A State Space Modelling approach for Time Series with Patches of Unusual Observations

Jeremy Penzer

Department of Statistics, London School of Economics
Houghton Street, London, WC2A 2AE, UK.
www.lse.ac.uk/collections/statistics
j.penzer@lse.ac.uk

August 28, 2003

Summary
An alternative to leave-k-out diagnostics for detecting patches of outlying points in time series is developed. We propose that unusual behaviour should be modelled by the addition of shocks. By including shocks in the transition equation of our state space model, we allow for the possibility of persistent changes in structure associated with a patch of outliers. This persistent change may take the form of a level shift or a change in seasonal pattern. We provide an efficient mechanism for computing diagnostic statistics associated the addition of k shocks using a simple adaptation of the Kalman filter. Statistics for detecting unspecified patterns of shocks and an interpretation of the output of the associated smoothing algorithm are derived. Illustrations using real series are given.

KEYWORDS: Kalman filter; Leave-k-out; Outliers; Shock detection; Smoothing.

1 Introduction

Time series data often contain outliers or changes in structure. The development of techniques to detect and account for outlying points is motivated by the observation that their presence may have a strong influence on parameter estimates and forecasts (Harvey 1989, p.330). In the seminal work in this field, Fox (1972) characterises outliers as additive or innovative. A number of authors (Tsay 1986; Chang, Tiao, and Chen 1988) develop iterative procedures based on this categorisation. The possibility of permanent changes in structure is acknowledged by Box and Tiao (1975) who use step function input to as a means of generating intervention structure. Level shifts are included in the additive/innovative outlier framework by Tsay (1988). Harvey and Durbin (1986) provide a practical illustration of intervention analysis using structural time series models while Harvey and Koopman (1992) put forward an efficient means of detecting structural change using smoothed disturbances. De Jong and Penzer (1998) establish the equivalence of detection methods based on the introduction dummy intervention variables and, more efficient, smoother based approaches. The focus of the work cited above is on behaviour which can be associated with an individual time point. This point may be the location of a transient effect, such as an additive outlier, or the start of a permanent change in structure, such as a level shift.

The idea that outliers may occur in patches is accepted in regression analysis. Techniques based on deleting observations, sometimes referred to as leave-k-out diagnostics, are described in standard regression diagnostics texts, such as Cook and Weisberg (1982) or Atkinson (1985). Methods for detecting patches of unusual points in time series are less well established although the subject receives comprehensive treatment in Bruce and Martin (1989). Bruce and Martin derive statistics based on the measuring the influence of sections of data on estimates of ARMA model parameters and innovations variance. Working with autoregressive processes, Justel, Peña, and Tsay (2001) handle patches of outliers using an adaptation of a Markov chain Monte Carlo method for detecting individual outlying points given by McCulloch and Tsay (1994). In a recent paper, Proietti (2003) puts forward an efficient means of generating statistics associated with deletion residuals for state space models. All leave-k-out approaches assume implicitly that the process dynamics are identical either side of any unusual points. However, in time series data, a patch of outliers may be associated with a level shift, seasonal break, or some other non-transient alteration in structure.

We propose that a patch unusual behaviour is best modelled by the introduction of k shocks. By allowing shocks to the transition equation of a state space model, persistent alterations in structure can be represented. Thus, statistics for detecting this sort of departure from the null model are *put-k-shock-in diagnostics*. The put-k-shocks-in framework includes leave-k-out as a special case; diagnostics associated with deleting k observations are identical to those generated by introducing k measurement shocks. We proceed by generalizing Proietti's (2003) pseudo-model. By exploiting and extending results from De Jong and Penzer (1998), very efficient means of generating diagnostics for patches of unusual points in time series are derived. The presence of outliers and structural changes will distort estimates of the null model parameters. However, we do not advocate re-estimation of these parameters during the detection process. Using our method, the distortion in null model parameters does not mask the presence of outliers and structural changes. The computational expense of repeated parameter estimation is not justified.

Consider a univariate time series of n observations, $y = (y_1, \ldots, y_n)'$. Assume that we have fitted a linear Gaussian model thought, at least initially, to be an adequate representation of the data. We refer to this as the null model, denoted $y \sim N(0, \Sigma)$; zero mean is used for efficiency of notation. The alternative admits the possibility of outliers and structural changes occuring in a patch, say from time $i-k+1$ to i. These are modelled via introduction of an additive dummy regressor $D_{(i)}$ with associated magnitude $\delta_{(i)}$. For a given i, the alternative model is $y \sim N(D_{(i)}\delta_{(i)}, \Sigma)$. The significance of the outliers and structural changes represented by $D_{(i)}$ can be determined using statistics based on the least squares estimate of the magnitude parameter, $\hat{\delta}_{(i)}$. We focus on the statistic $\tau^2_{(i)} = \hat{\delta}'_{(i)} \mathrm{Var}(\hat{\delta}_{(i)})^{-1} \hat{\delta}_{(i)}$. Our interest lies in detecting and locating unusual behaviour. Diagnostic tests and informative plots can be constructed from the sequence $\{\tau^2_{(k)}, \ldots, \tau^2_{(n)}\}$. A simple and efficient method for computing these diagnostic statistics is the key contribution of this paper.

In §2, §3 and §4 we focus on the evaluation of $\tau^2_{(i)}$ for a given value of i. §2 develops a state space model for a patch of k unusual points based on the introduction of shocks while §3 describes efficient evaluation of $\tau^2_{(i)}$ using a novel filtering result. The results of §2 and §3 depend on knowledge of $D_{(i)}$. In §4 we derive diagnostic statistics for the case where the nature of the intervention is not known in advance. §5 describes practical aspects of detecting and locating patches of unusual points including a method for choosing k. Applications to real data are given in §6 and §7 concludes.

2 Intervention model for patches of unusual observations

Our choice of null model is motivated by the fact that all linear time series models have a state space representation (Anderson and Moore 1979; Harvey 1989). In this form, the disturbance vectors $\{\epsilon_t\}$ is related to the observations $\{y_t\}$ via a Markov process $\{\alpha_t\}$. We adopt the state space representation used by De Jong (1991). This representation admits the possibility of correlation between the error term in the measurement equation (1) and the error term in the transition equation (2). For $t = 1, \ldots, n$, we take

$$\begin{aligned} y_t &= Z_t \alpha_t + G_t \epsilon_t, & (1) \\ \alpha_{t+1} &= T_t \alpha_t + H_t \epsilon_t, & (2) \end{aligned}$$

where $\{\epsilon_t\} \sim \text{NID}(0, I)$ are the disturbances and NID denotes normal and independently distributed. The system matrices Z_t, T_t, G_t and H_t are deterministic and any unknown elements they contain are estimated using maximum likelihood. For a series with an $m \times 1$ state vector α_t and $p \times 1$ error vector ϵ_t, the matrices Z_t, T_t, G_t and H_t are respectively $1 \times m$, $m \times m$, $1 \times p$ and $m \times p$. The initial state is taken to be $\alpha_1 \sim N(a_1, P_1)$. For a stationary process, a_1 and P_1 are the unconditional mean and variance of the initial state. Non-stationarity requires a diffuse initialisation handled by adaptations to the filter recursions (Ansley and Kohn 1990; De Jong 1991; Koopman 1997; Koopman and Durbin 2003).

