Put-k-shocks-in time series diagnostics

Jeremy Penzer

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

August 13, 2001

Abstract

This paper puts forward an alternative to leave-k-out diagnostics for detecting patches of unusual points in time series. The proposed approach, in which shocks are introduced, allows for more general patterns of unusual behaviour. Efficient means of evaluating statistics associated the addition of k shocks, based on partition filtering, are suggested. Maximal statistics and and interpretation of the output of the partition smoother are also derived.

KEYWORDS: Kalman filter smoother; Leave-k-out; Outliers; Shock detection; State space model.

1 Introduction

In regression analysis, leave-k-out diagnostics (Atkinson 1985; Cook and Weisberg 1982) are a well established means of detecting patches of outliers. The idea is applied to time series, in particular ARMA models, by Bruce and Martin (1989). Using a state space framework, Proietti (2000) puts forward efficient means of generating statistics associated with deleting a section of data. Leave-k-out assumes implicitly that the model's 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, which has a persistent effect, is best modeled by the introduction of k shocks. 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 pseudo-model and associated filter. 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 paper is organized as follows. §2 and §3 deal with background material, covering respectively, Kalman filtering and smoothing, and Proietti's approach to leave-k-out diagnostics. §4 establishes the equivalence of leave-k-out and put-k-measurement-shocks in. A general framework for modeling patches of unusual points and efficient means of generating diagnostic statistics are developed in §5. §6 describes maximal statistics generated from the output of the Kalman filter and smoothing algorithm without the requirement for any additional filtering. In addition to being accurate in the case of persistent shock effects, using maximal statistics is more efficient than leave-k-out. §7 provides a mechanism which allows the nature of departures from the null model to be determined by the data. Practical issues, such as the choice of k, are dealt with in §8 with real data examples given in §9.

2 An interpretation of smoothing errors

All linear time series models admit state space representation. In this form, the disturbance vector $\{\epsilon_t\}$ is related to the observation vector $\{y_t\}$ via a Markov process $\{\alpha_t\}$. A convenient expression of the state space form is, for $t = 1, \ldots, n$,

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

where $\{\epsilon_t\} \sim (0, I)$ are the disturbances and $\alpha_1 \sim (a_1, P_1)$ is the initial state. The disturbances are mutually uncorrelated variables which are also uncorrelated with the initial state. The system matrices Z_t, T_t, G_t and H_t are deterministic and any unknown elements which they contain are estimated. For notational convenience we assume that $E(y_t) = 0$, for all t, although mean effects can be readily incorporated in (1).

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

$$\begin{aligned} v_t &= y_t - Z_t a_t, & F_t &= Z_t P_t Z_t' + G_t G_t', \\ & & K_t &= \left(T_t P_t Z_t' + H_t G_t'\right) F_t^{-1}, \\ a_{t+1} &= T_t a_t + K_t v_t, & P_{t+1} &= T_t P_t L_t' + H_t J_t', \end{aligned} \tag{2}$$

where $L_t = T_t - K_t Z_t$, $J_t = H_t - K_t G_t$ and v_t are referred to as *innovations.* De Jong (1989) gives a smoothing algorithm from which linear estimators based on the full sample Y_n can be computed. Smoothing takes the form of a backwards recursion, $t = n, \ldots, 1$,

$$\begin{aligned} u_t &= F_t^{-1} v_t - K_t' r_t, & M_t &= F_t^{-1} + K_t' N_t K_t, \\ r_{t-1} &= Z_t' u_t + T_t' r_t, & N_{t-1} &= Z_t' F_t^{-1} Z_t + L_t' N_t L_t, \end{aligned} \tag{3}$$

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

It is useful to describe two basic properties of smoothing errors. Proof of these results can be found in De Jong and Penzer (1998). Let $u = (u_1', \ldots, u_n')'$, be the vector of smoothing errors, and $M = \text{Var}(u)$. If $y = (y_1', \ldots, y_n')'$ and $\Sigma = \text{Var}(y)$ then

$$u = \Sigma^{-1} y. \tag{4}$$

Suppose that Jy is a selection of elements of y and Ky a vector containing the remaining elements. If $P(Jy|Ky)$ denotes the MMSLE of Jy based on Ky, we have

$$Ju = (JMJ')\{Jy - P(Jy|Ky)\}. \tag{5}$$

Suppose we want to calculate a residual based on deleting k observations up to and including the i^{th}, that is $y_{i-k+1}, \dots, y_i$. A quantity of interest is then $y_{(I)} - P(y_{(I)}|y^{(I)})$ where $y_{(I)} = (y'_{i-k+1}, \dots, y'_i)'$ and $y^{(I)} = (y'_1, \dots, y'_{i-k}, y'_{i+1}, \dots, y'_i)'$. By equation (5),

$$u_{(I)} = M_{(I)}\{y_{(I)} - P(y_{(I)}|y^{(I)})\},$$

where $u_{(I)} = (u'_{i-k+1}, \dots, u'_i)'$ and $M_{(I)} = \text{Var}(u_{(I)})$. Thus, smoothing errors can be interpreted as interpolation residuals.

