Diagnosing Shocks in Time Series

Piet de Jong and Jeremy Penzer
London School of Economics, London, WC2A 2AE, UK

August 11, 2000

Abstract

Departures from a fitted model can be represented by the introduction of shocks. In this paper efficient means of modelling aberant behaviour in times series are developed. The method allows test statistics for interventions to be computed from a single Kalman filter smoother run. New diagnostic statistics and a generalisation of jackknife estimation are put forward. Detailed discussion of the implications for structural models, ARIMA models and models with explanatory variables is given.

Key words: ARIMA models; Dynamic regression models; Explanatory variables; Interventions; Jackknife estimation; Kalman filter; Outliers; Regression; State space representation; Smoothing; Structural changes; Structural time series models.

1 Introduction

Time series processes are often characterised by their autocorrelation. However, many series are subject to external influences. These may take the form of a relationship with an observable quantity which is represented by the introduction of an explanatory variable. For example, much of the seasonal variability in the UK domestic gas demand is explained by seasonal weather patterns, particularly changes in temperature. Our interest lies with another type of exogenous process; that of intervention. Sudden or unexpected events often influence the values taken by a time series. These may take the form of natural disasters, strikes, wars or the introduction of new legislation. Interventions are, by nature, hard to characterize in terms of quantities which can be measured. They are often modeled by the introduction of dummy regression variables which represent the influence that an unquantifiable, external event has on the series.

The problem is conventionally viewed as one of diagnostic checking. Having specified and fitted a model, we may want to know whether there are movements in the series which are not adequately accounted for. Early methods to detect outliers in time series were based on assumptions of independence. Fox (1972) points out that this is inadequate and suggests an approach in which the structure of the series is taken into account. He develops two parametric models for outliers; additive, representing a single spurious observation, and innovative, in which a pulse shock to the noise sequence propagates through the observed series. Fox uses likelihood ratio criteria as a means of judging an outlier's significance. Box and Tiao (1975) include the possibility of step function input as a means of generating intervention structures.

Iterative procedures for detecting additive and innovative outliers and distinguishing between them are proposed by Tsay (1986) and Chang, Tiao and Chen (1988). Tsay (1988) extends his method to include level shifts and changes of variance. All of the methods described above are based on the use of autoregressive integrated moving average (ARIMA) models. Many authors, including Tsay (1986) and LeFrançois (1991) point out that the presense of outliers or structural changes can introduce serious bias in the sample autocorrelation function leading to problem with model identification. Balke (1993) establishes that, using Tsay's approach, level shifts may be labelled as innovative outliers. Harvey and Durbin (1986) provide a practical example of intervention analysis using structural models. Details of the general approach are given by Harvey (1989). The use of smoothed disturbances as a means of diagnostic checking is put forward by Harvey and Koopman (1992).

Outlier detection in regression is often carried out using deletion diagnostics (Cook 1977, Cook & Weisberg 1982, Atkinson 1985). This involves comparison of the estimate of the regression parameter for the model fitted to the full data set with an estimate based on the data with a single observation removed. Statistics, such as Cook's distance (Cook 1977), can be used to identify outlying points. In time series processes, aberant observations sometimes occur in patches. In order to account for this and to detect stretches of over influential observation, Bruce and Martin (1989) put forward leave-k-out diagnostics. The parameters of the model fitted to the full data set are compared with those generated by fitting the model to the data when a stretch of k points are taken to be missing. This approach is somewhat limited. There are many types of intervention, such as level shifts and slope changes, which do not fit into the leave-k-out approach. The method also requires a great deal of computationally intensive parameter re-estimation. These problems are, to a certain extent, addressed by Atkinson, Koopman and Shephard (1995) who use score statistics to approximate the change in hyperparameters when interventions are introduced to a series.

Diagnostic checking is usually one step in the model building cycle, the other steps being identification and estimation. Unsatisfactory diagnostics lead to changes in the null model and re-estimation of all hyperparameters. Recently a number of methods have been proposed (McCulloch & Tsay 1993) in which all hyperparameters, including those associated with the location and size of shocks, are estimated simultaneously within a single framework. The models used are state space models with the addition of indicator variables, where the indicators flag time points with large variance parameters. Estimation is based on Markov Chain Monte Carlo methods using Bayesian priors, an approach significantly more complex and computation expensive than conventional parameter estimation. The method also demands a great deal of user sophistication. The diagnostic tools described in this paper can be used to suggest appropriate generalizations of the standard state space models. Once the generalized models are estimated, they are, given the indicator variables, standard state space models and can be checked using the methods described below.

Interventions which model single outlying points represent a special case. The associated statistics can be generated by sucessively jackknifing each observation in the series. This fact is recognized by many authors in a number of different contexts. Jacknifing results for stationary, infinite sample time series are put forward by Whittle (1984). For finite sample state space models De Jong (1988a, 1989) gives the relevant formulae. It is clear from this work that the statistics associated with single outliers can be generated from one run of the Kalman filter smoother (KFS). Harrison and West (1991) give special cases of jackknifing results.

Diagnostic checking implicitly involves comparison of a fitted null model to an alternative. The alternative model should reflect nature of the inadequacy which we suspect in the null. We demonstrate that the addition of shocks can be used to model a large range of potential structural changes. This paper establishes that all statistics associated with these interventions can be generated from a single null model run of the KFS. The null model can be any time series model that has a state space representation. Shocks are assumed to enter linearly at any point along the length n of the time series. Diagnostics include the size and variance of the shock, as well as the impact of the assumed shocks on state, signal, disturbance, regression, error variance and other estimates. Thus, a single null model run of the KFS is informative for an large array of alternative models, where the alternative involves the addition of a shock at any of the n points along the time series.

The layout of this paper is as follows. The next section defines the terminology of shocks, interventions, impulse response functions and related concepts. The estimation of shocks is discussed in section 3 where we also introduce the key concepts of *smoothations* and *intervention contrasts*. Intervention contrasts play the same role in estimating shock effects as contrasts play in the estimation of treatment effects in the conventional design of experiments. The state space model, filter and smoother are given in section 4. Section 5 establishes that intervention contrasts, relevant to the diagnosis of outliers and structural changes, can be defined in terms of null model KFS output. New diagnostic statistics are put forward in section 6. Measurement interventions modelling single outlying values have particular properties. These are discussed, with reference to jackknife estimation, in section 7. This work is extended to the general shock case in section 8. Examples of the application of our results to structural, ARIMA and periodic models are given in section 9. Models with explanatory variables, also known as dynamic regression models, are discussed in Section 10. In section 11 the general *composite* intervention case is discussed. Section 12 includes a brief discussion of further work. All proofs are given in the appendix.

2 Shocks and Interventions

Typically, time series analysis is conducted within a parametric framework of which the model is a key component. A large number of time series models have been suggested in the literature. This paper is not concerned with the procedures used to choose models. Rather, it is assumed that there is a model which is thought, at least initially, to be an appropriate representation of the process which generates the data $y = (y_1' \cdots y_n')'$. This model is referred to as the null model. Under the null, the data are thought of as a function

of random disturbances or errors. Explanatory variables, such as mean effects or exogenous observed series, may also be included. The relationship is written

$$y = \text{null model}.$$

The null model can be any one of a large class of models conventionally used in time series. These include ARIMA, structural models and regression models with autocorrelated errors. All such models are linear in the sense that they linearly transform disturbances and, if present, explanatory variables, into observation y.

