Assessing the Predictive Influence of Cases in a State-Space Process

by

Joseph E. Cavanaugh

Department of Statistics, 222 Mathematical Sciences Building,

University of Missouri, Columbia, MO 65211

email: cavanaugh@stat.missouri.edu

and

Wesley O. Johnson

Division of Statistics, University of California, Davis, CA 95616

email: wojohnson@ucdavis.edu

Summary

An important inferential objective in state-space modelling is to recover unobserved states using fixed-interval smoothing. Thus, the identification of cases which have a substantial influence on the smoothers is a relevant practical problem. To facilitate this identification, we propose a case-deletion diagnostic which can be easily computed using the outputs of the standard filtering and smoothing algorithms. Our diagnostic is defined as the Kullback-Leibler directed divergence between two versions of the conditional density which determines the smoothers, one based on all the data, the other based on all the data except for the case or cases in question. We investigate the detection performance of the diagnostic in a practical application.

Some key words: Case-deletion diagnostic; EM algorithm; Fixed-interval smoothing; Kalman filtering; Kullback-Leibler divergence; Prediction; Predictive influence function; State-space modelling; Time series analysis.

1. Introduction

Influence diagnostics have been widely discussed in the linear regression setting (see Belsley, Kuh & Welsch, 1980; Cook & Weisberg, 1982). In time series analysis, such diagnostics have received limited attention, although extensive research has appeared on the characterisation and classification of outliers. References on influence assessment in time series include Chernick, Downing & Pike (1982), Lattin (1983), Martin & Yohai (1986), Li & Hui (1987), Peña (1987, 1990, 1991), Abraham & Chuang (1989), Bruce & Martin (1989), Ledolter (1989), LeFrançois (1991) and Van Hui & Lee (1992).

In the state-space setting, a primary inferential goal is to recover unobserved states using fixed-interval smoothing (de Jong, 1988; Kohn & Ansley, 1989); we introduce a diagnostic which assesses the influence of a case or set of cases on the smoothers. Our diagnostic is defined as the Kullback-Leibler directed divergence (Kullback, 1968, p. 5) between two versions of the conditional density which governs the smoothers, one based on all the data, and the other based on all the data except for the case or cases in question. This type of measure was first suggested by Johnson & Geisser (1983) for the detection of influential cases in linear regression.

Our diagnostic can be easily obtained using only the ordinary outputs of the Kalman filtering and fixed-interval smoothing algorithms, often jointly referred to as the Kalman filter smoother. Moreover, if the model parameter estimates are obtained using the EM algorithm (Shumway & Stoffer, 1982), the evaluation of the diagnostic is particularly simple.

Very little work has been published on state-space modelling diagnostics. Kohn & Ansley (1989) extend the concept of leverage to the state-space framework, and propose natural definitions for studentised and deleted residuals. Harrison & West (1991) introduce an influence diagnostic in a Bayesian context which, though related to ours, is distinctly different; see §2.

In §2, our diagnostic is introduced and a computational formula is developed for its evaluation. Important related results are also discussed. In §3, the detection performance of the diagnostic is illustrated in a practical application.

2. Assessing Predictive Influence in a State-Space Process

A q-dimensional state-space process y_t can be represented as

$$y_t = A_t x_t + v_t, \qquad x_t = \Phi x_{t-1} + w_t, \qquad t = 1, \ldots, n, \tag{2.1}$$

where x_t is an unobserved p-dimensional state process, A_t is a $q \times p$ design matrix, assumed known for each t, Φ is a $p \times p$ transition matrix, and v_t and w_t are zero-mean white noise processes (c.f. Shumway, 1988, p. 175).

Let $R = \text{cov}(v_t), Q = \text{cov}(w_t), \mu = E(x_o)$ and $\Sigma = \text{cov}(x_o)$. Let θ denote a parameter vector that uniquely determines the model coefficients and correlation structure. We assume that x_o, the v_t and the w_t are mutually independent and multivariate normal.