3 Pseudo-model and partition filter

Proietti (2000) uses the statistic $u'_{(I)}M^{-1}_{(I)}u_{(I)}$ to test whether the observations $y_{i-k+1}, \dots, y_i$ are jointly outlying. Evaluating this statistic directly involves evaluating and inverting the matrix $M_{(I)}$. This is potentially time consuming, particularly since $M_{(I)}$ is not evaluated by the KFS. Proietti's solution is based on the observation that the updating equations for the smoother can be used to define a pseudo-model for the smoothing errors. From (3)

$$r_{t-1} = Z'_t u_t + T'_t r_t = L'_t r_t + Z'_t F_t^{-1} v_t.$$

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

$$\beta_{t-1} = L'_t \beta_t + Z'_t F_t^{-1/2} \eta_t, \tag{6}$$

and $\{\eta_t\} \sim \text{WN}(0, I)$. The smoothing errors are thought of as being generated by

$$u_t = -K'_t \beta_t + F_t^{-1/2} \eta_t. \tag{7}$$

Equations (6) and (7) are respectively transition and measurement equations of our pseudo-model for $\{u_t\}$. The associated filter is given by

$$\begin{aligned} w_t &= u_t + K'_t b_t, & F^*_t &= K'_t P^*_t K_t + F_t^{-1}, \\ & & K^*_t &= \left(-L'_t P^*_t K_t + Z'_t F_t^{-1}\right) F_t^{*-1}, \\ b_{t-1} &= L'_t b_t + K^*_t w_t, & P^*_{t-1} &= L'_t P^*_t L_t + Z'_t F_t^{-1} Z_t - K^*_t F^*_t K^{*\prime}_t, \end{aligned} \tag{8}$$

Running filter (8) for $t = i, \dots, i-k+1$ initialized with $b_i = 0$ and $P^*_i = N_i$, implicitly diagonalizes $M_{(I)}$. The statistic of interest is then given by

$$\tau^2_{i,k} = u'_{(I)} M^{-1}_{(I)} u_{(I)} = \sum_{t=i-k+1}^{i} w'_t F_t^{*-1} w_t. \tag{9}$$

For a Gaussian process with all model hyperparameters known, $\tau^2_{i,k} \sim \chi^2_{k\ell}$ under the null hypothesis of no outliers where ℓ is the dimension of the observations.

Proietti does not suggest a name for filter (8) directly although, as its originator, this is his prerogative. *Psuedo-model filter* is an obvious choice. Another possibility is suggested by considering the filter in terms of the variance matrices involved. If $w_{(I)} = (w'_{i-k+1}, \ldots, w'_i)'$ and $F_{(I)} = \text{Var}(w_{(I)})$ then

$$(F_{i-k+1}^{-1/2} w'_{i-k+1}, \ldots, F_i^{-1/2} w'_i)' = F_{(I)}^{-1/2} w_{(I)} = M_{(I)}^{-1/2} u_{(I)}$$

where $M_{(I)}^{-1/2}$ is the inverse of the Cholesky decomposition of $M_{(I)}$. We know, from the definition of $M_{(I)}$ and (4), that $M_{(I)} = J\Sigma^{-1}J'$. Thus, filter (8) inverts the Cholesky decomposition of a section of the matrix Σ^{-1} where Σ^{-1} is partitioned appropriately. We conclude, a useful and descriptive name for (8) is *partition filter.*

4 Put-k-measurement-shocks-in diagnostics

Proietti's leave-k-out approach is an important development in the treatment of unusual points by deletion. An alternative diagnostic paradigm views outliers as being associated with 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 model becomes

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

The equivalent of leaving k out is to take $X_t = I$ for $t = i-k+1, \ldots, i$ and $X_t = 0$ otherwise. Writing this as a stacked regression we have

$$y = D\delta_{(I)} + z, \tag{10}$$

where D is the transpose of the selection matrix for elements $i-k+1$ to i and $\delta_{(I)} = (\delta'_{i-k+1}, \ldots, \delta'_i)'$. The vector $z = (z'_1, \ldots, z'_n)'$ has the state space form given by (1), so $\text{Var}(z) = \Sigma$. Generalized least squares (GLS) and the property of smoothing errors given by (4) yield

$$\hat{\delta}_{(I)} = \left(D'\Sigma^{-1}D\right)^{-1} D'\Sigma^{-1}y = M_{(I)}^{-1}u_{(I)}.$$

Hence, the statistic $\tau_{i,k}^2$, associated with leaving k observations out, can also be thought of in term of the addition of k measurement shocks. In fact,

$$\tau_{i,k}^2 = \hat{\delta}'_{(I)} \text{Var}(\hat{\delta}_{(I)})^{-1} \hat{\delta}_{(I)},$$

where $\hat{\delta}_{(I)}$ is the vector of estimated shock magnitudes.

5 A general put-k-shocks-in framework

Leave-k-out, and the equivalent put-k-measurement-shocks-in, do 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 induce behaviour such as level shifts and seasonal changes. The model is

$$
\begin{aligned}
y_t &= X_t\delta_t + Z_t\alpha_t + G_t\epsilon_t, \\
\alpha_{t+1} &= W_t\delta_t + T_t\alpha_t + H_t\epsilon_t.
\end{aligned}
\tag{11}
$$

where both $X_t = 0$ and $W_t = 0$ when $t \leq i-k$ or $t > i$. There is no obvious analogue in a leave-k-out framework. Using an adaptation of a result in De Jong and Penzer (1998), it can be shown that, $D = (D_1', \ldots, D_n')'$ in the stacked form (10) of model (11) is given by

$$
D_t = \begin{cases} 0, & t = 1, \ldots, i-k, \\ X_t^* + Z_t \sum_{j=i-k+1}^{t-1} T_{t-1,j+1} W_j^*, & t = i-k+1, \ldots, i, \\ Z_t \sum_{j=i-k+1}^{i} T_{t-1,j+1} W_j^*, & t = i+1, \ldots, n, \end{cases}
\tag{12}
$$

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 $t \leq j$. For $t = i-k+1, \ldots, i$, X_t^* has column dimension equal to the row dimension of $\delta_{(I)}$ and $(t-i+k)^{\text{th}}$ block column entry X_t with all other entries zero. W_t^* is defined similarly.