We want to check whether there are features in the data which are not adequately explained by the null model. If these features take the form of sudden or unexpected movements in the series we can model them by the addition of shocks,

$$y = \text{null model} + \text{shock effects}.$$

A shock is an event which takes place at a particular point in the series. Its defining characteristics are location and magnitude. In time series the effect of a shock is not confined to the point at which it occurs. It is useful to distinguish between the *shock impact*, which is a shock's instantenous effect, and the subsequent effects, called the *after shock*. The after shock is determined by the form of the null model since it is the dynamics of the null model which determines how the shock transmits or manifests itself in subsequent observations.

Departures from the null are modeled by the addition of shocks. The intervention model describes how these shocks manifest themselves in the observations. An intervention is characterised by its origin, which is the location of the first shock, its shape and its magnitude. This can be represented by

$$y = \text{null model} + D(i)\delta.$$

The origin of the intervention is i and the magnitude is δ. $D(i)$ is the *intervention signature* which defines the intervention's shape. An intervention has no impact prior to its origin, so $D_t(i) = 0$ for $t < i$. Here $D_t(i)$ is the row, or in the case of vector time series, row block, of $D(i)$ corresponding to y_t. The shock impact and after shocks are then $D_i(i)\delta$ and $D_{i+1}(i)\delta, D_{i+2}(i)\delta, \ldots$ respectively. In the Econometrics literature these are called the *impact* and *delay multipliers*. This paper deals with methods for identifying the origin i and estimating the magnitude δ and shock effects.

Figure 2 illustrates the shape of four common interventions signatures. The simplest type of intervention models a single outlying value. These measurement interventions take the form

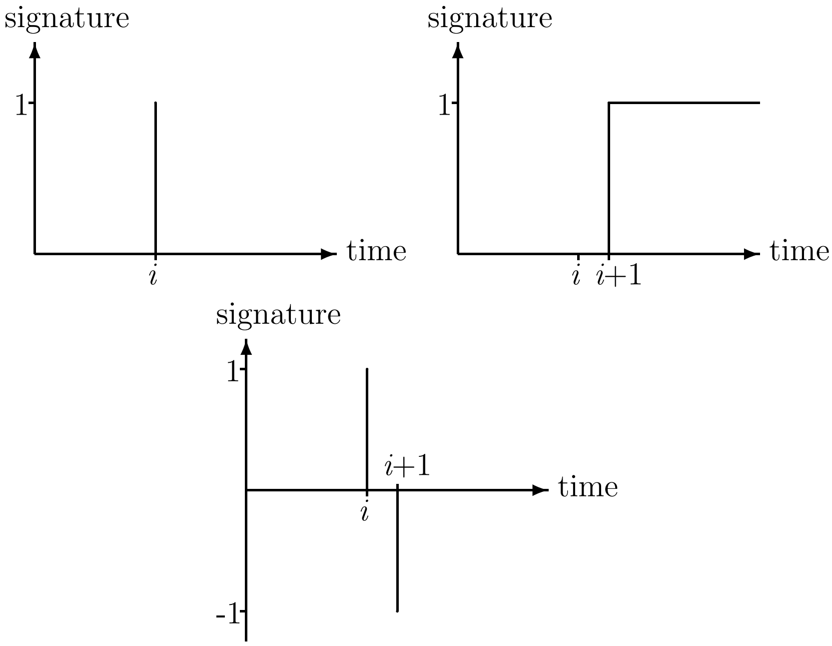

Figure 1: The signature for measurement, level and switch interventions

of a single shock. The effect of this shock is all instantaneous, there is no after shock. The shock signature is given by

$$D_t(i) = \begin{cases} 1, & t = i, \\ 0, & t \neq i. \end{cases}$$

Series A of Box, Jenkins and Reinsel (1994) and the Latin American exports data from Atkinson, Koopman and Shephard (1995) are examples of series with measurement outliers.

Another common intervention is a level shift. This is characterised by a permanent shift in the mean of the series and is modeled using the signature

$$D_t(i) = \begin{cases} 0, & t \leq i, \\ 1, & t > i. \end{cases}$$

Series with possible level shifts include the seat belt data in Harvey and Durbin (1986) and Nile data, see for example Balke (1993). The third intervention displayed in figure 2 is a switch intervention. This is characterised by consecutive extreme values either side of the mean of the series. They may arise from an increase in production after a strike or a collapse

in stock values after a sudden rise. The signature is

$$D_t(i) = \begin{cases} 0, & t = 1, \ldots, i-1, \\ 1, & t = i, \\ -1, & t = i+1, \\ 0, & t = i+2, \ldots, n. \end{cases}$$

Many practical departures from a null model can be represented by a single shock. We distinguish these cases by saying that they can be modeled using a *simple intervention*. Situations in which several shocks with different origins are required to generate the appropriate signature are referred to as *composite interventions*. In this case interventions are the cummulative effect of several shocks. The number of shocks required to represent a given intervention is model dependent. For example, a level intervention can be introduced into a structural, local linear trend model by a single shock. However, for a stationary ARMA model a composite intervention is required. Simple shock structures are desirable for economy of representation and ease of computation. More general null models allow interventions to be represented with fewer shocks. In some cases it is beneficial to use models which incorporate features such as non-stationarity or seasonality rather than trying to impose interventions with these properties via a large number of shocks.

Another useful distinction is between interventions which only affect a fixed set of points and those whose after shocks manifest themselves in all of the observations subsequent to their introduction. A *transient intervention* is characterised by a signature matrix whose entries $D_t(i)$ tend to zero as t increases. An intervention which does not have this property is *persistent*. Clearly, measurement and switch interventions are transient, whereas a level shift is persistent. However, a temporary level shift can be defined using the framework above. Further examples are given in sections 5, 9, 10 and 11.

3 Estimating and Testing Intervention Effects

In section 2 we described the representation of interventions via $D(i)$ and its associated scale parameter δ. The generalised least squares (GLS) estimate of the intervention parameter $\hat{\delta}_i$ can be used to devise statistics to measure the significance of an intervention at $t = i$. The null model described in section 2 defines the covariance matrix $\sigma^2\Sigma = \mathrm{cov}(y)$. Standard GLS arguments yield

$$\hat{\delta}_i = \{D(i)'\Sigma^{-1}D(i)\}^{-1}D(i)'\Sigma^{-1}y, \quad \mathrm{cov}(\hat{\delta}_i) = \sigma^2\{D(i)'\Sigma^{-1}D(i)\}^{-1}. \tag{1}$$

By noting that $\sigma^2 D(i)'\Sigma^{-1}D(i) = \text{cov}\{D(i)'\Sigma^{-1}y\}$ equation (1) simplifies to

$$\hat{\delta}_i = S_i^{-1}s_i, \quad \text{cov}(\hat{\delta}_i) = \sigma^2 S_i^{-1}, \tag{2}$$

where

$$s_i = D(i)'\Sigma^{-1}y, \quad S_i = D(i)'\Sigma^{-1}D(i) = \sigma^{-2}\text{cov}(s_i). \tag{3}$$

The s_i are referred to as the *intervention contrasts*. They are sufficient statistics for estimating the intervention effects. The intervention contrasts are the analogue of treatment contrasts for estimating treatment effect in the standard design of experiments. If the data is uncorrelated, Σ is a diagonal matrix, allowing Σ^{-1} and thus $\hat{\delta}_i$ to be calculated directly from (2). This is equivalent to the usual approach, regressing the data y on the intervention effect $D(i)$. In the extreme case, where $\Sigma = I$ and $D(i)$ is 0 everywhere except in a single position where it is one, the estimate of δ is the observation y_i.