Modern work on the state-space model often considers a more general version of (2.1) (de Jong, 1989). We utilise (2.1) since it is the same as or similar to the form of the model usually employed in EM algorithm implementations (c.f. Shumway & Stoffer, 1982; Watson & Engle, 1983; Koopman, 1993).

Let $Y_t = (y_1, \ldots, y_t)$, $X_t = (x_o, x_1, \ldots, x_t)$, $Y = Y_n$, $X = X_n$ and let Y^i represent Y with case y_i omitted.

To predict the unobserved x_t, the smoother computes $\tilde{x}_t(\theta) = E(x_t \,|\, Y)$ for each t. Under the previously mentioned independence and normality assumptions, the $\tilde{x}_t(\theta)$ are linear in the data Y, and can be efficiently evaluated along with their error covariance matrices $P_t(\theta) = E[\{x_t - \tilde{x}_t(\theta)\}\{x_t - \tilde{x}_t(\theta)\}' \,|\, Y]$ using well-known recursions developed in de Jong (1988) and Kohn & Ansley (1989).

Unknown parameters in θ are generally estimated via maximum likelihood, by computing the likelihood $L(\theta \,|\, Y)$ using the Kalman filter and numerically maximising it, or by using the EM algorithm. We will use $\hat{\theta}$ and $\hat{\theta}^{i}$, respectively, to denote the estimators of θ which maximise the likelihoods $L(\theta \,|\, Y)$ and $L(\theta \,|\, Y^{i})$.

The conditional density which determines the smoothers is $f(X \,|\, Y, \hat{\theta}\,)$. Thus, the influence of the case y_i on the smoothers might be judged by measuring the disparity between $f(X \,|\, Y, \hat{\theta}\,)$ and $f(X \,|\, Y^{i}, \hat{\theta}^{i}\,)$. We gauge this disparity by the Kullback-Leibler directed divergence (c.f. Johnson & Geisser, 1983; Johnson, 1985).

We define the predictive influence function for assessing the impact of y_i on the smoothers as

$$PIF(i) = \int \left[\log \left\{ \frac{f(X \,|\, Y, \hat{\theta}\,)}{f(X \,|\, Y^{i}, \hat{\theta}^{i}\,)} \right\} \right] f(X \,|\, Y, \hat{\theta}\,) \; dX. \tag{2.2}$$

It is well known that (2.2) is nonnegative (Kullback, 1968, pp. 14–15). Moreover, its magnitude will be indicative of the impact the deletion of y_i has on $f(X \,|\, Y, \hat{\theta}\,)$.

A formula for the exact evaluation of (2.2) is provided in the following proposition. The derivation of the result is presented in the Appendix.

Let $L(\theta \,|\, X, Y)$ denote the complete-data likelihood of Y and X (Shumway & Stoffer, 1982, p. 256; Shumway, 1988, p. 179), and define

$$Q(\theta \,|\, \theta_*) = E_*\{\log L(\theta \,|\, X, Y) \,|\, Y\}, \tag{2.3}$$

a familiar tool in the EM algorithm, where $E_*(\;\cdot\;|\, Y)$ denotes the expectation with respect to the conditional density $f(X \,|\, Y, \theta_*)$. Similarly, let

$$q_i(\theta \,|\, \theta_*) = E_*\{\log f(y_i \,|\, x_i, \theta\,) \,|\, Y\}. \tag{2.4}$$

Proposition 1.

$$PIF(i) = \left\{ Q(\hat{\theta} \,|\, \hat{\theta}\,) - Q(\hat{\theta}^{i} \,|\, \hat{\theta}\,) + q_i(\hat{\theta}^{i} \,|\, \hat{\theta}\,) \right\} + \left\{ \log L(\hat{\theta}^{i} \,|\, Y^{i}) - \log L(\hat{\theta} \,|\, Y) \right\}. \tag{2.5}$$

As illustrated by the formulae which follow, the functions $Q(\theta \,|\, \theta_*)$ and $q_i(\theta \,|\, \theta_*)$

have convenient representations based on ordinary outputs from the smoother.