Generalized least squares and some manipulation yields

$$
\hat{\delta}_{(I)} = S_{(I)}^{-1} s_{(I)}
$$

where

$$
s_{(I)} = \begin{pmatrix} X_{i-k+1}' u_{i-k+1} + W_{i-k+1}' r_{i-k+1} \\ \vdots \\ X_i' u_i + W_i' r_i \end{pmatrix}
$$

and $S_{(I)} = \mathrm{Var}(s_{(I)})$. A statistic to test for the presence of k shocks is given by

$$
\tau_{i,k}^2 = \hat{\delta}_{(I)}' \mathrm{Var}(\hat{\delta}_{(I)})^{-1} \hat{\delta}_{(I)} = s_{(I)}' S_{(I)}^{-1} s_{(I)}.
$$

Assuming Gaussianity and that all hyperparameters of the model are known, this τ^2-statistic is χ_p^2 under the null, where p is the rank of $S_{(I)}$. Evaluating $\tau_{i,k}^2$ involves generating and inverting $S_{(I)}$, a problem analogous to that described in §3. We propose a solution which generalizes Proietti's pseudo-model given by equations (6) and (7). If $s_t = X_t' u_t + W_t' r_t$ then, for $t = i, \ldots, i-k+1$, we can think of s_t as generated by

$$
\begin{aligned}
s_t &= Q_t' \beta_t + X_t' F_t^{-1/2} \eta_t, \\
\beta_{t-1} &= L_t' \beta_t + Z_t' F_t^{-1/2} \eta_t,
\end{aligned}
\tag{13}
$$

where $Q_t = W_t - K_t X_t$. The filter associated with this form is, for $t = i, \ldots, i-k+1$,

$$
\begin{aligned}
w_t &= s_t - Q_t' b_t, & F_t^* &= Q_t' P_t^* Q_t + X_t' F_t^{-1} X_t, \\
& & K_t^* &= \left(L_t' P_t^* Q_t + Z_t' F_t^{-1} X_t \right) F_t^{*-1}, \\
b_{t-1} &= L_t' b_t + K_t^* w_t, & P_{t-1}^* &= L_t' P_t^* L_t + Z_t' F_t^{-1} Z_t - K_t^* F_t^* K_t^{*\prime},
\end{aligned}
\tag{14}
$$

where $b_i = 0$ and $P_i^* = N_i$. The filter defined by (14) can be viewed as a *generalized partition filter*. Running this filter orthogonalizes $s_{i-k+1}, \ldots, s_i$. This is equivalent to inverting the Cholesky decomposition of $S_{(I)}$.

6 Maximal put-k-shocks-in diagnostics

The results of the preceding sections allow statistics associated with patterns of k shocks to be evaluated efficiently. In practice, it is unlikely that the exact shape of the intervention is known prior to implementing the diagnostics. Therefore, maximal put-k-shocks-in statistics are useful. It can be shown that the maximal τ^2-statistic over all possible interventions is

$$\max_{X,W}(\tau_{i,k}^2) = \rho_{i,k}^2 = r_i' N_i^{-1} r_i + \sum_{t=i-k+1}^{i} v_i' F_i^{-1} v_i. \tag{15}$$

This is the maximum value that any τ^2-statistic can take for k shocks from times $i-k+1$ to i inclusive. The proof of this result is given in the Appendix. For a Gaussian process, under the null hypothesis of no shocks, $\rho_{i,k}^2 \sim \chi_{\ell k+m}^2$ where m is the dimension of the state vector. Result (15) is appealing as it indicates that maximal statistics can be generated directly from KFS output without the need for additional filtering.

There are several distinct collections of k shocks for which the maximum is attained. Perhaps the most convenient and readily interpretable is a collection of $k-1$ measurement shocks in addition to a shock to both measurement and transition equations at point i. Thus, $X_t = I$ for $t = i-k+1, \ldots, i-1$, $(X_i',\ W_i')' = I$, and $X_t = 0$ and $W_t = 0$ otherwise. This is intuitively appealing. The total of k measurement shocks can be thought of as deleting k observations. The transition shock at the end allows for a change in the dynamics of the process after a patch of k unusual points.

