eta_{(60,1)}}} \end{pmatrix} = \begin{pmatrix} -1.00 \\ -2.07 \end{pmatrix}.$$

This shows good agreement between the one-step and iterated parameter estimates. As would be expected, the outlier intervention has little effect on σ_η^2 but an appreciable effect on σ_ε^2 corresponding to identification of one of the two additive outliers.

The evidence for a level shift, given in Fig. 9b, is much more subtle, due to potential confounding of additive outliers and level breaks. There seems to be evidence for the fragility of $\widehat{\log}\,\sigma_\eta$ for a level shift intervention may change this value. However, at this stage we must be careful for the undoubted increase in the level of the series has come immediately after the additive outlier. The apparent level break may become less important when we remove the additive outlier at time 15. To check this we refit the model to the data with the two additive outlier regressors included. The new estimates are

$$\widehat{\log}\,\sigma_\varepsilon = -1.11 \qquad \widehat{\log}\,\sigma_\eta = -2.07,$$
$$\hat{J}^{-1} = \begin{pmatrix} 0.008 & -0.011 \\ -0.011 & 0.145 \end{pmatrix},$$
$$S = -0.11, \quad K = 2.85, \quad N(\chi_2^2) = 0.20, \quad Q_{10}(\chi_8^2) = 4.8.$$

There is thus, as would be expected, a further decrease in σ_ε^2. However, there is virtually no change in σ_η^2. Conventional aggregate statistics indicate that this is a good model. But we still need to look for a level change to check the implication of Fig. 9b. The statistics for the level change when the two outliers are removed are given in Fig. 9c. This picture is very interesting. Now the level change is slightly under the 1% bound. More importantly, $d_{(22,2)}$ is no longer solely outstandingly large and so it is unclear whether we should remove this feature from the data. Our own prejudice is to tend to leave these kinds of characteristics in the data and not to remove problem points unless there is substantial evidence for such removal. The reason for this is that we do not want to overmanipulate the data – which is obviously a tempting thing to do in such diagnostic analyses.

6.2. Coal consumption

A series on the quarterly consumption of coal in the UK from 60Q1 to 83Q4 is analysed in Harvey (1989, pp. 2, 96). The coal users are classified as 'other final users', a category including public administration, commerce and agriculture. The series shows a continual fall as the UK diversified its energy sources and a strong seasonal pattern, which seems quite steady. After taking a logarithmic transformation we fitted a model with a stochastic trend, trigonometric seasonal and an additive irregular component. The fitted model has fixed, that is non-stochastic, slope and seasonal terms. The presence of fixed terms is not unusual in applied work and, indeed, arose in the seasonal component of our model for the Latin-American export data. The rest of the model has

$$\widehat{\log}\,\sigma_\varepsilon = -2.19, \qquad \widehat{\log}\,\sigma_\eta = -3.36,$$
$$\hat{J}^{-1} = \begin{pmatrix} 0.009 & -0.004 \\ -0.004 & 0.125 \end{pmatrix},$$
$$S = -0.74, \quad K = 4.45, \quad N(\chi_2^2) = 16.5, \quad Q_{10}(\chi_8^2) = 6.11.$$

We will assess the effect on this fitted model of the addition of level and additive outlier interventions. Figs. 10a and b give our new statistics for these features of the data. From the first of these plots there seems strong evidence for additive outliers at 37 and 39. The movements around the 60th observation, the 74Q4 value, are a little more difficult to analyse. There seems substantial evidence for a level break at this time – not only from Fig. 10b but visual inspection of the raw data is suggestive of this. However, there may also be some additive outliers around this point, which corresponds to a major strike in the British coal mining industry and the resultant turbulence of the re-stocking period. Our modelling approach is to deal first with major difficulties, leaving the details until later. As a result we refit the model taking out the additive outliers at observations 37 and 39 and fitting a level change at 60. The corresponding new parameter estimates and aggregate diagnostics are

$$\widehat{\log \sigma_\varepsilon} = -2.28, \qquad \widehat{\log \sigma_\eta} = -3.43,$$

$$\hat{J}^{-1} = \begin{pmatrix} 0.008 & -0.0005 \\ -0.0005 & 0.113 \end{pmatrix},$$

$$S = -0.0366, \quad K = 0.128, \quad N(\chi_2^2) = 0.1645, \quad Q_{10}(\chi_8^2) = 6.52.$$