Following Shumway & Stoffer (1982, pp. 256–257) and Shumway (1988, p. 179), we have

$$
\begin{aligned}
q_t(\theta \,|\, \theta_*) \;\; = \;\; & -\frac{1}{2} \log |R| \\
& -\frac{1}{2} \operatorname{tr} \left(R^{-1} \left[\{y_t - A_t \tilde{x}_t(\theta_*)\}\{y_t - A_t \tilde{x}_t(\theta_*)\}' + A_t P_t(\theta_*) A_t' \right] \right). \qquad (2.6)
\end{aligned}
$$

Moreover, with $P_{t,t-j}(\theta) = E[\{x_t - \tilde{x}_t(\theta)\}\{x_{t-j} - \tilde{x}_{t-j}(\theta)\}' \,|\, Y]$,

$$
\begin{aligned}
Q(\theta \,|\, \theta_*) \;\; = \;\; & -\frac{1}{2} \log |\Sigma| - \frac{1}{2} \operatorname{tr} \left(\Sigma^{-1} \left[P_o(\theta_*) + \{\tilde{x}_o(\theta_*) - \mu\}\{\tilde{x}_o(\theta_*) - \mu\}' \right] \right) \\
& -\frac{n}{2} \log |Q| - \frac{1}{2} \operatorname{tr} \left[Q^{-1} \left\{ C(\theta_*) - B(\theta_*)\Phi' - \Phi B(\theta_*)' + \Phi A(\theta_*) \Phi' \right\} \right] \\
& + \sum_{t=1}^{n} q_t(\theta \,|\, \theta_*), \qquad (2.7)
\end{aligned}
$$

where

$$
A(\theta_*) = \sum_{t=1}^{n} \{P_{t-1}(\theta_*) + \tilde{x}_{t-1}(\theta_*) \tilde{x}_{t-1}(\theta_*)'\}, \quad B(\theta_*) = \sum_{t=1}^{n} \{P_{t,t-1}(\theta_*) + \tilde{x}_t(\theta_*) \tilde{x}_{t-1}(\theta_*)'\},
$$

$$
C(\theta_*) = \sum_{t=1}^{n} \{P_t(\theta_*) + \tilde{x}_t(\theta_*) \tilde{x}_t(\theta_*)'\}.
$$

Note that, aside from the data y_t and the model parameters μ, Σ, R, Q, and Φ, (2.6) and (2.7) involve only the smoothers $\tilde{x}_t(\theta_*)$, their error covariance matrices $P_t(\theta_*)$, and the cross-covariance matrices $P_{t,t-1}(\theta_*)$. The $\tilde{x}_t(\theta_*)$ and the $P_t(\theta_*)$ are the main outputs of the smoother. The basic smoother recursions can be easily augmented to provide for the evaluation of the $P_{t,t-1}(\theta_*)$; see de Jong (1989, p. 1087). Alternative expressions for $q_t(\theta \,|\, \theta_*)$ and $Q(\theta \,|\, \theta_*)$ may be derived based on the disturbance smoother of Koopman (1993).

The terms in (2.5) can be conveniently computed for each y_i. Once the maximum likelihood vector $\hat{\theta}$ is obtained, $\log L(\hat{\theta} \,|\, Y)$ and $Q(\hat{\theta} \,|\, \hat{\theta})$ are easily found, the latter from (2.7). The vectors $\hat{\theta}^i$ can then be successively determined, using $\hat{\theta}$ as the starting value for θ in the estimation algorithm. The computations are not prohibitive, even when the EM algorithm is used to obtain the $\hat{\theta}^i$. Although the EM algorithm is known to have a slow convergence rate, our investigations suggest $PIF(i)$ performs effectively even if the $\hat{\theta}^i$ are based on a relatively small number of iterations. For each $\hat{\theta}^i$, $\log L(\hat{\theta}^i \,|\, Y^i)$ can be found by evaluating $\log L(\theta \,|\, Y)$ with $\theta = \hat{\theta}^i$ and with y_i treated as missing data, and $Q(\hat{\theta}^i \,|\, \hat{\theta}\,)$ and $q_i(\hat{\theta}^i \,|\, \hat{\theta}\,)$ can be computed from (2.7) and (2.6).