We first consider modelling a patch of outliers. Suppose that we know that outliers occur at times $i - k + 1$ to i. We can accounted for these outliers via the alternative model $y \sim N(D_{(i)}\delta_{(i)}, \Sigma)$ where $D_{(i)} = (D_{i-k+1}, \ldots, D_i)$ is the $n \times k$ intervention matrix and $\delta_{(i)} = (\delta_{i-k+1}, \ldots, \delta_i)'$ is the $k \times 1$ vector of magnitude parameters. For $j = i - k + 1, \ldots, i$, the t^{th} element of D_j is given by

$$[D_j]_t = \begin{cases} 1, & t = j, \\ 0, & \text{otherwise}, \end{cases}$$

thus, one magnitude parameter is associated with each outlier. We adopt a paradigm in which outliers are viewed as a consequence of measurement shocks (Fox 1972; Box and Tiao 1975; Tsay 1986; De Jong and Penzer 1998). These shocks are introduced in the measurement equation, so the alternative model is

$$\begin{aligned} y_t &= X_t \delta_t + Z_t \alpha_t + G_t \epsilon_t, \\ \alpha_{t+1} &= T_t \alpha_t + H_t \epsilon_t, \end{aligned} \qquad (3)$$

where $X_t = 1$ for $t \in [i - k + 1, i]$. Thus, X_t is an indicator for measurement shocks.

The alternative model given by (3) does not allow for anything other than transient changes in structure, that is, a patch of k outliers. A more general approach is to allow shocks to the transition equation. These transition shocks can be used to model persistent behaviour such as level shifts and seasonal changes. The alternative model is then

$$\begin{aligned} y_t &= X_t\delta_t + Z_t\alpha_t + G_t\epsilon_t, \\ \alpha_{t+1} &= W_t\delta_t + T_t\alpha_t + H_t\epsilon_t. \end{aligned} \tag{4}$$

The quantity W_t is a vector of indicators for shocks to the components of the state. Outside the region of unusual behaviour, that is, for $t \notin [i-k+1, i]$, we take $X_t = 0$ and $W_t = 0$. For $t \in [i-k+1, i]$, some of the elements of X_t and W_t may be given the value 1. The choice of dimensions for X_t and W_t and the location of non-zero values among their elements determines the structure of the intervention.

Consider time j in the region of unusual behaviour, that is, $j \in [i-k+1, i]$. Let δ_j be $d_j \times 1$, so X_j is $1 \times d_j$ and W_j is $m \times d_j$. In addition define $d_{(i)} = \sum_{j=i-k+1}^{i} d_j$. The additive impact of the shock at time j is $D_j\delta_j$, where D_j is an $n \times d_j$ matrix. Iterating in (4) yields the result (De Jong and Penzer 1998, eqn. (11)) that the t^{th} row of D_j is given by

$$[D_j]_t = \begin{cases} 0, & t = 1, \ldots, j-1, \\ X_j, & t = j, \\ Z_t T_{t-1,j+1} W_j, & t = j+1, \ldots, n, \end{cases} \tag{5}$$

where $T_{t-1,j+1} = T_{t-1} \ldots T_{j+1}$ for $t > j+1$, $T_{t-1,j+1} = I$ for $t = j+1$ and $T_{t-1,j+1} = 0$ for $t \leq j$. The combined additive impact of putting k shocks into the model is $D_{(i)}\delta_{(i)}$ where $D_{(i)} = (D_{i-k+1}, \ldots, D_i)$ is $n \times d_{(i)}$ and $\delta_{(i)} = (\delta'_{i-k+1}, \ldots, \delta'_i)'$ is $d_{(i)} \times 1$.

3 Evaluation of put-k-in shock diagnostics

Consider the alternative model, $y \sim N(D_{(i)}\delta_{(i)}, \Sigma)$. The least squares estimate of the magnitude of the intervention can be written

$$\hat{\delta}_{(i)} = S_{(i)}^{-1} s_{(i)}, \tag{6}$$

where

$$\begin{aligned} s_{(i)} &= D'_{(i)}\Sigma^{-1}y \\ S_{(i)} &= D'_{(i)}\Sigma^{-1}D_{(i)}. \end{aligned} \tag{7}$$

We use a statistic based on the magnitude estimate to determine the significance of the intervention;

$$\tau_{(i)}^2 = \hat{\delta}_{(i)}\text{Var}(\hat{\delta}_{(i)})^{-1}\hat{\delta}_{(i)} = s'_{(i)}S_{(i)}^{-1}s_{(i)}. \tag{8}$$

Under the null hypothesis of no intervention, $\tau_{(i)}^2 \sim \chi_p^2$ where $p = \text{rank}(S_{(i)})$. We derive an efficient mechanism for evaluating (8) using the Kalman filter and smoother.

The Kalman filter and smoother (KFS) form the basis for statistical treatment of state space models. For a Gaussian model, the Kalman filter evaluates the minimum mean squared estimator (MMSE) of the state vector α_{t+1} given observations $\{y_1, \ldots, y_t\}$, denoted a_{t+1}, and the associated mean square error, $P_{t+1} = \text{MSE}(a_{t+1})$. The Kalman filter is, for $t = 1, \ldots, n$,

$$\begin{aligned} v_t &= y_t - Z_t a_t, & F_t &= Z_t P_t Z_t' + G_t G_t', \\ & & K_t &= \left(T_t P_t Z_t' + H_t G_t'\right) F_t^{-1}, \\ a_{t+1} &= T_t a_t + K_t v_t, & P_{t+1} &= T_t P_t L_t' + H_t J_t', \end{aligned} \tag{9}$$

where $L_t = T_t - K_t Z_t$ and $J_t = H_t - K_t G_t$. The quantities v_t are referred to as *innovations.* Notice that the form of the filter is entirely determined by the state space form given by (1) and (2). Only the innovations, $\{v_t\}$, and the estimates of the states, $\{a_t\}$, are dependent on the values of the input series, $\{y_t\}$. An efficient smoothing algorithm is given by De Jong (1989) and Kohn and Ansley (1989). From the output of this smoother, linear estimators based on the full sample can be computed. Smoothing takes the form of a backwards recursion, for $t = n, \ldots, 1$,

$$\begin{aligned} u_t &= F_t^{-1} v_t - K_t' r_t, & M_t &= F_t^{-1} + K_t' N_t K_t, \\ r_{t-1} &= Z_t' u_t + T_t' r_t, & N_{t-1} &= Z_t' F_t^{-1} Z_t + L_t' N_t L_t, \end{aligned} \tag{10}$$

where $r_n = 0$ and $N_n = 0$. The u_t are referred to as *smoothing errors.*

The Kalman filter implicitly defines a lower triangular matrix C such that $v = Cy$, where $v = (v_1, \ldots, v_n)'$. It is well known (Harvey 1989, p. 131) that C is the Cholesky decomposition of the inverse covariance matrix of the observations, that is, $\Sigma^{-1} = C'F^{-1}C$ where $F = \text{diag}\{F_1, \ldots, F_n\}$. Less well known (De Jong and Penzer 1998; Durbin and Koopman 2001) is the fact that C' defines the smoothing operation (10). Thus,

$$u = C'F^{-1}v = C'F^{-1}Cy = \Sigma^{-1}y, \tag{11}$$

where $u = (u_1, \ldots, u_n)'$. Iterative substitution in the equation for r_{t-1} from (10) yields

$$r_j = \sum_{t=j+1}^{n} T_{t-1,j+1}' Z_t' u_t. \tag{12}$$

Using the notational conventions of §2, we define $s_{(i)} = (s_{i-k+1}', \ldots, s_i')'$, where s_j is a $d_j \times 1$ vector. We refer to the s_j as *shock contrasts.* Starting from (7), substituting from (11) then using the results (5) and (12) yields, for $j = i - k + 1, \ldots, i$,

$$\begin{aligned} s_j &= D_j' \Sigma^{-1} y = D_j' u \\ &= \sum_{t=j}^{n} [D_j]_t' u_t = X_j' u_j + \sum_{t=j+1}^{n} W_j' T_{t-1,j+1}' Z_t' u_t \\ &= X_j' u_j + W_j' r_j. \end{aligned} \tag{13}$$

Thus, $s_{(i)}$ can be evaluated from smoother output using (13). However, our real interest lies in the diagnostic statistic $\tau_{(i)}^2$. Direct computation of this statistic, would involve evaluating and inverting the matrix $S_{(i)}$. This is prohibitively time consuming, particularly given that $S_{(i)}$ is not evaluated by the KFS.