The hypothesis of no shock, $\delta = 0$, is tested with the usual statistic,

$$\hat{\delta}_i'\{\text{cov}(\hat{\delta}_i)\}^{-1}\hat{\delta}_i = \sigma^{-2}s_i'S_i^{-1}s_i. \tag{4}$$

In practice, σ^2 is replaced by the normal based maximum likelihood estimate (MLE). The MLE of σ^2 assuming $\delta = 0$ is given by $\hat{\sigma}^2 = (y'\Sigma^{-1}y)/n$. This estimate is adjusted to give the MLE under the hypothesis that $\delta \neq 0$, assuming that all other hyperparameter estimates remain fixed; $\hat{\sigma}_i^2 = \hat{\sigma}^2 - n^{-1}s_i'S_i^{-1}s_i$. Substituting this adjusted estimate into (4) yields the test statistic

$$\tau_i^2 \equiv \hat{\sigma}_i^{-2}s_i'S_i^{-1}s_i = \{\hat{\sigma}^2(s_i'S_i^{-1}s_i)^{-1} - n^{-1}\}^{-1}, \tag{5}$$

which has an approximate χ_p^2 distribution where p is the rank of S_i. The estimate of the standard error of each component of $\hat{\delta}_i$ is given by the square root of the appropriate diagonal entry in $\sigma^{-2}S_i^{-1}$. Dividing each component by its estimated standard error gives a statistic which is analogous to the regression t-statistic.

Practical implementation of the methods requires specification of $\text{cov}(y) = \sigma^2\Sigma$ and the intervention signature $D(i)$. These are modeled using a state space description of the data y. An efficient method for computing the statistic τ_i^2 or the t-statistic is also crucial. Defining the *smoothation vector* $u = \Sigma^{-1}y$, equation (3) yields

$$s_i = D(i)'u. \tag{6}$$

In section 5 we establish that the smoothations u_t, all intervention contrasts s_i and $\text{cov}(s_i) = \sigma^2 S_i$ corresponding to a wide class of shock signatures can be generated with a single pass of the Kalman filter smoother. Thus, all of the interventions statistics τ_i^2 can be computed

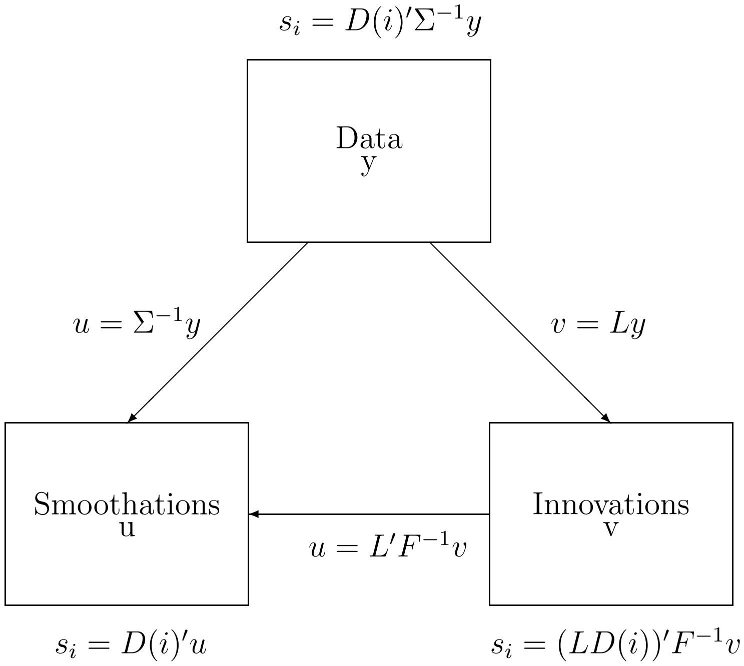

Figure 2: The relationship between observations, innovations and smoothations

from smoother output. Our method is direct, transparent and, in addition to computational benefits, provides insights into the nature of the detection process. It allows the simultaneous testing of a wide variety of shock effects.

Established approaches (Tsay 1986, Atkinson, Koopman & Shephard 1997) to computing (3) and estimating δ involve transforming the observations and the signature matrix. By applying the appropriate linear operations, the data can be transformed to the innovations, which are uncorrelated. Applying the same transformation to the regression variables $D(i)$ allows us to generate s_i and S_i and thus calculate the estimate $\hat{\delta}_i$ and its covariance matrix. Think about this procedure in the context of Kalman filtering. Using the lower triangular matrix L to denote the filtering transformation, write $v = Ly$, where $v = (\, v_1' \quad \cdots \quad v_n' \,)'$ is the vector of Kalman filter innovations. It is well known that $\Sigma^{-1} = L'F^{-1}L$ so $s_i = (LD(i))'F^{-1}v$ where $F = \sigma^{-2}\text{cov}(v) = \text{diag}\{F_1, \ldots, F_n\}$. The entries in $LD(i)$ can be calculated by applying the filter to the columns of the regression variable $D(i)$. Using this approach, for each origin i and each type of intervention, the columns of $D(i)$ must be filtered.

4 Observations, Innovations and Smoothations

The implementation of the methods outlined above exploits the recursive definition of the observations, innovations and smoothations. For the observations y the recursive definition is the state space model. This a flexible representation encompassing all linear time series models used in practice. The null state space model for the series $\{y_t : t = 1, \ldots, n\}$ is

$$y_t = Z_t \alpha_t + G_t \varepsilon_t, \tag{7}$$
$$\alpha_{t+1} = T_t \alpha_t + H_t \varepsilon_t, \quad t = 1, \ldots, n, \tag{8}$$

where $\varepsilon_t \sim (0, \sigma^2 I)$, $\alpha_1 \sim (a_1, \sigma^2 P_1)$ and the ε_t and α_1 are mutually uncorrelated. The system matrices Z_t, T_t, G_t and H_t are deterministic quantities which, as the notation indicates, may vary over time. Equations (7) and (8) serve to define $\text{cov}(y) = \sigma^2 \Sigma$.

Innovations $\{v_t : t = 1, \ldots, n\}$ are defined recursively in terms of the observations y via the usual Kalman filter. For $t = 1, \ldots, n$,

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

where $L_t = T_t - K_t Z_t$ and $J_t = H_t - K_t G_t$. The innovations v_t are orthogonal and

$$v_t = y_t - E(y_t | y_{t-1}, \ldots, y_1), \quad \sigma^2 F_t = \text{cov}(y_t | y_{t-1}, \ldots, y_1).$$

Here conditional expectation is defined in the linear predictor sense. Conditional covariance denotes the covariance of the prediction error. Non-stationary models require diffuse initial conditions. These can be handled using adaptations to the recursion (Ansley & Kohn 1985, Ansley & Kohn 1990, De Jong 1988b, De Jong 1991). The normal based MLE of σ^2 is given by $\hat{\sigma}^2 = n^{-1} \sum_{t=1}^{n} v_t' F_t^{-1} v_t$ which is conveniently accumulated in parallel with (9).

De Jong (1988a) and Kohn and Ansley (1989) put forward a set of efficient recursions which can be used to calculate all of the quantities normally associated with smoothing. The recursions are initialised with $r_n = 0$ and $N_n = 0$ and then, for $t = n, \ldots, 1$,

$$r_{t-1} = Z_t' F_t^{-1} v_t + L_t' r_t, \quad N_{t-1} = Z_t' F_t^{-1} Z_t + L_t' N_t L_t. \tag{10}$$

The combination of the filtering forward pass and smoothing back pass is referred to as the Kalman filter smoother (KFS). The smoothing recursions (10) are used to generate the smoothations u_t and associated covariances matrices $\sigma^2 M_t$.