In time series, outliers and influential values often appear in 'patches,' i.e. in sets of contiguous cases. Bruce & Martin (1989) introduce diagnostics that can detect such patches in the ARIMA framework, and discuss the issue of choosing a suitable patch length. An important property of $PIF(i)$ is that it can be used to measure influence for any subset of cases in a series. If the subsets of interest consist of adjacent cases, the derivation and implementation of the computational formula (2.5) follow simply by writing the state-space model (2.1) so that the y_t vectors represent consecutive patches. If the subsets do not consist of adjacent cases, the modifications needed to derive and implement (2.5) are straightforward.

Variations on the diagnostic $PIF(i)$ may be proposed based on related divergence measures considered by Kullback (1968, pp. 6–7). We might use the alternative directed divergence that reverses the roles of $f(X \,|\, Y, \hat{\theta}\,)$ and $f(X \,|\, Y^i, \hat{\theta}^i\,)$ in (2.2), or the symmetric divergence, defined as the sum of the two directed divergences. We will refer to these three divergences by $PIF_1(i)$, $PIF_2(i)$, and $PIF_{12}(i)$, respectively. It is straightforward to derive computational expressions for $PIF_2(i)$ and $PIF_{12}(i)$ which are analogous to (2.5). Of the three measures, we prefer $PIF_1(i)$, since it is based on the averaging density $f(X \,|\, Y, \hat{\theta}\,)$ which remains fixed as i varies.

Harrison & West (1991) introduce an influence diagnostic which is related to $PIF_2(i)$. They consider a state-space model where the parameter matrices R, Q and Φ are allowed to vary over time. Their initial diagnostic presumes all model parameters are known. A more general variant is then derived which allows for uncertainty in the time-varying R. Their diagnostic assesses the effect that the deletion of a case y_i has on the smoother for the associated state x_i. The evaluation of their measure is accomplished using 'dynamic model jacknife' recursions which are a special case of the formulae provided in de Jong (1989). We introduce their diagnostic in the context of model (2.1) using our notation.

With θ assumed known, the Harrison & West diagnostic (1991, p. 801, exp. (7)) can be written as

$$HW^{\theta}(i) = \int \left[\log \left\{ \frac{f(x_i \,|\, Y^{i}, \theta\,)}{f(x_i \,|\, Y, \theta\,)} \right\} \right] f(x_i \,|\, Y^{i}, \theta\,)\; dx_i.$$

If θ is unknown, we might use either $HW^{\hat{\theta}}(i)$ or $HW^{\hat{\theta}^{i}}(i)$. It may seem appealing to define a variant of $HW^{\theta}(i)$ by replacing θ with $\hat{\theta}^{i}$ in $f(x_i \,|\, Y^{i}, \theta\,)$ and θ with $\hat{\theta}$ in $f(x_i \,|\, Y, \theta\,)$, but exact evaluation seems impossible.

The following proposition relates $HW^{\theta}(i)$ to the analogue of $PIF_2(i)$ with θ assumed known, denoted by $PIF_2^{\theta}(i)$. The proposition applies to a case or a collection of cases denoted by y_i and to the associated state(s) denoted by x_i. The proof is presented in the Appendix.

PROPOSITION 2. *Assuming θ is known, we have*

$$\begin{aligned} HW^{\theta}(i) &= \int \left[\log \left\{ \frac{f(x_i \,|\, Y^{i}, \theta\,)}{f(x_i \,|\, Y, \theta\,)} \right\} \right] f(x_i \,|\, Y^{i}, \theta\,)\; dx_i \\ &= \int \left[\log \left\{ \frac{f(X \,|\, Y^{i}, \theta\,)}{f(X \,|\, Y, \theta\,)} \right\} \right] f(X \,|\, Y^{i}, \theta\,)\; dX = PIF_2^{\theta}(i). \qquad (2.8) \end{aligned}$$

Thus, when θ is known, the effect of deleting the case y_i on the conditional density $f(X \,|\, Y, \theta\,)$ for the entire collection of states X is the same as the effect of deleting y_i on the conditional density $f(x_i \,|\, Y, \theta\,)$ for only the corresponding state x_i. In such a setting, $HW^{\theta}(i) = PIF_2^{\theta}(i)$ can be computed using either the appropriate analogue of (2.5) or the recursions employed in Harrison & West (1991). The latter approach is arguably more efficient, since (2.5) involves likelihood-based terms which need not be evaluated when θ is known.