7 Estimating shock magnitude associated with k shocks

In §5 we establish that $\hat{\delta}_{(I)} = S_{(I)}^{-1} s_{(I)}$ where $\delta_{(I)} = (\delta_{i-k+1}', \ldots, \delta_i')'$ is the vector of shock magnitudes. We can avoid evaluating these quantities directly by exploiting the property of smoothing errors given by (4) applied to the pseudo-model defined by (13). By definition $S_{(I)} = \mathrm{Var}(s_{(I)})$, so $S_{(I)}^{-1} s_{(I)}$ can be thought of as smoothing errors from the smoother associated with the generalized partition filter (14). We conclude that $\hat{\delta}_{(I)} = x_{(I)} = (x_{i-k+1}', \ldots, x_i')'$ where, for $t = i-k+1, \ldots, i$, x_t is generated by

$$\begin{aligned} x_t &= F_t^{*-1} w_t - K_t^{*\prime} q_t, & M_t^* &= F_t^{*-1} + K_t^{*\prime} N_t^* K_t^*, \\ q_{t+1} &= Q_t x_t + L_t q_t, & N_{t+1}^* &= Q_t F_t^{*-1} Q_t' + L_t^{*\prime} N_t^* L_t^*, \end{aligned} \tag{16}$$

where $L_t^* = L_t' - K_t^* Q_t'$ and the smoother is initialised at $q_{i-k+1} = 0$ and $N_{i-k+1}^* = 0$. When we have k measurement shocks as in §4, the partition smoother implicitly inverts a part of Σ^{-1}.

For the maximal case the partition smoother reduces to

$$\begin{aligned} x_t &= v_t - Z_t q_t, & M_t^* &= F_t + Z_t N_t^* Z_t', \\ q_{t+1} &= T_t q_t - K_t v_t, & N_{t+1}^* &= K_t F_t K_t' + T_t N_t^* T_t' \end{aligned} \tag{17}$$

for $t = i-k+1, \ldots, i-1$ and

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

All of the system matrices involved in partition smoothing for the maximal case come directly from KFS output. The algorithm is a straightforward consequence of the proof of (15) given in the Appendix.

8 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

$$y_t = \mu_t + \gamma_t + \varepsilon_t, \qquad \{\varepsilon_t\} \sim \mathrm{NID}(0, \sigma_\varepsilon^2), \tag{18}$$

where NID denotes normal and independently distributed. Here μ_t is the trend, γ_t is the seasonal component and ε_t is a error term. The trend is taken to be locally linear

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

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

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

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 $\mathrm{NID}(0, \sigma_\omega^2)$ error terms. All of the error terms in the model are mutually independent.

8.1 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 $\rho_{i,k}^2$ in (19) with $\tau_{i,k}^2$ for the inclusion of k measurement shocks. For any given value of k, calculate $\rho_{i,k}^2$ statistics for $i = k+1, \ldots, n$. Let the largest of these be λ_k so

$$\lambda_k = \max_i (\rho_{i,k}^2). \tag{19}$$

Repeat this, for a plausible set of values of k, to give $\lambda_1, \ldots, \lambda_q$, 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 $\lambda_1, \ldots, \lambda_q$ is efficient since it only requires null model KFS output. Each λ_k represents the most extreme event when adding k shocks and the sequence $\lambda_1, \ldots, \lambda_q$, 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 $k = 1, \ldots, q$, taking $\lambda_0 = 0$. If $\Delta\lambda_k < c_k$ for $k > k^*$ then k^* is the appropriate number of shocks, where $\{c_k\}$ is a set of critical values. By definition $\Delta\lambda_1 = \lambda_1 = \chi_{1+m}^2$ and, for $k > 1$, the upper 5% point of a χ_1^2 provides a guide to significance. Taking $c_1 = \chi_{1+m,0.95}^2$ and $c = 4$ provides sensible results in real examples although larger values may be chosen to account for simultaneous testing.

8.2 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 maximal statistic defined by (19). Let λ_k be an instance of the random variable Λ_k. Bonferonni criteria (David 1981) define an upper limit for $P(\Lambda_k > \lambda_k)$ under the null hypothesis of no shocks. This is interpreted as an approximation to the probability value (p-value) associated with out test for unusual behaviour. Our test statistic, λ_k, is the largest order statistic of a sample from $n+1-k$ correlated random variables each distributed as χ^2_{k+m} under H_0. The Bonferonni upper bound for $P(\Lambda_k > \lambda_k)$ is given by $p = (n+1-k)(1-F(\lambda_k))$ 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 $\rho^2_{i,k}$ 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 smoother associated with the generalised partition filter was given in §7. In §6 we establish that the maximal τ^2-statistic is attained for an intervention which consists of k measurement shocks followed by a shock to all of the elements of the state. The shapes associated with interventions 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 partition smoother (17), 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 partition smoother also generates standard errors which are used to scale the estimated magnitudes to approximate standard normals. We can use the largest element 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.

9 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).

9.1 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 the 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 $\tau^2_{i,2}$- 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.

The largest $\rho^2_{i,2}$-statistic is 58.51 occurring in 1970 quarter 4; see figure 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 to include, we evaluate the null model based, estimated magnitudes associated with shocks to the components of the state in 1970 quarter 4 using the partition smoother. These values are given in table 3. The largest impact is on the second seasonal component. Using (12), we establish that a shock to the second seasonal component at $t = i$ corresponds to an additive intervention of the form

$$D = \begin{cases} -1, & \text{for } \; t = i+2, \; i+6, \ldots, \\ 1, & \text{for } \; t = i+4, \; i+8, \ldots, \\ 0, & \text{otherwise.} \end{cases} \tag{20}$$

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 (20). 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.

9.2 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 $\tau_{i,6}^2$ 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 $k > 4$. The maximum value of $\rho_{i,4}^2$ 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 folowing 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 over 615 000 lower than before the war.

10 Conclusion

Leave-k-out diagnostics are useful for detecting patches of outliers. However, they may give misleading results if unusual behaviour is a series is persistent. For example, in §9.2, 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 the maximal put-k-shocks-in diagnostic, ρ^2, does not require additional filtering. Thus, put-k-shocks-in is a good generic procedure which reduces to leave-k-out where appropriate.