We propose an efficient method for evaluating $\tau_{(i)}^2$ that generalizes the approach described by Proietti (2003). Proietti observes that the updating equations for the smoother can be used to define a pseudo-model for the smoothing errors and that this pseudo-model has a state space form. We can apply a similar argument to the contrast s_j. From (10)

$$r_{j-1} = Z_j' u_j + T_j' r_j = L_j' r_j + Z_j' F_j^{-1} v_j.$$

We define a Markov process $\{\beta_j\}$ where

$$\beta_{j-1} = L_j' \beta_j + Z_j' F_j^{-1/2} \eta_j, \tag{14}$$

and $\{\eta_j\} \sim N(0, I)$. The contrast s_j is then an instance of

$$s_j = Q_j' \beta_j + X_j' F_j^{-1/2} \eta_j, \tag{15}$$

where $Q_j = W_j - K_j X_j$. Equations (14) and (15) are respectively transition and measurement equations of our pseudo-model for $\{s_j\}$. The associated Kalman filter is of the same form as (9), with the system matrices replaced by the corresponding quantities from (14) and (15); for $j = i, \ldots, i-k+1$,

$$\begin{aligned} v_j^* &= s_j - Q_j' a_j^*, & F_j^* &= Q_j' P_j^* Q_j + X_j' F_j^{-1} X_j, \\ & & K_j^* &= \left(L_j' P_j^* Q_j + Z_j' F_j^{-1} X_j \right) F_j^{*-1}, \\ a_{j-1}^* &= L_j' a_j^* + K_j^* v_j^*, & P_{j-1}^* &= L_j' P_j^* L_j + Z_j' F_j^{-1} Z_j - K_j^* F_j^* K_j^{*\prime}, \end{aligned} \tag{16}$$

initialised with $a_i^* = 0$ and $P_i^* = N_i$. The $*$ superscript is used to denote quantities associated with the pseudo-model. We refer to the recursions defined by (16) as the pseudo-model filter. The pseudo-model filter generates an orthogonal sequence $v_{(i)}^* = (v_{i-k+1}^*, \ldots, v_i^*)'$ and $\mathrm{Var}(v_{(i)}^*) = F_{(i)}^* = \mathrm{diag}\{F_{i-k+1}^*, \ldots, F_i^*\}$. Here, $v_{(i)}^*$ and $F_{(i)}^*$ are analogous to the innovations vector v and the innovation variance matrix F in the null model filter. Diagnostic statistics are evaluated using

$$\tau_{(i)}^2 = s_{(i)}' S_{(i)}^{-1} s_{(i)} = v_{(i)}^{*\prime} F_{(i)}^{*\,-1} v_{(i)}^* = \sum_{j=i-k+1}^{i} v_j^{*\prime} F_j^{*-1} v_j^*, \tag{17}$$

avoiding the need to evaluate and invert the matrix $S_{(i)}$.

We have established that the pseudo-model filter provides an efficient mechanism for evaluating the diagnostic statistic $\tau_{(i)}^2$. The associated smoother is, for $j = i-k+1, \ldots, i$,

$$\begin{aligned} u_j^* &= F_j^{*-1} v_j^* - K_j^{*\prime} r_j^*, & M_j^* &= F_j^{*-1} + K_j^{*\prime} N_j^* K_j^*, \\ r_{j+1}^* &= Q_j u_j^* + L_j r_j^*, & N_{j+1}^* &= Q_j F_j^{*-1} Q_j' + L_j^{*\prime} N_j^* L_j^*, \end{aligned} \tag{18}$$

where $L_j^* = L_j' - K_j^* Q_j'$ and the smoother is initialised at $r_{i-k+1}^* = 0$ and $N_{i-k+1}^* = 0$. The pseudo-model smoother (18) is used to evaluate the least squares estimate of the shock magnitude. Compare $\hat{\delta}_{(i)} = S_{(i)}^{-1} s_{(i)}$, given by (6), with the general smoothing error vector, $u = (\mathrm{Var}(y))^{-1} y$, given by (11). Since $S_{(i)} = \mathrm{Var}(s_{(i)})$, the magnitude estimate is equivalent to the smoothing error associated with the contrasts $s_{(i)}$. Thus $\hat{\delta}_{(i)} = (u_{i-k+1}^{*\prime}, \ldots, u_i^{*\prime})'$ is efficiently computed using the pseudo-model smoother. A practical application for $\hat{\delta}_{(i)}$ values is described in §5.

4 Diagnostics for unspecified patterns of shocks

The preceding sections put forward an efficient method for evaluating statistics associated with patterns of k shocks is developed. However, in practice it is unlikely that the exact nature of the intervention is known prior to implementing the diagnostics. In this section we derive statistics for detecting a patch of k unusual points that do not depend on an exact specification of the way in which shocks enter the system.

We start by allowing for an outlier at every point in $[i-k+1, i]$. These k measurement shocks can be thought of as deleting k observations. We are not concerned with the location of structural changes within $[i-k+1, i]$, however, to allow for a change in the dynamics of the process after a patch of k unusual points, a shock is introduced to each component of the state at time i. This set up can be summarised as follows:

$$X_j = 1 \text{ for } j = i-k+1, \ldots, i-1, \quad \text{and} \quad (X_i', W_i')' = I_{m+1}, \tag{19}$$

where I_{m+1} is the $(m+1) \times (m+1)$ identity. Introducing the shocks defined by (19) takes into account all available information about the difference in structure before and after a patch of k unusual points. The following lemma clarifies this point.

Lemma 4.1 *The shock magnitude for the shocks defined by* (19) *cannot be estimated if a non-null value W_j is introduced, where $j \in [i-k+1, i-1]$.*

Proof: From (13) and the shocks defined by (19) we have $s_{(i)} = (u_{i-k+1}', \ldots, u_i', r_i')'$. A non-null W_j introduces a new component $W_j' r_j$ into $s_{(i)}$. However, from (10) $r_j = T_{i,j+1}' r_i + \sum_{t=j+1}^{i} T_{t-1,j+1}' Z_t' u_t$, so $W_j' r_j$ is a linear combination of other elements of $s_{(i)}$. We conclude that $S_{(i)} = \mathrm{Var}(s_{(i)})$ is singular and that $\hat{\delta}_{(i)} = s_{(i)}' S_{(i)}^{-1} s_{(i)}$ cannot be evaluated.□

The $\tau_{(i)}^2$ statistic associated with (19) can be evaluated without resorting to the pseudo-model filter. This particular $\tau_{(i)}^2$ statistic will be referred to frequently and is given its own identifier, $\rho_{(i)}^2$. Under the null hypothesis of no shocks, $\rho_{(i)}^2 \sim \chi_{k+m}^2$.