Theorem 4.1 *The smoothations u_t and* $\mathrm{cov}(u_t) = \sigma^2 M_t$ *satisfy*

$$\begin{aligned} u_t &= F_t^{-1} v_t - K_t' r_t, \qquad M_t = F_t^{-1} + K_t' N_t K_t, \\ r_{t-1} &= Z_t' u_t + T_t' r_t, \qquad N_{t-1} = Z_t' F_t^{-1} Z_t + L_t' N_t L_t, \quad t = n, \ldots, 1, \end{aligned} \tag{11}$$

where $r_n = 0$ and $N_n = 0$. Further $u_t = M_t\{y_t - E(y_t|y^t)\}$ and $M_t = \{\sigma^{-2}\mathrm{cov}(y_t|y^t)\}^{-1}$, where y^t is y excluding y_t.

5 Key Results for Simple Interventions

Section 2 shows how interventions can be represented by the introduction of shocks. In a state space model these shocks may affect both measurement and transition equations. This leads to the general specification

$$y_t = X_t \delta + Z_t \alpha_t + G_t \varepsilon_t, \tag{12}$$

$$\alpha_{t+1} = W_t \delta + T_t \alpha_t + H_t \varepsilon_t, \tag{13}$$

where X_t and W_t model the shock at time t and are referred to as the *shock design*. We state a key result which allows us to generate the GLS estimate of δ using a single run of the KFS. In this section the result is stated for simple interventions. These can be generated by the introduction of a single shock. To emphasis this point, the intervention contrasts s_i are refered to as *shock contrasts*. For a simple intervention $X_t = 0$ and $W_t = 0$ for $t \neq i$. By substituting recursively in the equations for the model (12) and (13), we see that the signature corresponding to a single shock is

$$D_t(i) = \begin{cases} 0, & t = 1, \ldots, i-1, \\ X_i, & t = i, \\ Z_t T_{t-1,i+1} W_i, & t = i+1, \ldots, n. \end{cases} \tag{14}$$

where, $T_{j,t} = T_j \ldots T_t$ for $j \geq t$, $T_{t-1,t} = I$ and $T_{j,t} = 0$ for $j < t-1$. The general, composite intervention case is dealt with in section 11.

Theorem 5.1 *For a simple intervention with origin i, the shock contrast s_i and associated covariance matrix $\sigma^2 S_i$ are such that, for $i = 1, \ldots, n$,*

$$s_i = X_i' u_i + W_i' r_i, \quad S_i = X_i' F_i^{-1} X_i + Q_i' N_i Q_i, \tag{15}$$

where $Q_i = W_i - K_i X_i$ and the quantities u_i, r_i, F_i and N_i are generated by the KFS applied to the null model.

Using the framework which we have developed, the expression for the shock contrasts is fairly easy to derive. From the recursive representation of r_t given in theorem 4.1, it is immediately apparent that, r_t can be written in terms of the smoothations. For $t = 0, \ldots, n-1$,

$$r_t = \sum_{j=t+1}^{n} T'_{j-1,t+1} Z'_j u_j, \tag{16}$$

From (6), $s_i = D(i)'u$, so (16) yields

$$s_i = X'_i u_i + W'_i \sum_{j=i+1}^{n} T'_{j-1,i+1} Z'_j u_j = X'_i u_i + W'_i r_i,$$

as required. The proof of expression (15) for S_i is given in the appendix.

For any simple intervention, theorem 5.1 can be used to generate the GLS estimate of δ and the test statistic τ_i^2 for origins i from 1 to n. No assumptions about the nature of the intervention are made in filtering and smoothing process. This allows us to generate intervention statistics for any number of different types of simple intervention without doing any additional KFS runs.

The following corollary of theorem 5.1 provides concrete interpretation of the quantities involved in the smoothing recursions.

Corollary 5.1 *For a measurement intervention, where $X_i = I$ and $W_i = 0$,*

$$\begin{aligned} s_i &= u_i, & S_i &= M_i, \\ \hat{\delta}_i &= M_i^{-1} u_i & \operatorname{cov}(\hat{\delta}_i) &= \sigma^2 M_i^{-1}. \end{aligned}$$

For a state intervention in which each component of the state is shocked independently, that is $X_i = 0$ and $W_i = I$,

$$\begin{aligned} s_i &= r_i, & S_i &= N_i, \\ \hat{\delta}_i &= N_i^{-1} r_i & \operatorname{cov}(\hat{\delta}_i) &= \sigma^2 N_i^{-1}. \end{aligned}$$

For an intervention in which every component in the measurement and transition equation of the state space representation is shocked independently, that is $(\, X'_i \quad W'_i \,)' = I$,

$$\begin{aligned} s_i &= \begin{pmatrix} u_i \\ r_i \end{pmatrix}, & S_i &= \begin{pmatrix} M_i & -K'_i N_i \\ -N_i K_i & N_i \end{pmatrix}, \\ \hat{\delta}_i &= \begin{pmatrix} v_i \\ K_i v_i + N_i^{-1} r_i \end{pmatrix}, & \operatorname{cov}(\hat{\delta}_i) &= \begin{pmatrix} F_i & F_i K'_i \\ K_i F_i & K_i F_i K'_i + N_i^{-1} \end{pmatrix}, \end{aligned}$$

where K_i is the null model Kalman gain matrix.

Proof: The results for pure measurement and pure state shocks are immediate consequences of theorem 5.1 and equation (2) for the GLS estimate of δ. The final result for a combined shock follows from theorem 5.1 and the applyication of standard partioned matrix inverse results. $\square$

Corollary 5.1 stresses that, in common with regression, shock estimates depend on which shocks are present. For example, in the absence of other shocks, the estimate of a measurement shock is the standardized smoothation. If a state shock is present, the estimate of the measurement component of the shock is the innovation.

The mechanism by which interventions are generated can be viewed in several different ways. The model defined by (12) and (13) encompasses these alternative perspectives. In certain instances, it may be desirable to think of data nonconformities being caused by aberant state vectors. This leads to a model of the form

$$\begin{aligned} y_i &= Z_i(\alpha_i + \delta) + G_i\varepsilon_i, \\ \alpha_{i+1} &= T_i(\alpha_i + \delta) + H_i\varepsilon_i. \end{aligned}$$

In this situation $X_i = Z_i$ and $W_i = T_i$ so $Q_i = T_i - K_iZ_i = L_i$. Applying theorem 5.1 yields $s_i = r_{i-1}$ and $S_i = N_{i-1}$. This is equivalent to taking $X_{i-1} = 0$ and $W_{i-1} = I$, that is, a state shock at $t = i - 1$.

An alternative to measurement or state shocks is to model the intervention as an aberant disturbance,

$$\begin{aligned} y_i &= Z_i\alpha_i + G_i(\varepsilon_i + \delta), \\ \alpha_{i+1} &= T_i\alpha_i + H_i(\varepsilon_i + \delta). \end{aligned}$$

It is useful to have an interpretation for the shock contrast s_i and associated covariance matrix S_i. This is provided by the following theorem.