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^2_ε	1.823	0.232	0.767
σ^2_η	0.000	0.338	0.249
σ^2_ζ	0.008	0.005	0.005
σ^2_ω	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_ε^2	178.40	0.0000	373.69
σ_η^2	3787.8	2177.2	1326.1
σ_ζ^2	0.8297	1.6462	1.8940
σ_ω^2	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)

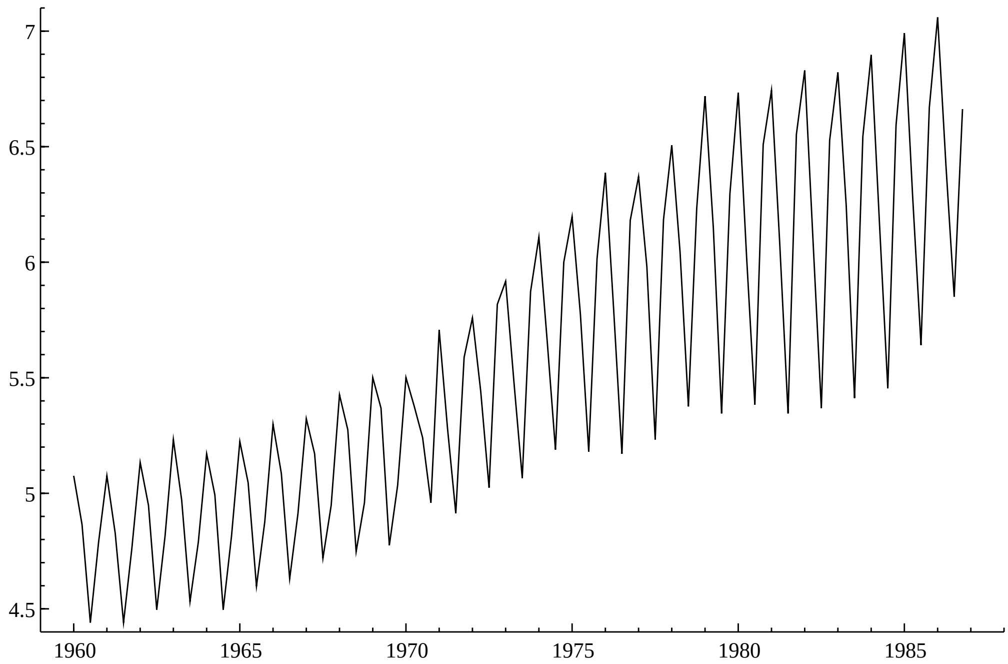