Proposition 4.1 *The $\tau_{(i)}^2$ statistic associated with* (19) *is given by*

$$\tau_{(i)}^2 = \rho_{(i)}^2 = r_i' N_i^{-1} r_i + \sum_{t=i-k+1}^{i} v_t' F_t^{-1} v_t. \tag{20}$$

Proof: Consider the first iteration of the pseudo-model filter. The intialisation is $a_i^* = 0$ and $P_i^* = N_i$. Since $(X_i', W_i')' = I_{m+1}$, we have $Q_i = W_i - K_i X_i = (-K_i, I_m)$. Thus,

$$\begin{aligned} v_i^* &= s_i = X_i' u_i + W_i' r_i = (u_i', r_i')', \\ F_i^* &= \begin{pmatrix} K_i' N_i K_i + F_i^{-1} & -K_i' N_i \\ -N_i K_i & N_i \end{pmatrix} \quad \Rightarrow \quad F_i^{*-1} = \begin{pmatrix} F_i & F_i K_i' \\ K_i F_i & N_i^{-1} + K_i F_i K_i' \end{pmatrix}, \end{aligned}$$

$$\begin{aligned} K_i^* &= (Z_i',\ T_i'), \\ a_{i-1}^* &= r_{i-1}, \\ P_{i-1}^* &= 0. \end{aligned}$$

The value of F_i^{*-1} is given by standard partition matrix inversion results.

Assume that for a value $j \in [i-k+1, i-1]$ we have $a_j^* = r_j$ and $P_j^* = 0$. One step of the filter recursions yields

$$\begin{aligned} v_j^* &= s_j - Q_j' r_j = X_j' u_j + W_j' r_j - (W_j' - X_j' K_j') r_j = X_j' F_j^{-1} v_j = F_j^{-1} v_j, \\ F_j^* &= X_j' F_j^{-1} X_j = F_j^{-1}, \\ K_j^* &= (Z_j' F_j^{-1} X_j) F_j^{*-1} = Z_j', \\ a_{j-1}^* &= L_j' r_j + K_j^* v_j^* = r_{j-1}, \\ P_{j-1}^* &= Z_j' F_j^{-1} Z_j - K_j^* F_j^* K_j^{*\prime} = 0. \end{aligned} \tag{21}$$

By induction $v_j^* = F_j^{-1} v_j$ for all $j \in [i-k+1, i-1]$. Now (17) gives the required value of $\tau_{(i)}^2$. □

For the shocks defined by (19) the pseudo-model smoother becomes, for $j = i-k+1, \ldots, i-1$,

$$\begin{aligned} u_j^* &= v_j - Z_j q_j, \qquad & M_j^* = F_j + Z_j N_j^* Z_j', \\ r_{j+1}^* &= T_j r_j^* - K_j v_j, \qquad & N_{j+1}^* = K_j F_j K_j' + T_j N_j^* T_j' \end{aligned} \tag{22}$$

and

$$u_i^* = \begin{pmatrix} v_i - Z_i q_i \\ N_i^{-1} r_i - T_i q_i + K_i v_i \end{pmatrix}, \qquad M_i^* = \begin{pmatrix} F_i + Z_i N_i^* Z_i' & F_i K_i' + Z_i N_i^* T_i' \\ K_i F_i + T_i N_i^* Z_i' & N_i^{-1} + K_i F_i K_i' + T_i N_i^* T_i' \end{pmatrix}.$$

All of the system matrices involved in this pseudo-model smoother come directly from KFS output. The algorithm is a straightforward consequence of the proof of proposition 4.1.

5 Practical consideration in detecting k shocks

This section deals with practical issues arising from real data applications of put-k-shocks-in diagnostics. The basic structural model (Harvey 1989) is used for illustration. The model is

$$y_t = \mu_t + \gamma_t + \varepsilon_t, \qquad \{\varepsilon_t\} \sim \text{NID}(0, \sigma_\varepsilon^2). \tag{23}$$

Here μ_t is the trend, γ_t is the seasonal component and ε_t is an error term. The trend is taken to be locally linear

$$\begin{aligned} \mu_{t+1} &= \mu_t + \beta_t + \eta_t, \quad & \{\eta_t\} \sim \text{NID}(0, \sigma_\eta^2), \\ \beta_{t+1} &= \beta_t + \zeta_t, \quad & \{\zeta_t\} \sim \text{NID}(0, \sigma_\zeta^2). \end{aligned}$$

Let s be the seasonal period. We use a dummy seasonal component,

$$\gamma_{t+1} = -\sum_{j=1}^{s-1} \gamma_{t+1-j} + \omega_t, \qquad \{\omega_t\} \sim \text{NID}(0, \sigma_\omega^2),$$

or a trigonometric seasonal component in which seasonal effects are generated by adding stochastic cycles with frequencies $2\pi j/s$, where $j = 1, \ldots, [s/2]$, and $\text{NID}(0, \sigma_\omega^2)$ error terms. All of the error terms in the model are mutually independent.

5.1 Heuristic determination of k

To draw meaningful inferences about real situations, we would like to be able to specify the particular value of k which is most appropriate for a given data set and null model. A heuristic for determing k, which is effective in practice, is described below. The approach can be adapted to the leave-k-out case by replacing $\rho^2_{(i)}$ in (24) with $\tau^2_{(i)}$ for the inclusion of k measurement shocks. For any given value of k, calculate $\rho^2_{(i)}$ statistics for $i = k, \dots, n$. Let the largest of these be λ_k so

$$\lambda_k = \max_i(\rho^2_{(i)}). \tag{24}$$

Replicating this calculation for a plausible set of values of k yields $\lambda_1, \dots, \lambda_q$, where q is the maximum number of consecutive shocks we are willing to add. Taking q to be the the closest integer to $\min(0.1n, 15)$ is reasonable for most applications. Generating the sequence $\lambda_1, \dots, \lambda_q$ is efficient as it only requires null model KFS output. Each λ_k represents the most extreme event when adding k shocks and the sequence $\lambda_1, \dots, \lambda_q$, is monotone increasing. If a value, k^* say, is the appropriate value of k, we would expect the increases in λ_k to be insignificant for $k > k^*$. In practice, we evaluate the first difference $\Delta\lambda_k$ for $k = 1, \dots, q$, taking $\lambda_0 = 0$. If $\Delta\lambda_k < c_k$ for all $k > k^*$ then k^* is the appropriate number of shocks, where $\{c_k\}$ is a set of critical values. By definition $\Delta\lambda_1 = \lambda_1 = \chi^2_{1+m}$. For $k > 1$, the upper 5% point of a χ^2_1 provides a guide to significance. Taking $c_1 = \chi^2_{1+m,0.95}$ and $c_k = 4$ for $k > 1$ provides sensible results in real examples although larger values may be chosen to account for simultaneous testing.