Corollary 5.2 *For a disturbance intervention, where $X_i = G_i$ and $W_i = H_i$,*

$$s_i = G_i'u_i + H_i'r_i = E(\varepsilon_i|y), \quad S_i = G_i'F_i^{-1}G_i + J_i'N_iJ_i = \{I - \sigma^{-2}\mathrm{cov}(\varepsilon_i|y)\},$$

where $E(\varepsilon_i|y)$ and cov$(\varepsilon_i|y)$ are calculated under the null model.

Harvey and Koopman (1992) refer to the scaled, smoothed disturbances as auxiliary residuals. De Jong (1988a) indicates that the smoothed measurement disturbance can be computed using (11). Koopman (1993) makes a similar observation about the smoothed state

disturbances. Harvey and Koopman use auxiliary residuals to detect aberant behaviour in time series. For structural component models, the auxiliary residuals have an appealing interpretation. However, thinking of diagnostics in terms of smoothed disturbances can be misleading. The coefficients of the shocks are constrained to be multiples of G_i and H_i. If a component has zero variance, useful diagnostic information which is available at the end of a smoothing run would be disregarded using the auxiliary residual approach.

Conventionally, interventions are represented by adding the signature and magnitude parameter to the measurement equation directly (Harvey & Durbin 1986, Harvey 1989, Atkinson et al. 1997). This leads to models of the form

$$\begin{aligned} y_t &= D_t\delta + Z_t\alpha_t + G_t\varepsilon_t, \\ \alpha_{t+1} &= T_t\alpha_t + H_t\varepsilon_t. \end{aligned}$$

It is clear that the model given by equations (12) and (13), is more general.

Outliers and structural changes can be detected by running an intervention along the series, that is, taking the starting point i between 1 and n (Tsay 1988, Atkinson et al. 1997). Statistics based on moving the origin of the intervention can be used to locate structural changes in the data. Running different type of interventions along the series allows different kinds of structural change to be distinguished. For state space models, Atkinson et al. argue that this requires the computation of $LD(i)$, the filtered columns of the $D(i)$ matrix. This is achieved by augmenting Kalman filter equations (9) with the additional recursions

$$V_t(i) = X_t(i) - Z_tA_t(i), \quad A_{t+1}(i) = -W_t(i) + K_tV_t(i) + T_tA_t(i),$$

where $V_t(i)$ is row block t of $LD(i)$. For each intervention type and for each origin i from 1 to n, the augmented filter is run, initialised with $A_i(i) = 0$. Using our approach only requires a single run of the KFS.

6 Test Statistics

The shock design matrices X_i and W_i determine the type of intervention resulting from a shock δ. One approach to detecting structural changes is to compute the statistics $\{\tau_i^2 : i = 1, \ldots, n\}$ for specified shock designs. A significant values of τ_i^2 suggest the location of a shock corresponding to a shock signature determined by the X_i and W_i. This diagnostic assumes that X_i and W_i are given, that is, the appropriate intervention type is known. Suppose,

for example, we suspect the presence of outliers and level shifts in data modelled by a local linear trend (see section 9). The statistic τ_i^2 would be generated for all values of i, first using $X_i = 1$, $W_i = 0$, to look for outliers, and then using $X_i = 0$, $W_i = (\,1 \quad 0\,)'$, to look for level shifts.

As an initial diagnostic procedure we can allow the data to determine the shock design X_i, W_i and the corresponding test statisic τ_i^2.

Theorem 6.1 *For given i, the maximum of $\rho_i^2 = s_i' S_i^{-1} s_i$ over all matrices (X_i', W_i') of fixed column dimension is*

$$\rho_i^{*2} = v_i' F_i^{-1} v_i + r_i' N_i^{-1} r_i, \tag{17}$$

where v_i, F_i, r_i and N_i are computed with the KFS applied to the null model. The maximum is attained when δ is scalar, $X_i = v_i$ and $W_i = K_i v_i + N_i^{-1} r_i$.

This result can be viewed from two different perspectives. If $(\,X_i' \quad W_i'\,)$ is a row vector, a given scalar shock is distributed across the components of $(\,y_i' \quad \alpha_{i+1}'\,)'$. The optimal distribution, in the sense of maximising the statistic ρ_i^2, is attained at $(\,X_i' \quad W_i'\,) = (\,v_i' \quad v_i' K_i' + r_i' N_i^{-1}\,)$. If $(\,X_i' \quad W_i'\,)$ is of full column rank, the shock design is equivalent to imposing independent scalar shocks on each of the components of $(\,y_i' \quad \alpha_{i+1}'\,)'$. Thus, the GLS estimate of the shock $\hat{\delta}_i$ in the full column rank case is equivalent to the optimal shock design $(\,X_i' \quad W_i'\,)'$ for the rank 1 case.

The statistic ρ_i^{*2} has useful interpretations. It is made up of two components $v_i' F_i^{-1} v_i$ and $r_i' N_i^{-1} r_i$. Under normality, these components are independent χ^2 random variables. From corollary 5.1, the term $r_i' N_i^{-1} r_i$ is the χ^2 statistic for testing a state shock in the absence of any other shock. Thus, $v_i' F_i^{-1} v_i$ is the χ^2 statistic for testing a measurement shock conditional on the presence of a state shock. In fact, the estimate of the measurement shock in this case is v_i. We can also consider ρ_i^{*2} as the sum of $u_i' M_i^{-1} u_i$, the χ^2 statistic for testing a measurement shock in the absence of any other shock, and $v_i' F_i^{-1} v_i + r_i' N_i^{-1} r_i - u_i' M_i^{-1} u_i$ which represents the χ^2 statistic for testing a state shock given a measurement shock.

The statistic ρ_i^{*2} can be scaled by dividing by $\hat{\sigma}^2$ or the adjusted estimate $\hat{\sigma}_i^2$. Notice that the test statistic τ_i^2 given by equation (5) has maximum value $\tau_i^{*2} = \{\hat{\sigma}^2 \rho_i^{*-2} - n^{-1}\}^{-1}$. A plot of τ_i^{*2} against i is an informative diagnostic tool. For each i, the plot shows the value of the test statistic under the shock design which has the greatest possible impact at that point. Large values of τ_i^{*2} indicate the location of potential shocks while the design matrices $X_i = v_i$ and $W_i = K_i v_i + N_i^{-1} r_i$ suggest the most plausible way that a shock enters the system.

Estimates for shock effects at different time points are correlated. The exact form of this correlation is given by the following theorem.

Theorem 6.2 *If $\hat{\delta}_i$ is the estimate of a shock at i and $\hat{\delta}_j$ is the estimate of a shock at j, for $i < j$*

$$\mathrm{cov}(\hat{\delta}_i, \hat{\delta}_j) = \sigma^2 S_i^{-1} Q_i' L_{j-1,i+1}' (Z_j' F_j^{-1} X_j + L_j' N_j Q_j) S_j^{-1}, \tag{18}$$

where, $L_{j,t} = L_j \ldots L_t$ for $j \geq t$, $L_{t-1,t} = I$ and $L_{j,t} = 0$ for $j < t - 1$.

Theorem 6.2 can be used to adjust levels of signficance when many tests are performed simultaneously.