In more practical settings where θ is unknown, one could substitute an estimate for θ into (2.8) and use the result as a diagnostic, e.g. $HW^{\hat{\theta}}(i) = PIF_2^{\hat{\theta}}(i)$. However, if θ is unknown, the influence of y_i on the prediction of the states is largely dictated by the effect of y_i on the estimation of θ. Our investigations suggest that a diagnostic which ignores the estimative influence of y_i tends to be over-sensitive, in that it flags an excessive number of cases as being influential. Our measure $PIF_1(i) = PIF(i)$ does not appear to have this difficulty.

3. An Application of the Diagnostic

Box, Jenkins & Reinsel (1994, p. 545) analyse 310 hourly viscosity measurements taken on a chemical process. They model the series both as a first-order autoregression and as an integrated moving average where the order of differencing and the order of the moving average are both one. Using the former model, they apply an outlier detection procedure suggested by Chang, Tiao & Chen (1988) to identify values in the series which substantially affect the fitted model. Based on this method, they identify an outlier at hour $t = 217$ (Box et al., 1994, pp. 473–474). The viscosity measurements are shown in Fig. 1(a), with case 217 specially marked.

We consider using a local-level model to describe the viscosity series, i.e.

$$\begin{aligned} y_t &= x_t + v_t, \quad v_t \sim N(0, \sigma_m^2), \\ x_t &= x_{t-1} + w_t, \quad w_t \sim N(0, \sigma_s^2), \quad t = 1, \ldots, 310 \end{aligned} \tag{3.1}$$

(c.f. Harvey, 1989, pp. 18–19). The model appears to provide an adequate fit to the series, although the discrete nature of the viscosity measurements makes the propriety of the normality assumptions debatable. The innovations from the fitted model appear approximately normal, except for one outlying value corresponding to case 217.

In fitting (3.1), we use a diffuse prior for x_o: we fix the variance of x_o at a relatively large value, and allow the mean of x_o to be estimated by $\tilde{x}_o(\hat{\theta})$ (c.f. Shumway, 1988, p. 180). The estimates of the variances of w_t and v_t are $\hat{\sigma}_s^2 = 0.08518$ and $\hat{\sigma}_m^2 = 0.00567$.

To determine if any of the observations exert a substantial influence on the smoothers, we compute $PIF(i)$ for $i = 1, \ldots, 310$; see Fig. 1(b). Note that $PIF(217)$ is relatively large, which supports the classification of case 217 as an outlier in the Box et al. (1994) analysis. However, $PIF(170)$ is appreciably larger than $PIF(217)$, even though case 170 was not flagged as suspicious in the Box et al. (1994) analysis and does not appear unusually prominent in Fig. 1(a).

One way in which an outlying value may influence the smoothers is by having an impact on the model parameter estimates. The estimates of σ_s^2 when case 217 and case 170 are individually omitted are respectively 0.08351 and 0.08852. The corresponding estimates of σ_m^2 are 0.00402 and 0.00255.

Although $\hat{\sigma}_s^2$ is only marginally affected by the omission of case 217 or 170, $\hat{\sigma}_m^2$ is substantially reduced when either case is omitted. Note also that both variance estimates change to a greater extent when case 170 is deleted than when case 217 is deleted. This reinforces the implication of the $PIF(i)$ results.

Because $\hat{\sigma}_m^2$ is deflated when either case 170 or 217 is deleted, the smoothers resulting from the omission of either case tend to track the original series more closely than the smoothers based on the complete set of cases. Since the smoothers are thought to represent a filtered reconstruction of the true viscosity levels, any case which has substantial impact on such a reconstruction warrants attention.