5.2 Shock detection and choosing appropriate interventions

Once an appropriate value of k has been chosen, diagnostic statistics are generated. Index plots of diagnostics may be used to detect and categorize shocks. A complementary approach is to use the largest order statistic defined by (24). Let λ_k be an instance of the random variable Λ_k. The Bonferonni criteria (David 1981) define an upper limit for $\text{pr}(\Lambda_k > \lambda_k)$ under the null hypothesis of no shocks. This is interpreted as an approximation to the probability value (p-value) associated with our test for unusual behaviour. The test statistic, λ_k, is the largest order statistic of a sample from $n - k + 1$ correlated random variables each distributed as χ^2_{k+m} under H_0. The Bonferonni upper bound for $\text{pr}(\Lambda_k > \lambda_k)$ is given by $p = (n - k + 1)(1 - F(\lambda_k))$ where $F(.)$ is the distribution function of a χ^2_{k+m}. The null hypothesis is rejected in a test with significance level α provided $\alpha > p$. Note that the true p-value will always be smaller than p. When the test indicates that shocks are present, the location of the last of the k shocks is determined by the value of i at which $\rho^2_{(i)}$ attains the value λ_k.

To account for unusual behaviour intervention variables, that is dummy regressors, are included in the model. Choice of the appropriate shape of intervention is important to ensure accurate parameter estimation and to allow us to make meaningful statements about the shocks in the context of the series which we are analyzing. An interpretation of the output of the pseudo-model smoother was given in §3. In §4 we use k measurement shocks followed by a shock to all of the elements of the state for situations where the nature of the intervention is not known. The shape associated with an intervention generated by shocking all of the state are neither intuitively appealing nor readily interpretable. However, using the estimated magnitudes, $\hat{\delta}_{i-k+1}, \ldots, \hat{\delta}_i$, generated by the pseudo-model smoother (18), we can identify the element of the state on which the shock has the largest impact. The lower m elements of $\hat{\delta}_i$ correspond to the elements of the state. The pseudo-model smoother also generates standard errors which are used to scale the estimated magnitudes to approximate standard normals. We use the largest of these scaled magnitudes to identify the element of the state on which the shock has greatest impact. For example, for a basic structural model, if the first element is the largest we categorize the unusual behaviour as a level shift. If none of the scaled magnitudes associated with the states are significant, we conclude that an intervention consisting of measurement shocks alone is appropriate.

6 Illustrative real data examples

This section gives two real data examples to illustrate the use of put-k-shocks-in diagnostics. In each instance we conduct a leave-k-out analysis for comparison. Computational results were produced with code written in Ox (Doornik 1998) using Ssfpack (Koopman, Shephard, and Doornik 1999).

6.1 Gas data

Figure 1 shows quarterly gas consumption (logs) by other final users, which includes domestic consumption, between 1960 quarter 1 and 1986 quarter 4. This series is used by Durbin and Koopman (2000) to illustrate a non-Gaussian model. A basic structural model with quarterly dummy seasonal component fits these data. The hyperparameter estimates for the null model are given in table 4. The scaled innovations in figure 2 indicate some unusual behaviour in 1970 and 1971. However, innovations do not accurately pinpoint the location of the shock nor do they give an indication of which type of intervention is appropriate.

From table 1 we see that, for both leave-k-out and put-k-shocks-in, $\Delta\lambda_k < 4$ for $k > 2$ indicating that $k = 2$ is appropriate for both methods. Figure 3 is an index plot of leave-2-out diagnostics with Bonferroni 5% and 10% critical values marked. The largest value $\tau^2_{(i)}$-statistic is at $i = 44$ indicating that 1970 quarter 3 and 1970 quarter should be deleted from the sample or, equivalently, measurement outliers should be included at these points. The results from fitting a model which takes these outliers into account are given, under the heading LkO, in table 4. The fitted model suggests much higher than expected gas consumption in 1970 quarter 3 and much lower than expected consumption in 1970 quarter 4 with both outliers highly significant. Inclusion of interventions to account for these points reduces dramatically the estimated variance of the irregular component.

For $k = 2$, the largest $\rho^2_{(i)}$-statistic is 58.51 occurring in 1970 quarter 4; see figure 4. From table 4, the associated Bonferroni p-value is 3.21×10^{-8} so the null hypothesis of no shocks is rejected at all reasonable significance levels. In order to identify the appropriate intervention, we evaluate the null model based, estimated magnitudes associated with shocks to the components of the state in 1970 quarter 4 using the pseudo-model smoother. These values are given in table 3. The largest impact is on the second seasonal component. Using (5), we establish that a shock to the second seasonal component at $t = i$ corresponds to an additive intervention of the form

$$D = \begin{cases} -1, & \text{for } \ t = i+2, \ i+6, \ldots, \\ 1, & \text{for } \ t = i+4, \ i+8, \ldots, \\ 0, & \text{otherwise.} \end{cases} \tag{25}$$

Under the column heading PkSI, table 4 give the estimated hyperparameters and intervention t-statistics from fitting a model with outliers in 1970 quarter 3 and 1970 quarter 4 and the seasonal intervention given by (25) in 1970 quarter 4. The two measurement outliers are still highly significant although they now indicate a smaller decrease in quarter 4. This can be attributed to the fact that the seasonal shift caused by a decrease in consumption in quarter 2 and an increase in quarter 4 from 1971 quarter 2 onwards is now accounted for. The estimated variances of both the irregular and seasonal components are markedly lower than in the null model. The seasonal shift may be explained by the introduction of North Sea gas and the accompanying increased use of gas for domestic heating.

6.2 New air passengers data

Figure 5 is a time series plot of the monthly number of passengers (thousands) on UK airlines from January 1980 to December 1996. A basic structural model with trigonometric seasonal component fits these data. Hyperparameter estimates are given in the null model column of table 8. Both the time series plot and the scaled innovations of figure 6 show unusual behaviour at the end of 1990 and start of 1991. However, neither accurately locates nor categorizes the shocks.

The $\Delta\lambda_k$ statistics, in the LkO row of table 5, indicate that leave-6-out is appropriate for this data set. Note that, when persistent effects are present in the data, this method of determining k may not be reliable in the leave-k-out context. The maximum $\tau^2_{(i)}$ value occurs in April 1991 corresponding to inclusion of measurement outliers from November 1990 to April 1991. The intervention t-statistics for this model are given in column LkO of table 8. All are negative suggesting a six month period of lower than expected air passenger numbers followed by a return to null model behaviour in May 1991.

For the put-k-shocks-in approach, $\Delta\lambda_k < 4$ for $k > 4$ so a put-4-shocks-in analysis is conducted. The maximum value of $\rho^2_{(i)}$ is 76.00 in February 1991; see table 6. Inspection of table 7, the scaled magnitudes associated with a shock at this point, indicates that the largest impact is on the level component. Thus, the appropriate interventions are 4 measurement shocks, November 1990 to February 1991, and a level shift in February 1991. By definition, this level shift will have the effect of changing the level from March 1991 onwards. Table 8 shows that including these interventions reduces dramatically the estimated variance of the level component. All of the t-statistics are negative and highly significant, suggesting 4 months of much lower than expected passenger numbers followed by a permanent shift downwards in level. With these interventions in the model, there is no significant unusual behaviour left unaccounted for.

The obvious explanation for the drop in airline passenger numbers is the Gulf War. Iraq invaded Kuwait on 2nd August 1990. Over the following months it became increasingly clear that the situation would not be resolved diplomatically. By the beginning of November war seemed inevitable, fuel prices increased dramatically and on November 8th the US decision to double the size of its military force in Saudi Arabia was made public. This coincides with a slump of around 240 000 in airline passenger numbers for November 1990, a drop which is maintained in December 1990. The Allied air attack started on 17th January 1991, on 24th February the ground assault began and by the 28th February the cease-fire took effect. During January 1991 the number of air passengers fell further and reached a low in February 1991 nearly 920 000 passengers short of what might have been expected. After the cease-fire, numbers rose but only to a level 615 000 lower than before the war.