7 Measurement Shocks

A measurement intervention is the addition of a shock to the measurement equation and models a single outlying value. It is represented in equation (12) and (13) by taking $X_i = I$ and $W_i = 0$. From corollary 5.1, $\hat{\delta}_i = M_i^{-1} u_i$ and so $\mathrm{cov}(\hat{\delta}_i) = \sigma^2 M_i^{-1}$. Using the expressions in theorem 4.1 yields

$$\hat{\delta}_i = y_i - E(y_i|y^i), \quad \mathrm{cov}(\hat{\delta}_i) = cov(y_i|y^i). \tag{19}$$

Thus the estimate of the shock is the jackknife residual at $t = i$, while the variance of the shock is estimated by the jackknife residual variance. The result is intuitively appealing since it states that the estimate $\hat{\delta}_i$ is the difference between the observation y_i and the estimate of y_i based on all the other data points. In Peña (1990) this is stated, in a less general context, as a well known result.

The measurement shock illustrates an important aspect of the method. The GLS approach to estimating δ is formally equivalent to treating δ as a random vector with infinitely large covariance matrix and constructing the linear predictor of δ given the data y. Thus $E(\delta|y) = S_i^{-1} s_i$ and $\mathrm{cov}(\delta|y) = \sigma^2 S_i^{-1}$ where the expectation and covariance are now computed with respect to the model which includes δ as diffuse random effect, that is, $y = D(i)\delta + w$ where $\mathrm{cov}(w) = \mathrm{cov}(y|\delta) = \sigma^2 \Sigma$. The expectation $E(y_i|y^i)$ and covariance $\mathrm{cov}(y_i|y^i)$ in (19) are computed with respect to the null model, that is, the model which does not contain any shocks. Using the framework in which δ is thought of as diffuse, (19) is equivalent to

$$E_a(\delta|y) = y_i - E_o(y_i|y^i), \quad \mathrm{cov}_a(\delta|y) = \mathrm{cov}_o(y_i|y^i), \tag{20}$$

where the subscripts indicate the model under which the expectation and covariance are computed: 'o' for null model and 'a' for alternative. Result of this type hold for general state space random vectors γ.

Theorem 7.1 *Suppose that the alternative model differs from the null by imposition of a measurement shock at $t = i$. Further let $E_a(\gamma|\delta) = \Gamma\delta$ and put $\Lambda_i = \sigma^{-2}\mathrm{cov}_o(\gamma, u_i)$. Then*

$$\begin{aligned} E_a(\gamma|y) &= E_o(\gamma|y^i) + \Gamma M_i^{-1}u_i = E_o(\gamma|y) + (\Gamma - \Lambda_i)M_i^{-1}u_i, \\ \mathrm{cov}_a(\gamma|y) &= \mathrm{cov}_o(\gamma|y^i) + \sigma^2\Gamma M_i^{-1}\Gamma' = \mathrm{cov}_o(\gamma|y) + \sigma^2(\Gamma - \Lambda_i)M_i^{-1}(\Gamma - \Lambda_i)', \end{aligned} \tag{21}$$

where u_i and M_i are given by the smoothing recursion (11).

The proof of this theorem follows from a more general result discussed in section 8. The role of the matrix Γ is as follows. Any state space quantity of practical interest can be written as a linear function of δ and the ε_ts. The matrix Γ indicates the manner in which γ depends on δ. If $\gamma = \delta$, then $\Gamma = I$, $E_o(\gamma|y^i) = 0$ and $\mathrm{cov}(\gamma|y^i) = 0$ so equations (21) reduce to (20). For state or disturbance estimation $\Gamma = 0$ and theorem 7.1 then implies

$$\begin{aligned} E_a(\gamma|y) &= E_o(\gamma|y^i) = E_o(\gamma|y) - \Lambda_i M_i^{-1}u_i, \\ \mathrm{cov}_a(\gamma|y) &= \mathrm{cov}_o(\gamma|y^i) = \mathrm{cov}_o(\gamma|y) + \sigma^2\Lambda_i M_i^{-1}\Lambda_i'. \end{aligned} \tag{22}$$

Thus when $\Gamma = 0$, the smoothed estimate of γ under the alternative is equal to the jackknife estimate. Equation (22) also gives the adjustments to null model smoothed estimates needed to yield smoothed estimates under the alternative.

Specific cases of interest are γ equal to the state α_t, the signal $Z_t\alpha_t$ or the disturbance ε_t, in particular for t near i. The following proposition specifies the form of Λ_i if $t = i$. The case $t \neq i$ is dealt with in section 8.

Proposition 7.1 *When $\gamma = \alpha_i$, $Z_i\alpha_i$ or ε_i,*

$$\Lambda_i = \sigma^{-2}\mathrm{cov}_o(\gamma, u_i) = \begin{cases} P_i Z_i' F_i^{-1}, & \gamma = \alpha_i, \\ G_i' F_i^{-1}, & \gamma = \varepsilon_i, \\ I - G_i G_i' F_i^{-1}, & \gamma = Z_i\alpha_i, \end{cases}$$

where the matrices are defined as in the Kalman filter smoother (9) and (11).

8 Generalized Jackknifing

Many of the established methods for dealing with outliers are based on statistics derived from jackknife residuals. In section 7 we establish that this approach is equivalent to the addition of a measurement shock and derive expressions for adjusting smoothed estimates. This technique is effective for individual outlying points. When dealing with more complex structural changes, simple jackknifing of the sample will not excise the aberant behaviour. In this section we establish that, for any shock, expectations and covariances under the alternative model can be computed from null model KFS output. Motivation is provided by the following example.

Consider the local linear trend model given by equations (24). The state vector at time t is given by $\alpha_t = (\,\mu_t \quad \beta_t\,)'$, the stack of the level and slope components. Suppose that the diagnostic statistics described in section 6 suggest the presence of a level shift at $t = i$. We want a quick means of judging how the introduction of a variable to model this structural change would affect our smoothed estimate of the level component. Thus, quantities such as $E_a(\alpha_{i-1}|y)$, $E_a(\alpha_i|y)$ and $E_a(\alpha_{i+1}|y)$ are of interest. The computation of various alternative model smoothed estimates is described below.

Theorem 8.1 *Suppose that s_i and $\hat{\delta}_i$ are defined with respect to shock design X_i, W_i. Further suppose that $E_a(\gamma|\delta) = \Gamma\delta$ and $\Lambda_i = \sigma^{-2}\mathrm{cov}(\gamma, s_i)$ then*

$$E_a(\gamma|y) = E_o(\gamma|y) + (\Gamma - \Lambda_i)\hat{\delta}_i, \quad \mathrm{cov}_a(\gamma|y) = \mathrm{cov}_o(\gamma|y) + (\Gamma - \Lambda_i)\mathrm{cov}_o(\hat{\delta}_i)(\Gamma - \Lambda_i)'. \quad (23)$$

where $\hat{\delta}_i$ is the GLS estimate of δ.

Note that equations (23) coincide with those in (21) if $\hat{\delta}_i = M_i^{-1}u_i$ and $\mathrm{cov}(\hat{\delta}_i) = \sigma^2 M_i^{-1}$, that is, if the shock is a measurement shock. Thus theorem 8.1 extends theorem 7.1 to more general shock designs.

The form of Λ_i depends on γ. Noteworthy cases include γ as the state vector α_t, the signal $Z_t\alpha_t$ or the disturbance ε_t. Values around $t = i$ are of interest since this is typically where the shock has most impact. The next result specifies the form of Λ_i in terms of null model KFS quantities. The proposition implies that all adjustments associated with introducing a shock to the system can be computed from the output of a null model run.