Figure 2: Gas example – scaled innovations

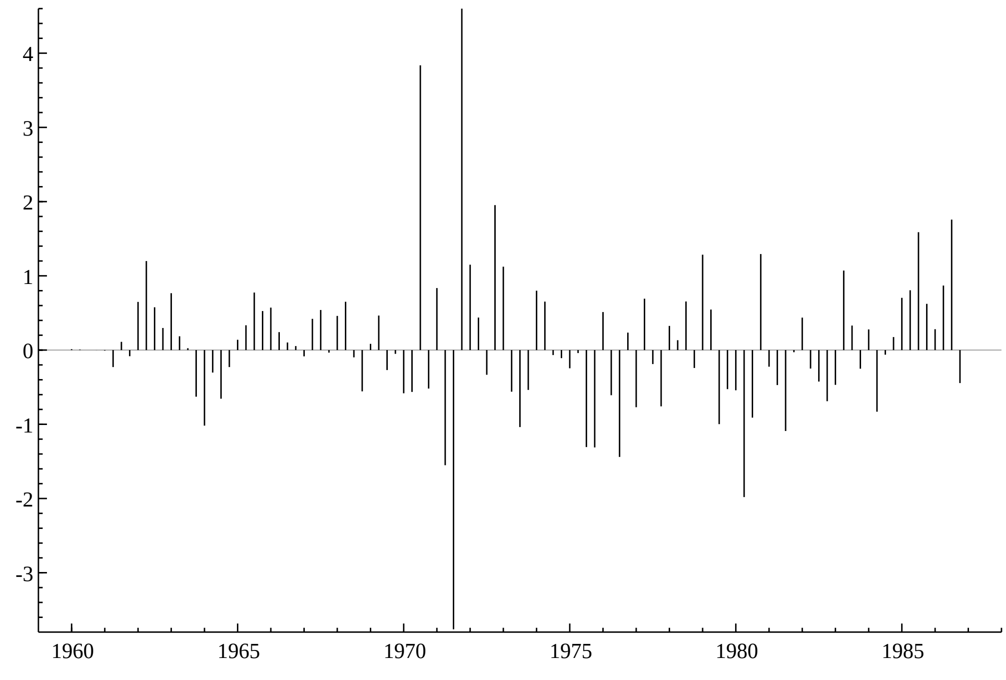


Figure 3: Gas example – leave-2-out τ^2-statistics

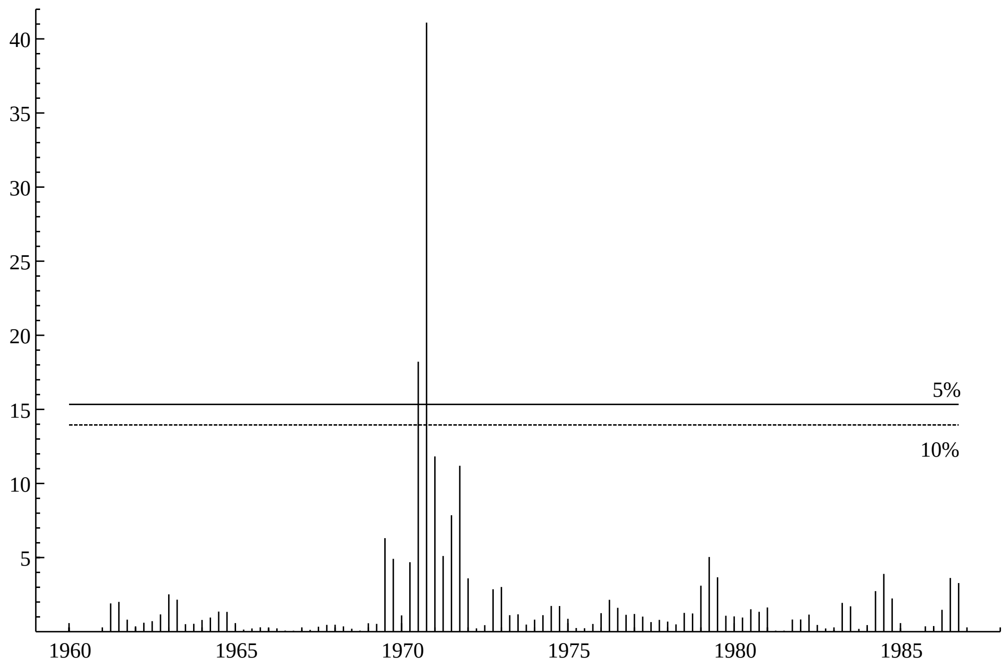


Figure 4: Gas example – put-2-shocks-in

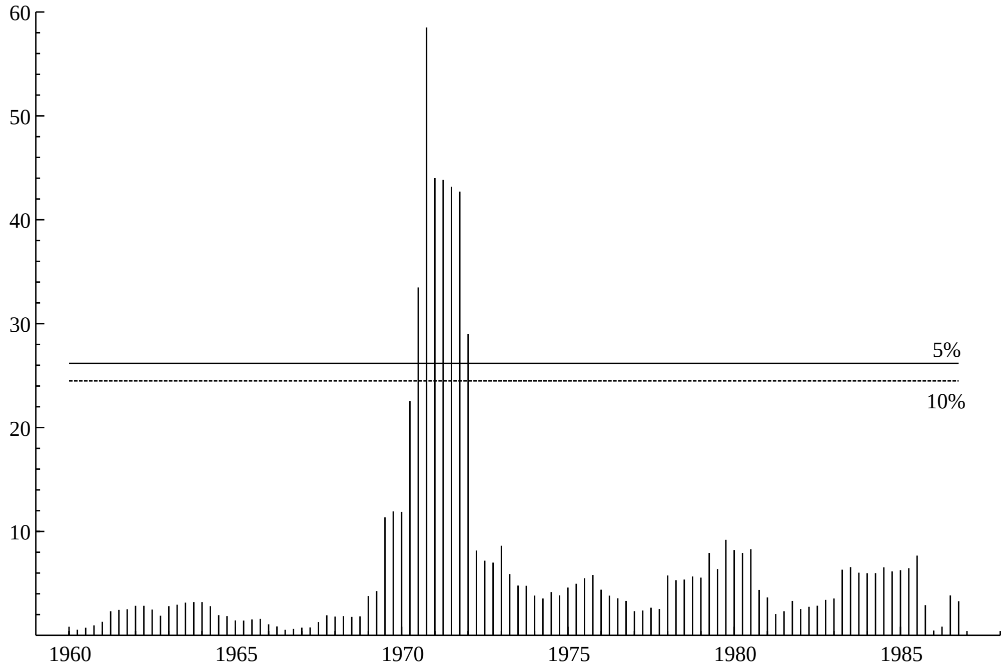


Figure 5: Air passengers example – monthly air passengers on UK airlines (thousands)