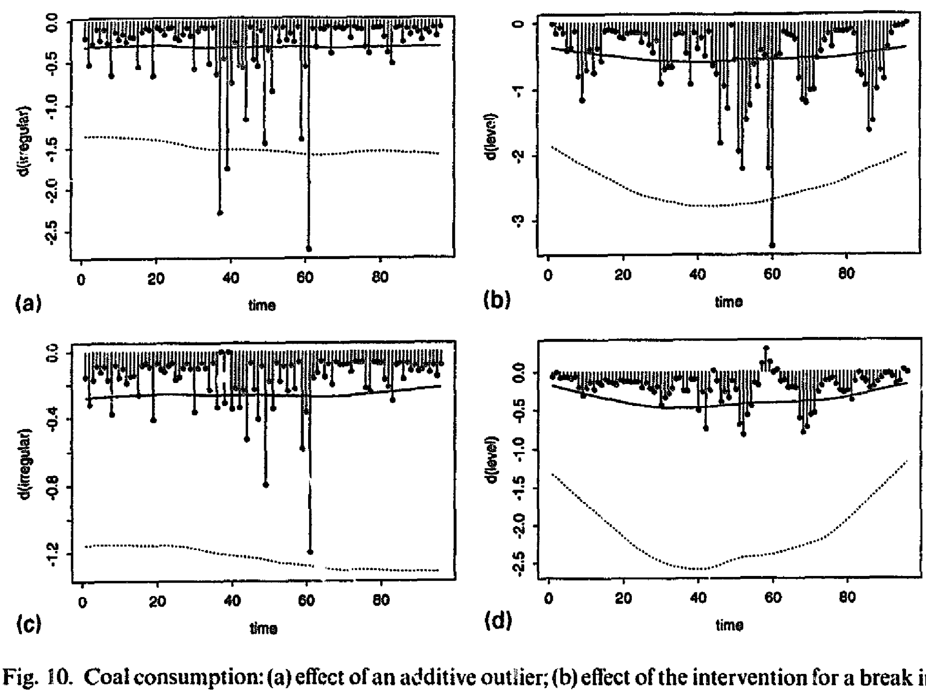

Fig. 10. Coal consumption: (a) effect of an additive outlier; (b) effect of the intervention for a break in level; (c) effect of a further additive outlier and (d) effect of a second break in level when two outlier interventions and one level break are already included in the model.

The addition of these three interventions thus produces a model with similar parameter values which however now satisfies the normality tests. In particular, the value of the Bowman–Shenton statistic is reduced to one-tenth of its previously significant value. To check the details we recalculate the two diagnostic plots as shown in Figs. 10c and d. The plot for the level break is transformed dramatically – it is now virtually structure free: addition of a second change in level close to observation 60 causes a slight increase in the estimated parameter, which is possible since we are using a marginal likelihood. Otherwise there is a negligible decrease. In Fig. 10c there is still some evidence of an additive outlier at observation 61, but the significance of the evidence is appreciably reduced from that shown in Fig. 10a before the level intervention was introduced.

6.3. *The problem of the Nile*

Cobb (1978) gives a series of readings on the annual volume of discharge from the Nile River at Aswan, Egypt for the years 1871–1970. He uses the data to illustrate a conditional technique for inference about a change point. Later analyses include those of Carlstein (1988) and of Balke (1993). All three conclude that a permanent decline in volume has taken place from 1899 onwards. It is surprising that none of them detect the outlier for 1913, observation 43, which is clearly visible from the time-series plot Fig. 11a.

For the moment we disregard the level shift and outlier and fit an unobserved component model with a stochastic level and an irregular component. The estimated parameters and standard diagnostics are

$$\widehat{\log \sigma_\varepsilon} = 4.81, \qquad \widehat{\log \sigma_\eta} = 3.64,$$

$$\hat{J}^{-1} = \begin{pmatrix} 0.0101 & -0.0263 \\ -0.0263 & 0.188 \end{pmatrix},$$

$$S = -0.07, \quad K = 0.25, \quad N(\chi_2^2) = 0.34, \quad Q_{10}(\chi_8^2) = 13.64. \tag{19}$$

Although the value for Q is a little large, it is not significantly so. Despite the outlier and the omitted level break, the aggregate statistics give no indication of departures from normality. Some time-series analysts would therefore conclude that this was a satisfactory model.