7 Conclusion

Leave-k-out diagnostics are useful for detecting patches of outliers. However, they may give misleading results if the unusual behaviour in a series is persistent. For example, in §6.2, the model suggested by leave-k-out estimates a drop of 630 000 in passenger number for February 1991 while put-k-shocks-in put the figure at 919 000. Leave-k-out underestimates the effect of the Gulf War because it does not take into account the non-transient downward shift in the level of airline passengers which the war induces. Put-k-shocks-in includes leave-k-out as a special case and evaluation of general put-k-shocks-in diagnostic, $\rho^2_{(i)}$, does not require additional filtering. Thus, put-k-shocks-in is a good generic procedure which reduces to leave-k-out where appropriate.

Acknowledgement: This work is based on an original suggestion by Piet de Jong.

References

Anderson, B. D. O. and J. B. Moore (1979). *Optimal filtering.* Englewood Cliffs, New Jersey: Prentice-Hall.

Ansley, C. F. and R. Kohn (1990). Filtering and smoothing in state space models with partially diffuse initial conditions. *Journal of Time Series Analysis 11*, 275–294.

Atkinson, A. C. (1985). *Plots, Transformations and Regression.* Oxford: Clarendon Press.

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

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

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

Cook, R. D. and S. Weisberg (1982). *Residuals and Influence in Regresssion.* New York: Chapman and Hall.

David, H. A. (1981). *Order Statistics* (2nd ed.). New York: John Wiley.

De Jong, P. (1989). Smoothing and interpolation with the state-space model. *Journal of the American Statistical Association 84*, 1085–1088.

De Jong, P. (1991). The diffuse Kalman filter. *Annals of Statistics 19*, 1073–1083.

De Jong, P. and J. R. Penzer (1998). Diagnosing shocks in time series. *Journal of the American Statistical Association 93*, 796–806.

Doornik, J. A. (1998). *Ox: an Object-Oriented Matrix Programming Language*. West Wickham, Kent: Timberlake Consultants Press.

Durbin, J. and S. J. Koopman (2000). Time series analysis of non-gaussian observations based on state space models from both classical and Bayesian perspectives. *Journal of the Royal Statistical Society, Series B 62*, 3–56.

Durbin, J. and S. J. Koopman (2001). *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.

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

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

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

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

Justel, A., D. Peña, and R. S. Tsay (2001). Detection of outlier patches in autoregressive time series. *Statistica Sinica 11*, 651–673.

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

Koopman, S. J. (1997). Exact initial Kalman filter and smoother for non-stationary time series models. *Journal of the American Statistical Association 92*, 1630–1638.

Koopman, S. J. and J. Durbin (2003). Filtering and smoothing of state vector for diffuse state space models. *Journal of Time Series Analysis 24*, 85–98.

Koopman, S. J., N. Shephard, and J. A. Doornik (1999). Statistical algorithms for models in state space form using Ssfpack 2.2. *Econometrics Journal 2*, 113–166.

McCulloch, R. E. and R. S. Tsay (1994). Bayesian analysis of autoregressive time series via the Gibbs sampler. *Journal of Time Series Analysis 15*, 235–250.

Proietti, T. (2003). Leave-k-out diagnostics in state space models. *Journal of Time Series Analysis 24*, 221–236.

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

Tsay, R. S. (1988). Outliers, level shifts and variances changes in time series. *Journal of Forecasting 7*, 1–20.

Table 1: Gas example – $\Delta\lambda_k$ statistic

	k										
	1	2	3	4	5	6	7	8	9	10	11
LkO	18.06	23.04	0.12	3.29	3.34	2.62	0.05	1.87	0.13	1.74	0.06
PkSI	43.79	14.72	0.32	0.34	0.22	1.43	0.45	0.32	0.34	0.31	1.43

Table 2: Gas example – maximum τ^2-statistics

	L2O	P2SI
Maximum value	41.10	58.51
Bonferroni p-value	1.28×10^{-7}	3.21×10^{-8}
Location	1970Q4	1970Q4

Table 3: Gas example – scaled magnitudes for shocks to the state in 1970Q4

Components of the state				
Level	Slope	Seas1	Seas2	Seas3
1.769	1.657	0.452	2.682	-0.875

Table 4: Gas example – estimated hyperparameters and intervention t-statistics

Hyperparameters		Model suggested by	
($\times 10^3$)	Null model	LkO	PkSI
σ_ε^2	1.823	0.232	0.767
σ_η^2	0.000	0.338	0.249
σ_ζ^2	0.008	0.005	0.005
σ_ω^2	3.308	2.038	1.014
Intervention t-statistics			
Outlier 1970Q3	-	7.992	7.890
Outlier 1970Q4	-	-6.053	-3.756
Seasonal shift 1970Q4	-	-	5.916
Seasonal shift 1979Q4	-	-	-

Table 5: Air passengers example – $\Delta\lambda_k$ statistic

	k														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
LkO	35.94	6.91	1.46	8.92	0.25	4.19	0.31	0.36	2.39	2.34	0.90	0.86	0.28	0.81	0.61
PkSI	55.29	8.06	4.60	8.06	0.52	3.10	1.69	1.42	0.61	2.61	0.45	1.06	0.45	0.65	0.03

Table 6: Air passengers example – maximum τ^2-statistics

	L6O	P4SI
Maximum value	57.66	76.00
Bonferroni p-value	2.69×10^{-8}	3.93×10^{-7}
Location	1991Apr	1991Feb

Table 7: Air passengers example – scaled magnitudes for shocks to the state in 1991Feb

Level	Slope	Seas1	Seas2	Seas3	Seas4	Seas5	Seas6	Seas7	Seas8	Seas9	Seas10	Seas11
-3.74	0.72	-1.45	3.32	0.70	0.20	-0.26	0.31	0.20	-0.16	0.89	-0.53	-0.66

Table 8: Air passengers example – estimated hyperparameters and intervention t-statistics

		Model suggested by	
Hyperparameters	Null model	LkO	PkSI
σ^2_ε	178.40	0.0000	373.69
σ^2_η	3787.8	2177.2	1326.1
σ^2_ζ	0.8297	1.6462	1.8940
σ^2_ω	13.473	15.842	17.271
Intervention magnitude (t-statistics)			
Outlier 1990Nov	-	-145.21 (-2.49)	-241.44 (-3.98)
Outlier 1990Dec	-	-119.64 (-1.75)	-276.75 (-3.84)
Outlier 1991Jan	-	-335.40 (-4.58)	-564.88 (-6.82)
Outlier 1991Feb	-	-630.60 (-8.60)	-919.19 (-10.02)
Outlier 1991Mar	-	-246.10 (-3.60)	-
Outlier 1991Apr	-	-245.20 (-4.21)	-
Level shift 1991Feb	-	-	-615.33 (-6.25)

Figure 1: Gas example – quarterly gas consumed by other final users (logs)