In multivariate settings, where even dramatically atypical values may be difficult to spot visually, our diagnostic should serve as a valuable tool for detecting potentially problematic cases.

Acknowledgement

The authors wish to thank the two referees for constructive critiques which served to improve the paper. We also thank Professor D. M. Titterington for helpful suggestions, and Professor Robert Shumway for furnishing part of the software used for the application.

The work of the first author was supported by grants from the Research Council of the University of Missouri-Columbia and the U.S. National Science Foundation.

Appendix

Proofs of Propositions 1 and 2

Proof of Proposition 1. Note that

$$\log f(X \mid Y, \theta) = \log f(X, Y \mid \theta) - \log f(Y \mid \theta), \tag{A.1}$$

$$\begin{aligned} \log f(X \mid Y^{i}, \theta) &= \log f(X, Y^{i} \mid \theta) - \log f(Y^{i} \mid \theta) \\ &= \{\log f(X, Y \mid \theta) - \log f(y_i \mid X, Y^{i}, \theta)\} - \log f(Y^{i} \mid \theta). \end{aligned} \tag{A.2}$$

Additionally, the properties of the state-space process (2.1) allow us to argue that

$$f(y_i \mid X, Y^{i}, \theta) = f(y_i \mid x_i, \theta). \tag{A.3}$$

Using (A.1) along with the definition (2.3), we have

$$\int \{\log f(X \mid Y, \hat\theta)\}\ f(X \mid Y, \hat\theta)\ dX = Q(\hat\theta \mid \hat\theta) - \log L(\hat\theta \mid Y). \tag{A.4}$$

Using (A.2) and (A.3) along with the definitions (2.3) and (2.4), we have

$$\int \{\log f(X \mid Y^{i}, \hat\theta^{i})\}\ f(X \mid Y, \hat\theta)\ dX = Q(\hat\theta^{i} \mid \hat\theta) - q_i(\hat\theta^{i} \mid \hat\theta) - \log L(\hat\theta^{i} \mid Y^{i}). \tag{A.5}$$

Expression (2.5) then follows from utilising (A.4) and (A.5) in conjunction with the definition of $PIF(i)$ provided by (2.2). □

Proof of Proposition 2. We suppress θ in our notation since θ is fixed throughout. Note that

$$f(X \mid Y) = \frac{f(Y^{i}) f(X \mid Y^{i}) f(y_i \mid X, Y^{i})}{f(Y)}, \tag{A.6}$$

$$f(x_i \mid Y) = \frac{f(Y^{i}) f(x_i \mid Y^{i}) f(y_i \mid x_i, Y^{i})}{f(Y)}. \tag{A.7}$$

The properties of the state-space process (2.1) also allow us to argue that

$$f(y_i \mid X, Y^{i}) = f(y_i \mid x_i, Y^{i}) = f(y_i \mid x_i). \tag{A.8}$$

Utilising (A.8) in conjunction with (A.6) and (A.7), we establish the relation

$$\frac{f(X \,|\, Y^{i})}{f(X \,|\, Y)} = \frac{f(x_i \,|\, Y^{i})}{f(x_i \,|\, Y)} = \frac{f(Y)}{f(Y^{i})} \frac{1}{f(y_i \,|\, x_i)} = \frac{f(y_i \,|\, Y^{i})}{f(y_i \,|\, x_i)}. \tag{A.9}$$

Note that (A.9) allows us to write the two divergences in (2.8) as

$$PIF_2(i) = \log f(y_i \,|\, Y^{i}) - \int \{\log f(y_i \,|\, x_i)\} f(X \,|\, Y^{i}) \; dX, \tag{A.10}$$

$$HW(i) = \log f(y_i \,|\, Y^{i}) - \int \{\log f(y_i \,|\, x_i)\} f(x_i \,|\, Y^{i}) \; dx_i. \tag{A.11}$$

Since the integrals in (A.10) and (A.11) are equal, it follows that $PIF_2(i) = HW(i)$.