We now apply our diagnostic procedure to these data. Fig. 11b indicates two additive outliers, one at 1887 and the other the already discussed outlier at 1913. Fig. 11c shows very clearly that a level intervention is required between 1897 and 1900, an interval including 1899 which was reported as a level shift in the earlier papers. If these three interventions are included and the model refitted the variance for the stochastic level becomes zero so that the model becomes deterministic with added white noise.

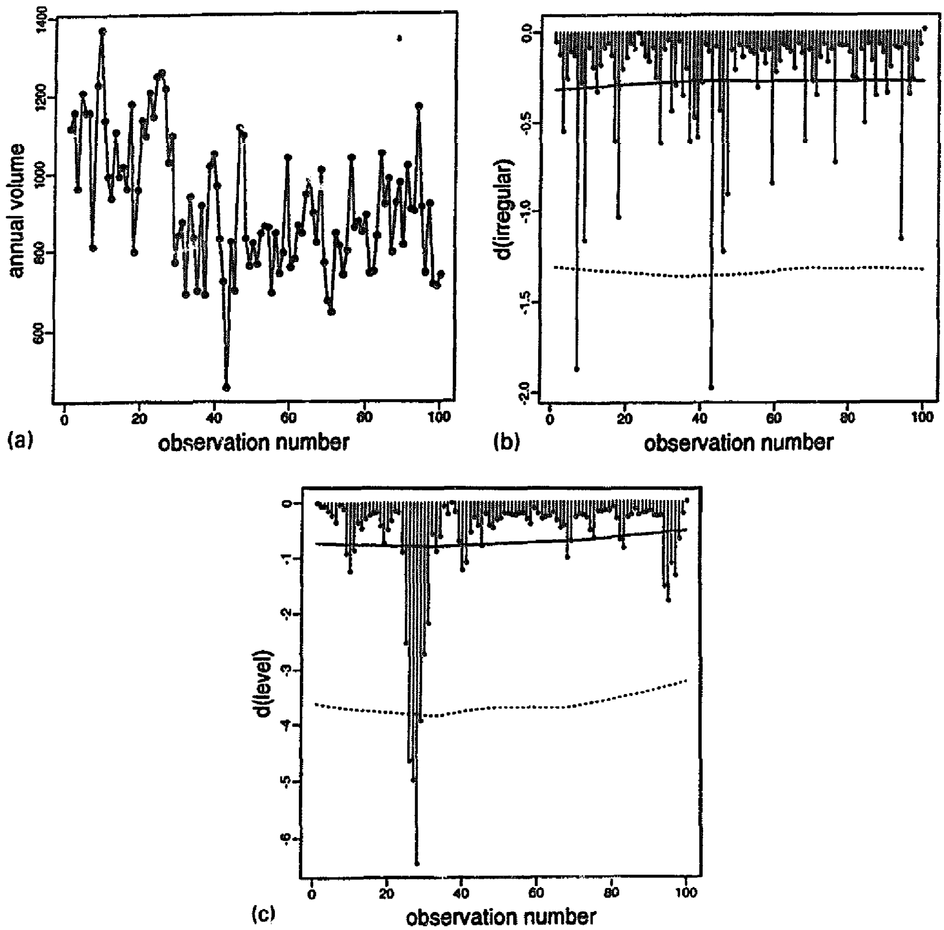

Fig. 11. The problem of the Nile: (a) annual flow at Aswan in cubic kilometres; (b) effect of an additive outlier; (c) effect of the intervention for a break in level.

A survey of analyses of the Nile data is given by MacNeill et al. (1991), who analyse three closely related series using ARMA models. For Cobb's data they also obtain white noise plus a level shift in 1899, but only for data up to 1907. Thereafter they find non-zero serial correlation. However, they fail to identify the outlier at 1913, which may perhaps be interacting with their detrending procedure to give the impression of correlation.

As well as technical statistical matters, MacNeill et al. (1991) give a brief history of the Nile and of the economic and political importance of its waters for irrigation. Their preferred explanation for the level shift around 1899 is that this is the date of the start of construction of the first Aswan dam. The relatively small dam cannot itself have had much effect on the flow of the river but, they argue, activities related to building the dam did lead to improved methods of measuring the flow and so to a reduction in earlier over-optimistic estimates of the annual volume of water.

7. Forecasting

Much of time-series modelling is motivated by the desire to produce a forecast of future observations and a measure of their likely precision. In this section we extend the use of the statistics developed and illustrated in the last three sections so that they are fully focused on the task of forecasting. This allows us to assess which features of the model are crucial in determining the forecasts.