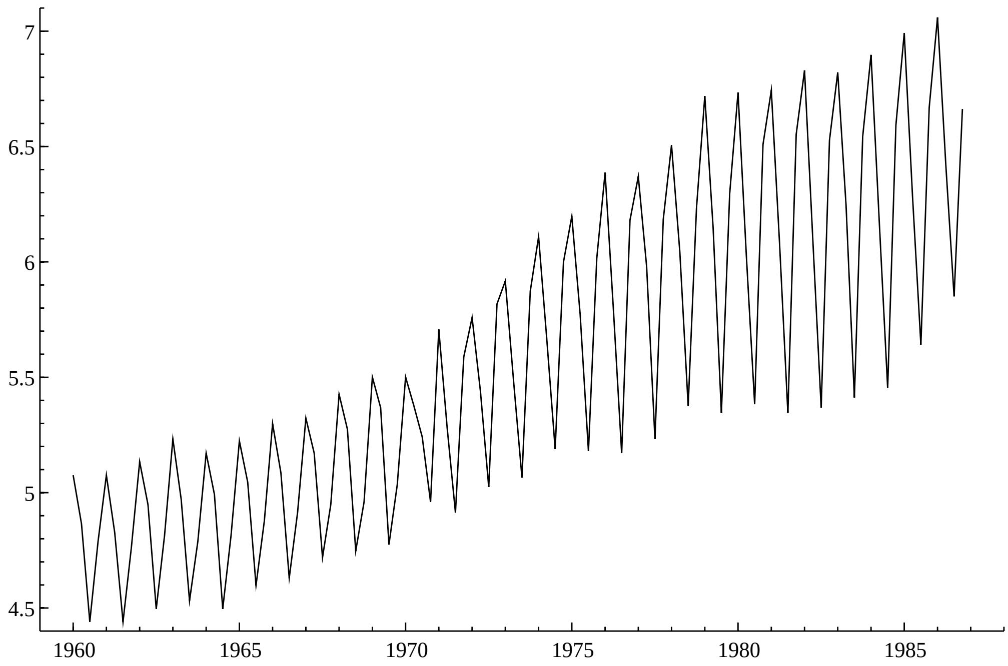

Figure 2: Gas example – scaled innovations

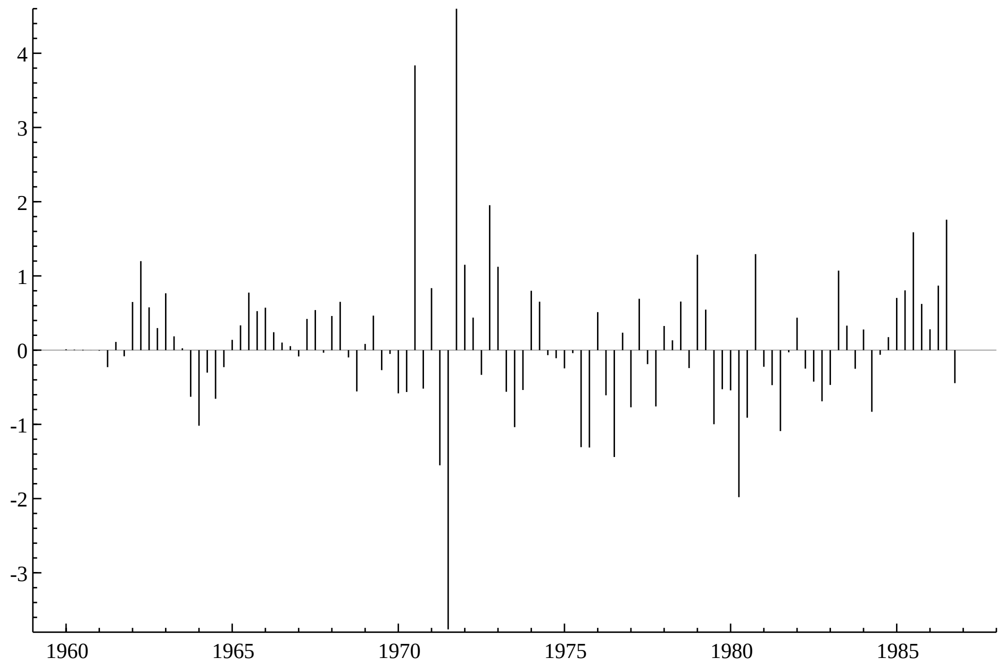

Figure 3: Gas example – leave-2-out diagnostics

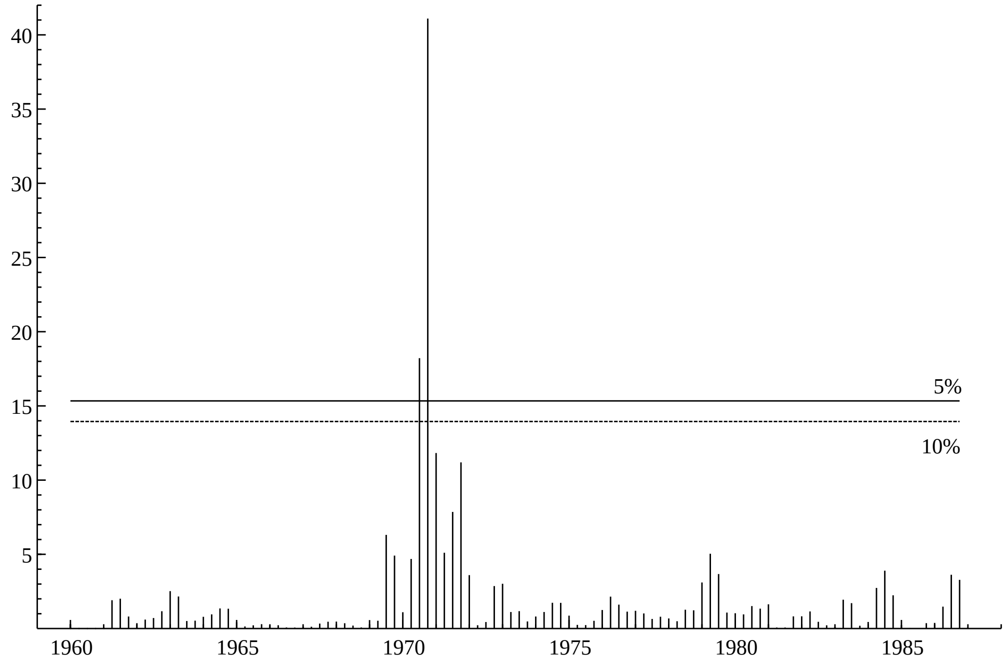

Figure 4: Gas example – put-2-shocks-in diagnostics

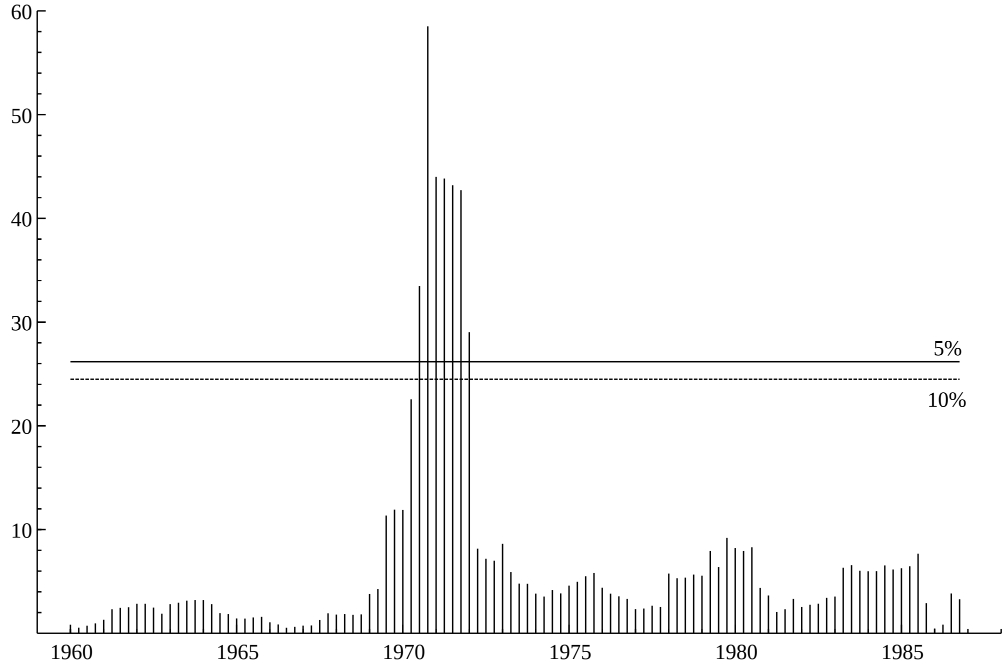

Figure 5: Air passengers example – monthly air passengers on UK airlines (thousands)