Proposition 8.1 *Define*

$$C_i = G_i' F_i^{-1} X_i + J_i' N_i Q_i, \quad R_i = Z_i' F_i^{-1} X_i + L_i' N_i Q_i,$$

then

$$\mathrm{cov}_o(\alpha_t, s_i) = \begin{cases} \sigma^2 P_t L'_{i-1,t} R_i, & t = 1, \ldots, i-1, \\ \sigma^2 P_i R_i, & t = i, \end{cases}$$

$$\mathrm{cov}_o(Z_t\alpha_t, s_i) = \begin{cases} -\sigma^2 G_t J'_t L'_{i-1,t+1} R_i, & t = 1, \ldots, i-1, \\ \sigma^2 (X_i - G_i C_i), & t = i, \end{cases}$$

and

$$\mathrm{cov}_o(\varepsilon_t, s_i) = \begin{cases} \sigma^2 J_t L'_{i-1,t+1} R_i, & t = 1, \ldots, i-1, \\ \sigma^2 C_i, & t = i, \\ \sigma^2 (G'_t F_t^{-1} Z_t + J'_t N_t L_t) L_{t-1,i+1} Q_i & t = i+1, \ldots, n. \end{cases}$$

The case $t > i$ is distinct since, provided $W_i \neq 0$, the state α_{t+1} carries the shock. In this case $\Gamma_{t+1} = T_{t,i+1} W_i = T_t \Gamma_{t-1}$ defines the shock signature on the state vector. Defining

$$\sigma^{-2}\mathrm{cov}(\alpha_{t+1}, s_i) \equiv \Lambda_{i,t+1} = T_t \Lambda_{i,t} + H_t \sigma^{-2}\mathrm{cov}(\varepsilon_t, s_i),$$

it is clear that

$$(\Gamma_{t+1} - \Lambda_{i,t+1}) = T_t(\Gamma_t - \Lambda_{i,t}) + H_t \sigma^{-2}\mathrm{cov}(\varepsilon_t, s_i).$$

This defines a recursion for computing the adjustment to any future smoothed state, relative to i, initialised at $t = i + 1$ using

$$\Gamma_{i+1} - \Lambda_{i,i+1} = W_i - (T_i P_i R_i + H_i C_i) = W_i - K_i X_i - P_{i+1} N_i Q_i = (I - P_{i+1} N_i) Q_i.$$

In turn $\mathrm{cov}(Z_t\alpha_t, s_i) = Z_t \mathrm{cov}(\alpha_t, s_i)$ for $t > i$. The input into the recursion is $\sigma^{-2}\mathrm{cov}(\varepsilon_t, s_i)$. Note that ε_t does not involve δ.

9 Simple Interventions in Practical Models

Many different interventions can be represented by a single shock to the transition and measurement equations. The precise effect of a shock is model dependent. We will consider examples drawn from two classes frequently used in practice, namely structural components (Harvey 1989) and autoregressive integrated moving average (ARIMA) models. The processes which we consider in this section can be given a time-invariant state space representation. For all t, we can write $T_t = T$ where T is constant over time. For an intervention with origin i, substituting $T_t = T$ into equation (14) yields the signature $D(i)$.

9.1 Local linear trend model

The local linear trend model is

$$
\begin{aligned}
y_t &= \begin{pmatrix} 1 & 0 \end{pmatrix} \alpha_t + \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \varepsilon_t, \\
\alpha_{t+1} &= \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \alpha_t + \begin{pmatrix} 0 & h_1 & 0 \\ 0 & 0 & h_2 \end{pmatrix} \varepsilon_t
\end{aligned} \tag{24}
$$

Here the first component of the state is interpreted as the level of the series and the second as the slope. Three types of shocks are immediately of interest: measurement shocks, dealt with in section 7, level shocks and slope shocks.

A level intervention corresponds to a sudden change in the level of the series. This can be modeled by the addition to the measurement equation of a step function with a single step at $t = i$. Alternatively, taking $X_i = 0$ and $W_i = \begin{pmatrix} 1 & 0 \end{pmatrix}'$ generates the appropriate shock signature. In this case s_i and S_i correspond to the first element of r_i and the first diagonal entry of N_i respectively. Changes in the slope of the series can be modeled using a slope intervention. Take $X_i = 0$, $W_i = \begin{pmatrix} 0 & 1 \end{pmatrix}'$ and substitute in equation (14) to yield the shock signature

$$
D_t(i) = \begin{cases} 0, & t = 1, \ldots, i, \\ (t - i - 1), & t = i + 1, \ldots, n. \end{cases}
$$

From theorem 5.1, s_i is now the second element of the vector r_i and S_i is the second diagonal entry in N_i.

Taking $X_i = 0$ and $W_i = I$, corollary 5.1 states that the estimate of the shock is $N_i^{-1} r_i$. This estimate is different from that given by separate level and slope shocks. If $\begin{pmatrix} X_i & W_i \end{pmatrix}' = I$ the estimates of the measurement and state components of the shock are v_i and $K_i v_i + N_i^{-1} r_i$ respectively. For the local linear trend model, a combined measurement and state shock is effectively equivalent to a disturbance shock. Thus, from corollary 5.2, an alternative expression for the estimate is $\{I - \sigma^{-2} cov(\varepsilon_i | y)\}^{-1} E(\varepsilon_i | y)$.

9.2 Autoregressive integrated moving average models

An autoregressive integrated moving average (ARIMA(p, d, q)) model can be written as

$$
\phi(B)(1 - B)^d y_t = \theta(B)\varepsilon_t,
$$

where B is the back-shift operator, $\phi(B) = 1 - \sum_{i=1}^{p} \phi_i B^i$, $\theta(B) = 1 + \sum_{i=1}^{q} \theta_i B^i$ and $\{\varepsilon_t\}$ is a sequence of independent $(0, \sigma^2)$ random variables. Consider the stationary case

when $d = 0$. The model can be given a time-invariant state space representation in which $Z = (1 \quad 0 \quad \cdots \quad 0)$, $G = 0$,

$$T = \begin{pmatrix} \phi_1 & 1 & 0 & \cdots & 0 \\ \phi_2 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \vdots & 0 & \cdots & 0 & 1 \\ \phi_m & 0 & \cdots & \cdots & 0 \end{pmatrix} \quad \text{and} \quad H = \begin{pmatrix} 1 \\ \theta_1 \\ \vdots \\ \vdots \\ \theta_{m-1} \end{pmatrix},$$

where $m = \max(p, q+1)$. An ARMA model in state space form immediately suggests two types of intervention which are of interest, namely measurement interventions and interventions corresponding to a shock to the state disturbance. Measurement interventions are described in section 7 above. A shock to the state disturbances can be modeled by taking $X_i = 0$ and $W_i = H$. Substituting in equation (14) yields

$$D_t(i) = ZT^{t-i}H = \sum_{j=0}^{q} \pi_{t-i-j}\theta_j. \tag{25}$$

where, for $k \geq 0$, π_k is the coefficient of B^k in the expansion of $1/\phi(B)$, and $\pi_k = 0$ for $k < 0$. The properties of interventions generated by shocks to the state disturbance are described in section 5. For both measurement and disturbance interventions we can apply theorem 5.1 and calculate estimates of δ, for $i = 1, \ldots, n$, using a single run of the KFS. Note that the final sum in (25) defines the coefficients in the polynomial expansion of $\theta(B)/\phi(B)$, that is, the coefficients in the infinite MA representation of the model.