Forecasts in state-space models are determined by $a_{n|n}$, while the associated precision is a function of $P_{n|n}$. As we vary the model these terms will change to $a_{n|n(t,j)}$. When the parameter values θ are constant the model changes can be assessed by simply running an auxiliary Kalman filter, Eq. (11). However, this will be a misleading assessment of the effect a change in the model will have on the forecast, for the change will also influence the parameter estimates themselves.

The implication of this discussion is that we define

$$a_{n|n(t,j,\tilde{\theta})} \simeq a_{n|n(t,j,\hat{\theta})} + \frac{\partial a_{n|n}}{\partial \theta}(\hat{\theta} - \tilde{\theta}_{(t,j)}) \tag{20}$$

and then measure the distance

$$a_{n|n} - a_{n|n(t,j,\tilde{\theta})}$$

as a first-order Taylor approximation to the fully iterated movement in the states. It is difficult to make analytic progress on the actual term $\partial a_{n|n}/\partial \theta$, although it is easily obtained by numerical differentiation. Unlike all the other terms in (20) this term does not depend on the type of intervention being used. It thus has to be calculated only once for all assessments of the sensitivity of the forecast. A similar argument can be made for the change in the mean square error of the estimate of α_n, $P_{n|n}$.

The forecasts themselves will be straightforward transformations of $a_{n|n}$

$$\hat{y}_{n+v|n} = Z_{n+v} T_{n+v} T_{n+v-1} \dots T_{n+1} a_{n|n}, \quad v = 1, 2, \dots$$

and so will, in general, depend on the forecast horizon v, although not for a local-level model. The difficulty with this is that we then have to decide what horizons are important so that we can plot them. This will be very dependent on the modelling task. In typical situations we suggest that we should look at both short and long horizons as these will focus on the robustness of both types of forecasts.

Similar types of arguments apply for the measure of accuracy of the forecasts. Here one could see if changing the model allows an important tightening of the model specification, thus constricting the forecast interval. Some of these points will be illustrated by reconsidering the data on the Nile time series.

7.1. Example: the problem of the Nile revisited

Fig. 12a plots $a_{n|n} - a_{n|n(t,j,\theta)}$ for the outlier intervention. Fig. 12b is the same plot for the level break. The discounting of the time-series model means that almost all the activity in the plot is at the end of the series. Figs. 12c and d, on the other hand, give the additional term which reflects the effect the changes in the parameters have on the forecasts. These terms dominate the ones shown in the previous figures. The influence of the additive outliers and level breaks is very clear from these pictures which are just rescaled versions of Fig. 11. The signs in the figures are determined purely by the sign of $\partial a_{n|n}/\partial\theta$. Fig. 13 gives the corresponding graphs for the forecasts which are, from (20), the sums of the relevant panels of Figs. 11 and 12. In practice, this is the most convenient and useful of these graphs as the forecast is of direct interest. Further, y_t tends to be univariate (as in the examples in this paper), even if the state α_t is multivariate.