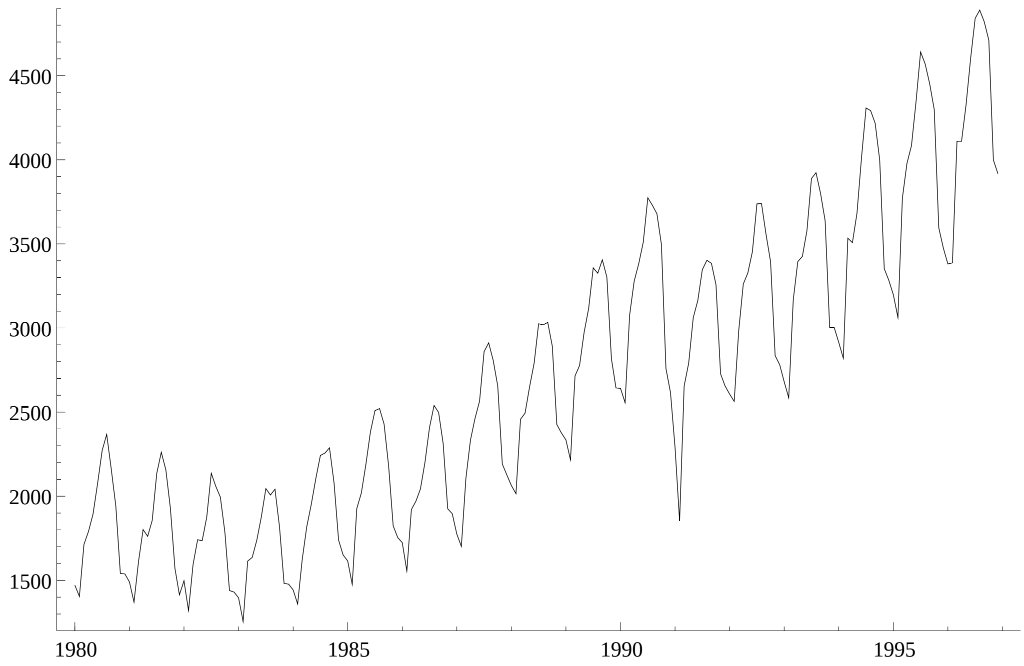

Figure 6: Air passengers example – scaled innovations

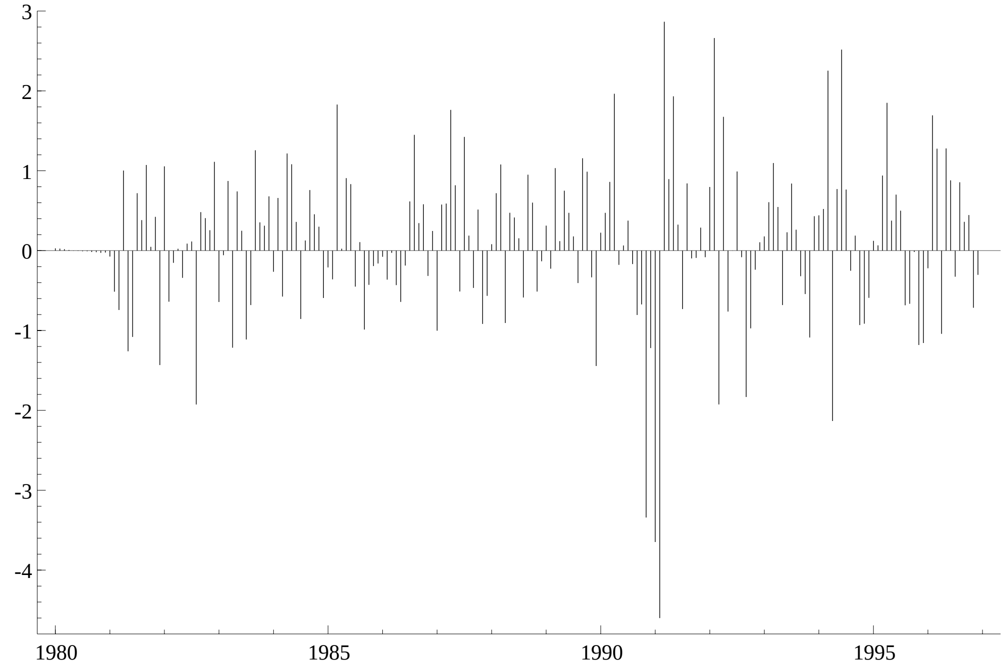

Figure 7: Air passengers example – leave-6-out diagnostics

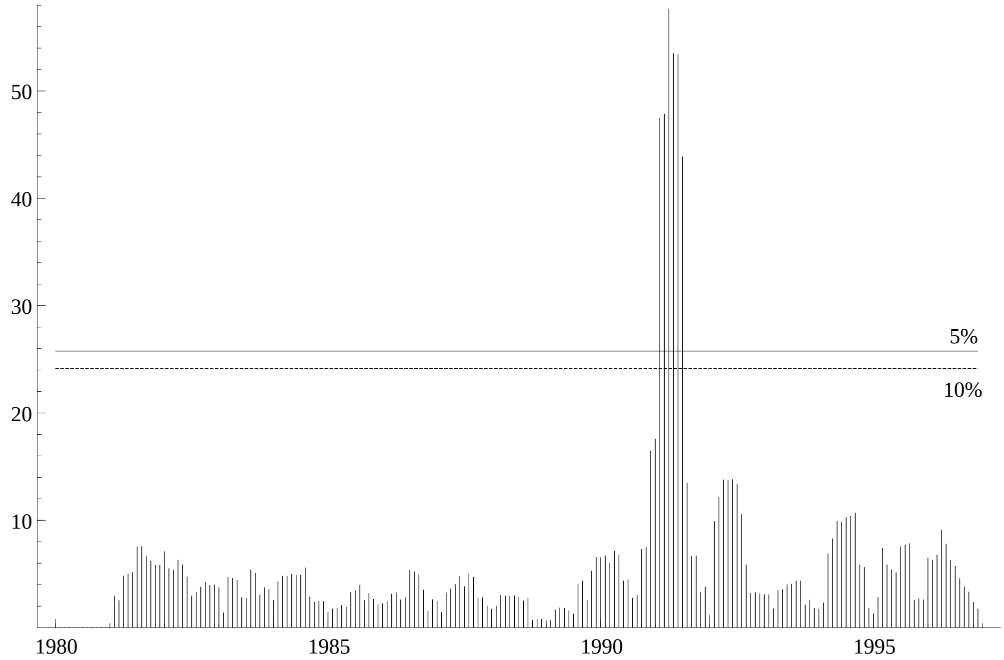

Figure 8: Air passengers example – put-4-shocks-in diagnostics