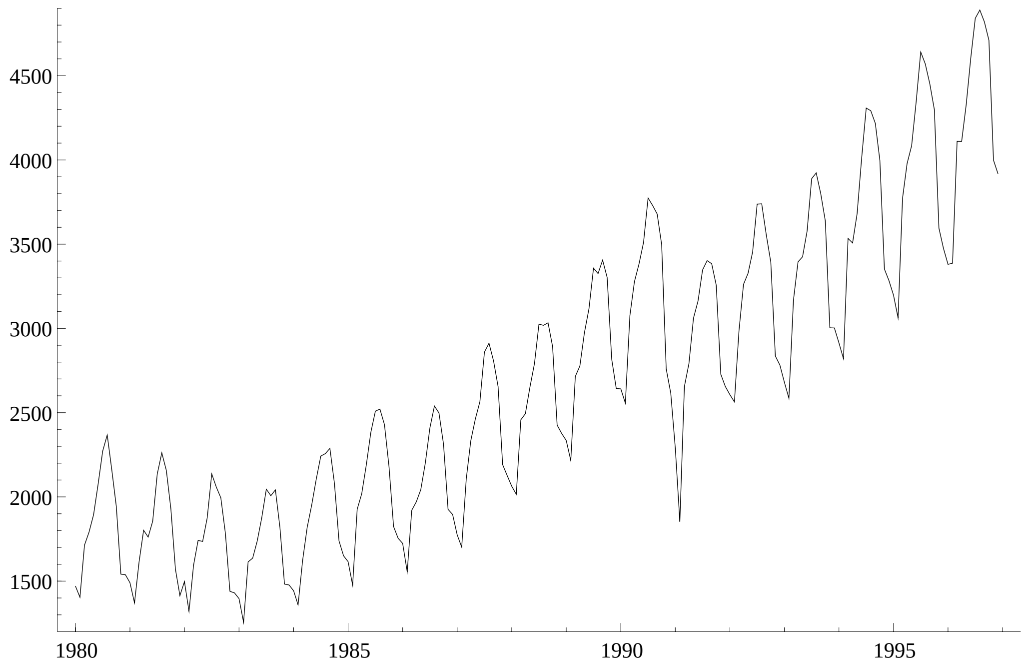


Figure 6: Air passengers example – scaled innovations

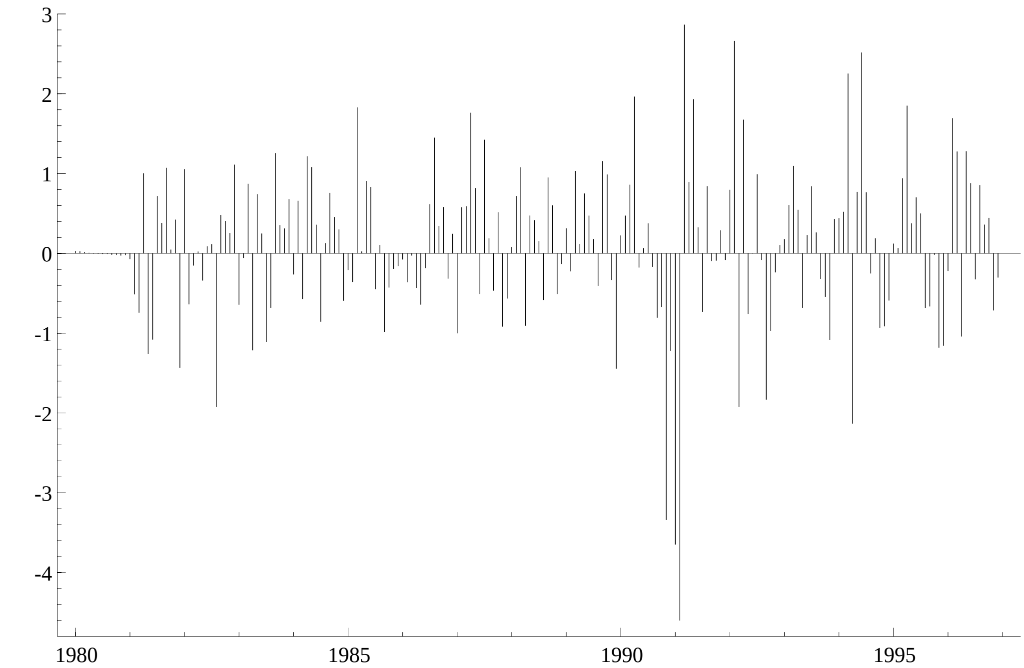


Figure 7: Air passengers example – leave-6-out diagnostics

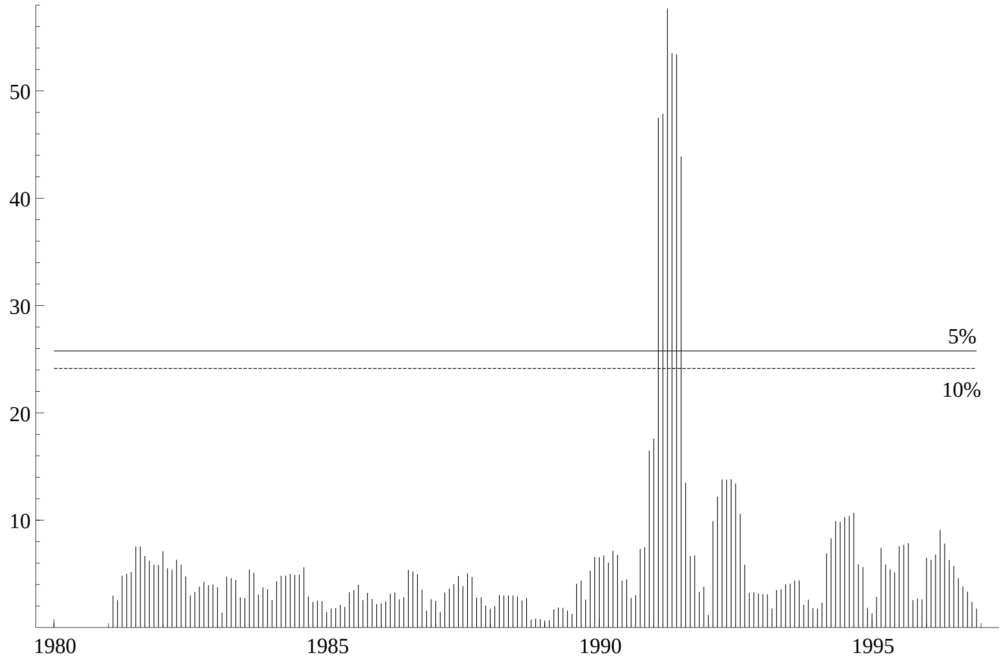
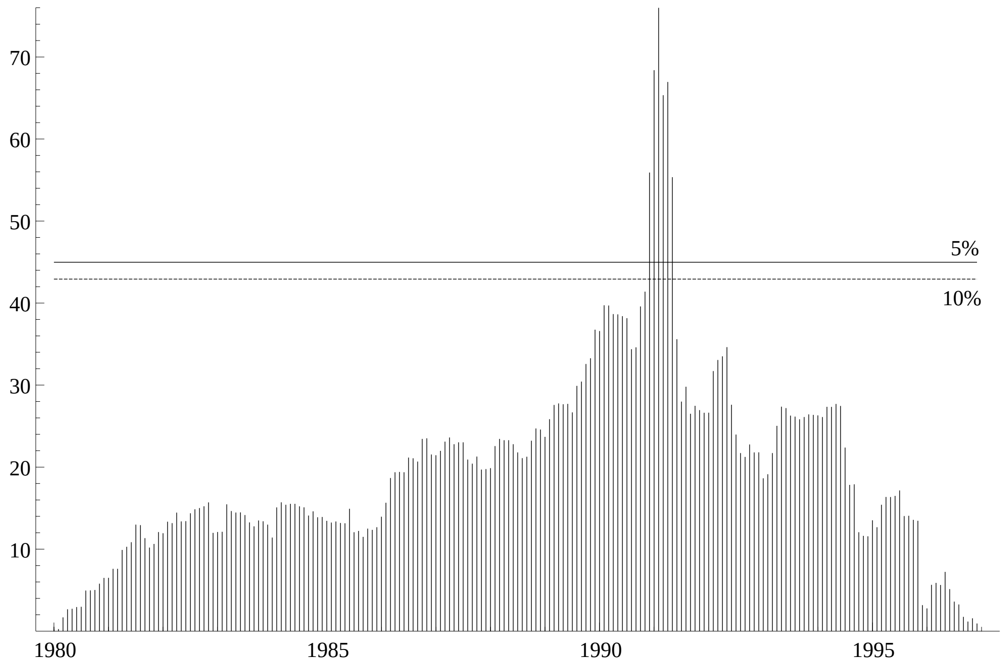