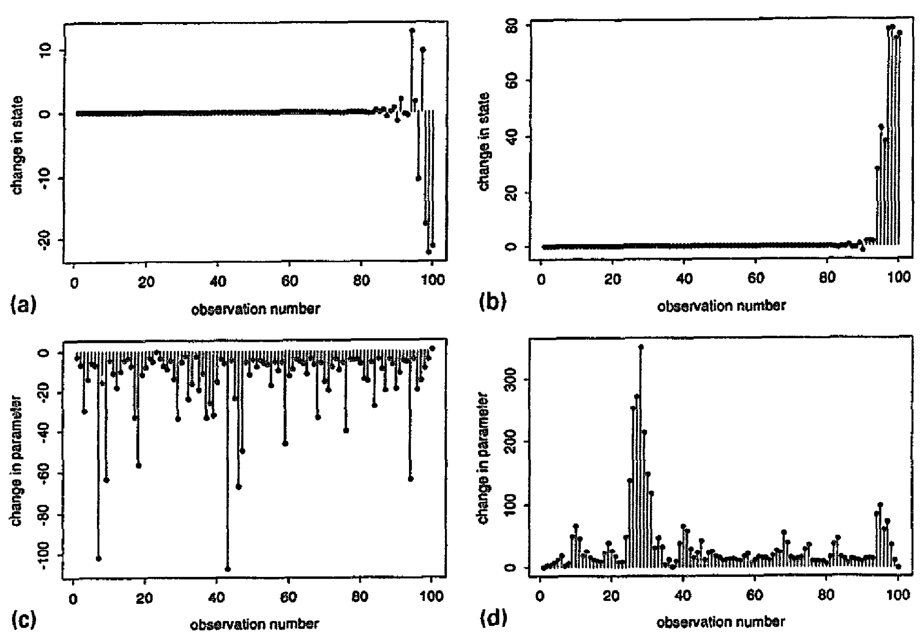

Fig. 12. The problem of the Nile: components of the diagnostic change in forecast, equation (20): change of state due to (a) the outlier intervention and (b) a change in level; change of parameter estimate due to (c) the outlier intervention and (d) a change in level.

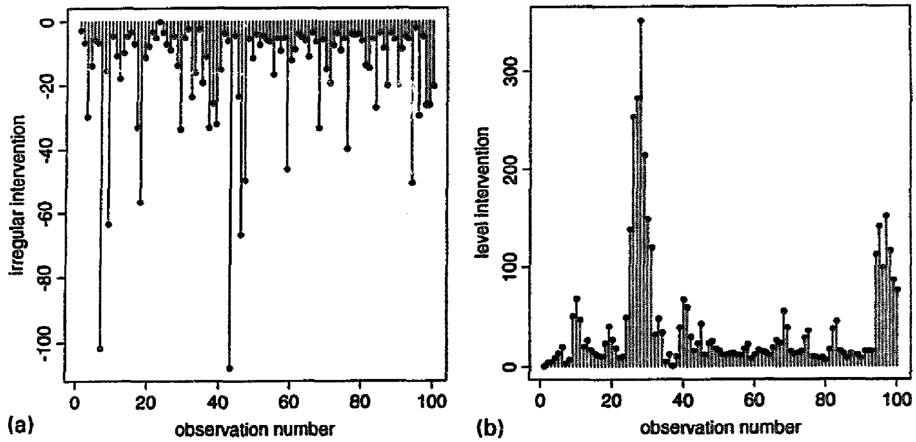

Fig. 13. The problem of the Nile: effects on the forecast of (a) an irregular intervention and (b) the intervention for a change in level.

8. Discussion

This paper addresses the issue of how modellers should check the sensitivity of their fitted models to possible shocks. We have exploited a new type of intervention analysis to suggest interesting directions for investigation and used score statistics and simulation envelopes to form a scaled distance to assess the importance of the interventions.

Our four empirical illustrations exhibit the problems of confounding and masking. In the purse data the additive outlier was confounded with a possible level shift to give a slightly misleading suggestion of the importance of the level break. The opposite effect, that of masking, was detected in the coal series where a coal strike increased the rate of decline in the size of the industry and led to a series of additive outliers in the catchup period as industry re-stocked. It is conceptually possible, as a referee has suggested, to extend our methods to the simultaneous investigation of two departures. However, we would not be happy with such a form of analysis unless there were some prior grounds for expecting several departures: a feature of unobserved components models is that they already allow for flexibility in the evolution of state values. Although we had success with all our sets of data by dealing with one feature and then tackling the next, it is nevertheless clear that there can be empirical problems where this strategy may fail.

The use of simulation envelopes requires a smoothing algorithm if the amount of computing is to be kept within reasonable limits. Our choice was the *S*-plus function 'lowess' (Becker et al., 1988, p. 497) with the parameter *f* which controls the fraction of data used in smoothing equal to 0.35. Although we experimented with values of f for our examples, we did not experiment with other smoothers. The widespread application of our method would seem to require a robust smoothing algorithm which can carry out this part of the analysis in an automatic and satisfactory way – in particular, in a way that automatically selects the bandwidth.

One advantage of the approach advocated in this paper is that it is straightforward to accommodate types of intervention other than those shown in Fig. 1. For example, Atkinson et al. (1994) apply a switching intervention to the analysis of the data on coal consumption discussed in Section 6.2. This models a situation in which lost consumption, perhaps due to a strike, is recovered in the following time period. Several examples are analysed, in an ARIMA framework, by Wu et al. (1993). Atkinson et al. (1994) find that two spaced-switch interventions of the form $\ldots, 0, 0, -1, 0, 1, 0, 0, \ldots$ explain the structure visible in Fig. 9a. A second advantage of our approach is that it can be extended in a relatively straightforward manner to deal with multivariate models. Of course, there would be a higher computational demand, but the complexity of the diagnostics should grow only linearly with the dimension of the series.