□

References

Abraham, B. & Chuang, A. (1989). Outlier detection and time series modeling. *Technometrics* **31**, 241–8.

Belsley, D. A., Kuh, E. & Welsch, R. E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity.* New York: Wiley.

Box, G. E. P., Jenkins, G. M. & Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control,* 3^{rd} *Edn.* Englewood Cliffs, New Jersey: Prentice-Hall.

Bruce, A. G. & Martin, R. D. (1989). Leave-k-out diagnostics for time series. *J. R. Statist. Soc.* B **51**, 363–401.

Chang, I., Tiao, G. C. & Chen, C. (1988). Estimation of time series parameters in the presence of outliers. *Technometrics* **30**, 193–204.

Chernick, M. R., Downing, D. J. & Pike, D. H. (1982). Detecting outliers in time series data. *J. Am. Statist. Assoc.* **77**, 743–7.

Cook, R. D. & Weisberg, S. (1982). *Residuals and Influence in Regression.* London: Chapman & Hall.

de Jong, P. (1988). A cross-validation filter for time series models. *Biometrika* **75**, 594–600.

de Jong, P. (1989). Smoothing and interpolation with the state-space model. *J. Am. Statist. Assoc.* **84**, 1085–8.

Harrison, P. 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: Cambridge University Press.

Johnson, W. (1985). Influence measures for logistic regression: Another point of view. *Biometrika* **72**, 59–65.

Johnson, W. & Geisser, S. (1983). A predictive view of the detection and characterization of influential observations in regression analysis. *J. Am. Statist. Assoc.* **78**, 137–44.

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**, 117–26.

Kullback, S. (1968). *Information Theory and Statistics.* New York: Dover.

Lattin, J. M. (1983). Identifying influential observations in time series data. In *Computer Science and Statistics: Proceedings of the 14th Symposium of the Interface*, Ed. K. W. Heiner, R. S. Sacher and J. W. Wilkinson, pp. 283–6. Berlin: Springer-Verlag.

Ledolter, J. (1989). The effect of additive outliers on the forecasts from ARIMA models. *Int. J. Forecasting* **5**, 231–40.

LeFrançois, B. (1991). Detecting over-influential observations in time series. *Biometrika* **78**, 91–9.

Li, W. K. & Hui, Y. V. (1987). On the empirical influence functions of residual autocorrelations in time series. In *Proc. Bus. Econ. Statist. Sec., Am. Statist. Assoc.*, pp. 465–8. Alexandria, Virginia: American Statistical Association.

Martin, R. D. & Yohai, V. J. (1986). Influence functionals for time series. *Ann. Statist.* **14**, 781–818.

Peña, D. (1987). Measuring the importance of outliers in ARIMA models. In *New Perspectives in Theoretical and Applied Statistics*, Ed. M. L. Puri, J. Perez Vilaplana & W. Wertz, pp. 109–18. New York: Wiley.

Peña, D. (1990). Influential observations in time series. *J. Bus. Econ. Statist.* **8**, 235–41.

PEÑA, D. (1991). Measuring influence in dynamic regression models. *Technometrics* **33**, 93–101.