Figure 8: Air passengers example – put-4-shocks-in diagnostics



Appendix: Proof of expression (15) for $\rho_{i,k}^2$

For any i and k, the maximum $\tau_{i,k}^2$ value will be attained when we allow independent shocks to each element of the measurement and transition equations for $t = i - k + 1, \ldots, i$. Starting at $t = i$ we take $(X_i',\ W_i')' = I_{\ell+m}$ the $(\ell+m) \times (\ell+m)$ identity matrix. After some manipulation, the generalised partition filter, initialised with $b_i = 0$ and $P_i^* = N_i$, yields:

$$
\begin{aligned}
w_i &= s_i = X_i'u_i + W_i'r_i = \begin{pmatrix} u_i \\ r_i \end{pmatrix}, \\
F_i^* &= \begin{pmatrix} K_i'N_iK_i + F_i^{-1} & -K_i'N_i \\ -N_iK_i & N_i \end{pmatrix} \quad \Rightarrow \quad F_i^{*-1} = \begin{pmatrix} F_i & F_iK_i' \\ K_iF_i & N_i^{-1} + K_iF_iK_i' \end{pmatrix}, \\
K_i^* &= (Z_i',\ T_i'), \\
b_{i-1} &= r_{i-1}, \\
P_{i-1}^* &= 0.
\end{aligned}
$$

Note $Q_i = (-K_i, I_m)$ and that the expression for F_i^{*-1} is given by partition matrix inversion results. Consider the next in the iterative process. At $t = i - 1$ we have

$$
\begin{aligned}
w_{i-1} &= s_i - Q_{i-1}'r_{i-1} = X_{i-1}'F_{i-1}^{-1}v_i, \\
F_{i-1}^* &= X_{i-1}'F_{i-1}^{-1}X_{i-1}, \\
K_{i-1}^* &= (Z_{i-1}'F_{i-1}^{-1}X_{i-1})F_{i-1}^{*-1}.
\end{aligned}
$$

These quantities are independent of the choice of W_{i-1}. The maximum will be attained when each of the elements of the measurement equations is shocked independently, that is $X_{i-1} = I_\ell$. Thus, for maximal k shocks,

$$
\begin{aligned}
w_{i-1} &= F_{i-1}^{-1}v_{i-1}, & F_{i-1}^* &= F_{i-1}, \\
& & K_{i-1}^* &= Z_{i-1}', \\
b_{i-2} &= r_{i-2}, & P_{i-2}^* &= 0.
\end{aligned}
$$

Using induction, we establish that $w_t = F_t^{-1}v_t$, $F_t^* = F_t^{-1}$ and $K_t = Z_t'$ for $t = i-k+1, \ldots, i-1$. We know that

$$
\rho_{i,k}^2 = \max_{X,W}(\tau_{i,k}^2) = \sum_{t=i-k+1}^{i} w_t'F_t^{*-1}w_t.
$$

so, noting that

$$
w_i'F_i^{*-1}w_i = v_i'F_i^{-1}v_i + r_i'N_i^{-1}r_i,
$$

we have the required result.

Acknowledgements

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

References

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.

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. 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.

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.

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.

Proietti, T. (2000). Leave-k-out diagnostics in state space models. Submitted to JTSA, available from http://sfb.wiwi.hu-berlin.de/.

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