A further generalization is to non-Gaussian, linear state-space models where the linear structure of the model is maintained, but the normality assumption is dropped. An example is the stochastic volatility model of Harvey et al. (1994), which is a multivariate time-series model with changing scale. The implication of this is that the Kalman filter no longer delivers the likelihood function, but rather a quasi-likelihood. However, the score vector computed by the Koopman and Shephard (1992) algorithm remains the derivative of the log-quasi-likelihood. Thus the approach of this paper can still be used. Of course, the simulated envelopes would have to be drawn from the correct non-Gaussian density.

Acknowledgements

Financial support from the ESRC is gratefully acknowledged by N. Shephard. The computations were done on equipment funded, in part, by a grant from the SERC to A.C. Atkinson. We thank Ola Elerian for research assistance.

References

Anderson, B.D.O., Moore, J.B., 1979. Optimal Filtering. Prentice-Hall, Englewood Cliffs, NJ.

Atkinson, A.C., 1981. Two graphical displays for influential and outlying observations in regression. Biometrika 68, 13–20.

Atkinson, A.C., 1985. Plots, Transformations, and Regression. Oxford University Press, Oxford.

Atkinson, A.C., 1986. Diagnostic tests for transformations. Technometrics 28, 29–37.

Atkinson, A.C., Koopman, S.J., Shephard, N., 1994. Outliers and switches in time series. In: Huskova, M. (Ed.), Proceedings of the Fifth Prague Symposium on Asymptotic Statistics, Prague, Artia, pp. 35–48.

Atkinson, A.C., Shephard, N., 1996. Deletion diagnostics for transformation of time series. Journal of Forecasting 15, to appear.

Balke, N.S., 1993. Detecting level shifts in time series. Journal of Business and Economic Statistics 11, 81–92.

Becker, R.A., Chambers, C., Wilks, A.R., 1988. The New S Language. Wadsworth and Brooks/Cole, Pacific Grove, CA.

Bowman, K.O., Shenton, L.R., 1975. Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2. Biometrika 62, 243–50.

Box, G.E.P., Tiao, G.C., 1975. Intervention analysis with applications to economic and environmental problems. Journal of the American Statistical Association 70, 70–79.

Bruce, A.G., Martin, R.D., 1989. Leave-k-out diagnostics for time series (with discussion). Journal of the Royal Statistical Society, Series B 51, 363–424.

Burman, P.J., 1985. Report on ASA fellowship 1984–5. Technical report, Statistical Research Division, US Bureau of the Census, Washington, DC.

Carlstein, E., 1988. Nonparametric change-point estimation. Biometrika 16, 188–197.

Chatterjee, S., Hadi, A.S., 1988. Sensitivity Analysis in Linear Regression. Wiley, New York.

Chen, C., Liu, L.-M., 1993. Joint estimation of model parameters and outlier effects in time series. Journal of the American Statistical Association 88, 284–297.

Christiano, L.J., 1992. Searching for a break in GDP. Journal of Business and Economic Statistics 10, 230–257.

Cobb, G.W., 1978. The problem of the Nile: conditional solution to a change point problem. Biometrika 65, 243–251.

Cook, R.D., 1977, detection of influential observations in linear regression. Technometrics 19, 15–18.

Cook, R.D., Weisberg, S., 1982. Residuals and Influence in Regression. Chapman and Hall, London.

Cox, D.R., Hinkley, D.V., 1974. Theoretical Statistics. Chapman Hall, London.

Cox, D.R., Reid, N., 1987. Parameter orthogonality and approximate conditional inference (with discussion). Journal of the Royal Statistical Society, Series B 49, 1–39.

Davison, A.C., Tsai, C.-L. 1992. Regression model diagnostics. International Statistical Review 60, 337–353.

de Jong, P., 1989. Smoothing and interpolation with the state space model. Journal of the American Statistical Association 84, 1085–88.

Dempster, A.P., Selwyn, M.R., Patel, C.M., Roth, A.J., 1984. Statistical and computational aspects of mixed model analysis. Applied Statistics 33, 203–214.

Flack, V.F., Flores, R.A., 1989. Using simulated envelopes in the evaluation of normal probability plots of regression residuals. Technometrics 31, 219–225.

Fox, A.J., 1972. Outliers in time series. Journal of the Royal Statistical Society, Series B 34, 350–363.