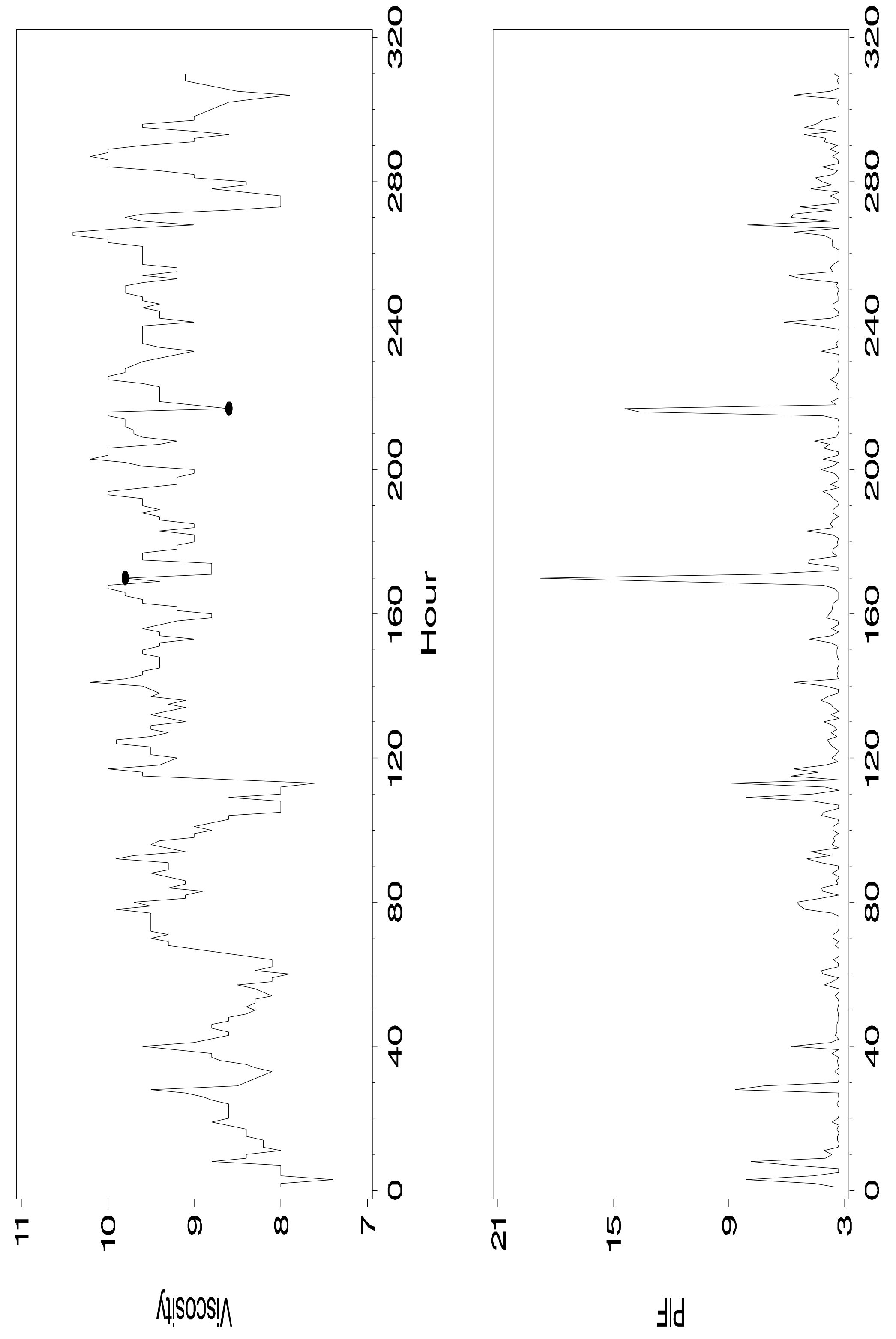
SHUMWAY, R. H. (1988). *Applied Statistical Time Series Analysis.* Englewood Cliffs, New Jersey: Prentice-Hall.

SHUMWAY, R. H. & STOFFER, D. S. (1982). An approach to time series smoothing and forecasting using the EM algorithm. *J. Time Ser. Anal.* **3**, 253–64.

VAN HUI, Y. & LEE, A. H. (1992). Assessing influence in time series. *Commn. Statist.* A **21**, 3463–78.

WATSON, M. W. & ENGLE, R. F. (1983). Alternative algorithms for the estimation of dynamic factor, MIMIC and varying coefficient regression models. *J. Economet.* **23**, 385–400.

Figure 1



Viscosity
11
10
9
8
7
PIF
21
15
9
3
0
40
80
120
160
200
240
280
320
Hour

Figure 1. (a) Hourly viscosity readings taken on a chemical process. Influential cases at hours 170 and 217 are marked with solid dots. (b) $PIF(i)$ values for viscosity readings modelled by (3.1). Largest peaks correspond to $i = 170$ and $i = 217$.