Gersch, W., Kitagawa, G., 1983. The prediction of time series with trends and seasonalities. Journal of Business and Economic Statistics 1, 253–264.

Hall, P., Titterington, D.M., 1989. The effect of simulation order on level accuracy and power of Monte Carlo tests. Journal of the Royal Statistical Society, Series B 51, 459–467.

Harrison, P.J., Veerapen, P.P., 1993. Incorporating and deleting information in dynamic models. In: Rao, T.S. (Ed.), Developments in Time Series Analysis: in Honour of Maurice B. Priestley. Chapman and Hall, London, pp. 37–49.

Harrison, J., West, M., 1991. Dynamic linear model diagnostics. Biometrika 78, 797–808.

Harvey, A.C., 1989. Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press, Cambridge.

Harvey, A.C., Durbin, J., 1986. The effects of seat belt legislation on British road casualities: a case study in structural time series modelling (with discussion). Journal of the Royal Statistical Society, Series A 149, 187–227.

Harvey, A.C., Koopman, S.J., 1992. Diagnostic checking of unobserved components time series models. Journal of Business and Economic Statistics 10, 377–389.

Harvey, A.C., Ruiz, E., Shephard, N., 1994. Multivariate stochastic variance models. Review of Economic Studies 61, 247–264.

Kitagawa, G., 1981. A nonstationary time series model and its fitting by a recursive filter. Journal of Time Series Analysis 2, 103–116.

Kohn, R., Ansley, C.F., 1989. A fast algorithm for signal extraction, influence and cross-validation in state space models. Biometrika 76, 65–79.

Koopman, S.J., 1993. Disturbance smoother for state space models. Biometrika 80, 117–126.

Koopman, S.J., Shephard, N., 1992. The exact score for time series models in state space form. Biometrika 79, 823–826.

Ljung, G.M., 1993. On outlier detection in time series. Journal of the Royal Statistical Society, Series B 55, 559–567.

Ljung, G.M., Box, G.E.P., 1978. On a measure of lack of fit in time series models. Biometrika 65, 297–303.

MacNeill, I.B., Tang, S.M., Jandhyala, V.K., 1991. A search for the source of the Nile's change-points. Environmetrics 2, 341–375.

McCulloch, R.E., 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.

Muirhead, C.R., 1986. Distinguishing outlier types in time series. Journal of the Royal Statistical Society, Series B 48, 39–47.

Ng, C.N., Young, P.C., 1990. Recursive estimation and forecasting of non-stationary time series. Journal of Forecasting 9, 173–204.

O'Hara Hines, R.J., Carter, E.M., 1993. Improved added variable and partial residual plots for the detection of influential observations in generalized linear models. Applied Statistics 42, 3–20.

Peña, D., 1990. Influential observations in time series. Journal of Economic and Business Statistics 8, 235–241.

Peña, D., 1991. Measuring influence in dynamic regression models. Technometrics 33, 93–101.

Perron, P., 1989. The great crash, the oil price shock and the unit root hypothesis. Econometrica 57, 1361–1401.

Rappoport, P., Reichlin, L., 1989. Segmented trends and non-stationary time series. Economic Journal 99, 168–177.

Reed, D., 1987. Whistlestop: a community alternative for crime prevention. Ph.D. Thesis, Department of Sociology, Northwestern University, Chicago.

Shephard, N., 1993a. Distribution of the ML estimator of a MA(1) and a local level model. Econometric Theory 9, 377–401.

Shephard, N., 1993b. Maximum likelihood estimation of regression models with stochastic trend components. Journal of the American Statistical Association 88, 590–595.

Theil, H., Wage, S., 1964. Some observations on adaptive forecasting. Management Science 10, 198–206.

Tsay, R.S., 1986. Time series model specification in the presence of outliers. Journal of the American Statistical Association 81, 132–141.

Tunnicliffe Wilson, G., 1989. On the use of marginal likelihood in time series model estimation. Journal of the Royal Statistical Society, Series B 51, 15–27.

Wecker, W.E., Ansley, C.F. 1983. The signal extraction approach to nonlinear regression and spline smoothing. Journal of the American Statistical Association 78, 81–89.

West, M., Harrison, P.J., 1989. Bayesian Forecasting and Dynamic Models. Springer, New York.

Wu, L.S.-Y., Hosking, J.R.M., Ravishanker, 1993. Reallocation outliers in time series. Applied Statistics 42, 301–313.