We can relate our approach to that described in Tsay (1988). Using the back shift operator notation, Tsay considers models of the form

$$y_t = \frac{\theta(B)}{\phi(B)}\varepsilon_t + \delta\omega(B)\xi_t(i),$$

where $\xi_t(i)$ is an indicator function which takes the value 1 when $t = i$ and is 0 otherwise. The ratio $\theta(B)/\phi(B)$ defines the null model, while $\omega(B)\xi_t(i)$ defines the signature and δ is the scale parameter. Different types of interventions can be represented by appropriate choice of the polynomial $\omega(B)$. Tsay defines additive outliers, for which $\omega(B) = 1$, and innovative outliers, for which $\omega(B) = \theta(B)/\phi(B)$. These are equivalent to our measurement and state disturbance shocks respectively.

To estimate δ, Tsay rewrites the model as

$$\frac{\phi(B)}{\theta(B)}y_t = \delta\frac{\phi(B)}{\theta(B)}\omega(B)\xi_t(i) + \varepsilon_t,$$

where the left hand side defines $\tilde{v}_t$ the infinite sample null model innovations. Defining $\lambda_t(i) = \{\phi(B)/\theta(B)\}\omega(B)\xi_t(i)$, a quantity which corresponds to the filtered shock signature, yields

$$\tilde{v}_t = \delta\lambda_t(i) + \varepsilon_t.$$

Tsay (1988) regresses the $\tilde{v}_t$s on $\lambda_t(i)$ for each origin i. The values of $\lambda_t(i)$ are determined by the choice of $\omega(B)$. In the innovational outlier case $\omega(B) = \theta(B)/\phi(B)$ implying $\lambda_t(i) = \xi_t(i)$. In this instance, the estimate of δ is $\tilde{v}_i$. In the additive outliers case $\omega(B) = 1$ so $\lambda_t(i) = \phi(B)/\theta(B)$, the coefficients of the infinite AR representation.

These two cases can be used to compare Tsay (1988) with our approach. The first clear difference is that our approach is exact. Tsay's defines the innovations as $\{\phi(B)/\theta(B)\}y_t$. This infinite sample approximation will give poor results when the roots of $\theta(B)$, $\phi(B)$ $\omega(B)$ or $1/\omega(B)$ are close to the unit disc and when the data series is short. Secondly, our approach does not involve explicit regression for each origin i but achieves this implicitly by appropriate interpretation of the null model KSF output. The addition of shocks to the state space representation allows for a wider class of shock designs than those considered by Tsay. Finally, our methods apply to any model that can be cast in the state space form, not just ARIMA models.

When $d \geq 1$ the ARIMA(p, d, q) process is non-stationary. We can derive a state space representation for these models by augmenting the system matrices and state vector associated with the stationary representation. One way of doing this is to treat the ARIMA(p, d, q) as a structural model with ARMA(p, q) disturbances. For example, an ARIMA$(p, 1, q)$ process $\{y_t\}$ could be written as $y_t = \mu_t$, where $\mu_{t+1} = \mu_t + w_t$ and $\{w_t\}$ is an ARMA(p, q). The first block of the T matrix associated with the state space representation for this form is equal to the T matrix for the structural model. Interventions, such as level shifts, can be modeled by introducing shocks to this state space form. Tsay's (1988) level intervention where $\omega(B) = (I - B)^{-1}$ corresponds to a level intervention introduced using the method described above.

9.3 Periodic models

Consider the cycle plus noise model where $Z = (1 \quad 0)$,

$$T = \phi \begin{pmatrix} \cos\omega & \sin\omega \\ -\sin\omega & \cos\omega \end{pmatrix} \quad \text{and} \quad T^j = \phi^j \begin{pmatrix} \cos\omega j & \sin\omega j \\ -\sin\omega j & \cos\omega j \end{pmatrix}.$$

Thus, apart from measurement noise, y_t is is a cosine function with fixed frequency ω but slowly varying amplitude and phase. At time t the amplitude and phase are the polar coordinates of α_t. By equation (14), the response to a state shock at i is, for $t = i, \ldots, n$,

$$D_t(i) = \phi^{t-i}\{\delta_1 \cos \omega(t-i) + \delta_2 \sin \omega(t-i)\}$$

This is equivalent to adding a damped cycle with amplitude $(\delta_1^2 + \delta_2^2)^{1/2}$, phase $\arctan(\delta_2/\delta_1)$ and damping factor ϕ. If $\phi = 1$, the shock effect is persistent.

The polar coordinates of δ measure the aberant changes in amplitude and phase. If $\delta_1 = 0$ then the only possible change is a plus or minus π radian shift in the phase, which is obviously of limited practical relevance. If $\delta_2 = 0$, there is a change in amplitude but no change in the positioning of the peaks and troughs of the cycle. Thus, a special case of interest is where the shock impacts only the first component of the state.

10 Intervention in Models with Explanatory Variables

Many time series are modeled as being dependent on a set of observed exogenous variables $X_\beta = (\, X'_{1,\beta} \quad \cdots \quad X'_{n,\beta}\,)'$. For example, the monthly number of car drivers killed in road accidents in the UK may be influenced by the price of fuel and the car traffic index, that is, the total number of kilometres travelled by all cars each month (Harvey 1989). The null model now includes explanatory variables

$$y_t = X_{t,\beta}\beta + Z_t\alpha_t + G_t\varepsilon_t.$$

This indicates the direct impact of explanatory variables X_β on y_t. Indirect effects may also be present. These are modeled by incorporating explanatory variables into the state equation,

$$\alpha_{t+1} = W_{t,\beta}\beta + T_t\alpha_t + H_t\varepsilon_t.$$

The integrated regressor model is an example of a model in which observed exogenous variables have an indirect impact on the data y. A single null model run of the KFS now includes the additional recursions relating to the presence of the explanatory variables. These can be run in parallel with the ordinary filter by augmenting the vectors a_t and r_t. The recursions now include, in the filter

$$V_t = X_{t,\beta} - Z_tA_t, \quad A_{t+1} = -W_{t,\beta} + K_tV_t + T_tA_t, \quad t = 1, \ldots, n, \tag{26}$$

where $A_1 = 0$, and in the smoother

$$U_t = F_t^{-1} V_t - K_t' R_t, \quad R_{t-1} = Z_t' U_t + T_t' R_t, \quad t = n, \ldots, 1, \tag{27}$$

where $R_n = 0$. Notice that the recursions given by (26) and (27) are just the data dependent part of the ordinary KFS applied to the explanatory variables and U_t contains the smoothations of the columns of X_β. The combination of these additional recursions with the ordinary KFS is refered to as the augmented KFS.

Shocks are included using the method described in section 5. The parameter vectors δ and β are stacked as $\gamma = (\,\delta' \quad \beta'\,)'$. The regression matrix incorporating all external effects is $X_\gamma = (\,D(i) \quad X_\beta\,)$. This allows us to write the GLS estimate of γ when there is an intervention with origin i as

$$\begin{aligned} \hat{\gamma}_i &= (X_\gamma' \Sigma^{-1} X_\gamma)^{-1} (X_\gamma' \Sigma^{-1} y) \\ &= \begin{pmatrix} \sigma^{-2}\mathrm{cov}(D(i)'u) & D(i)'\Sigma^{-1} X_\beta \\ X_\beta' \Sigma^{-1} D(i) & \sigma^{-2}\mathrm{cov}(X_\beta' u) \end{pmatrix}^{-1} \begin{pmatrix} D(i)'u \\ X_\beta' u \end{pmatrix} \\ &= \begin{pmatrix} S_i & S_{i,\beta} \\ S_{i,\beta}' & S_\beta \end{pmatrix}^{-1} \begin{pmatrix} s_i \\ s_\beta \end{pmatrix}, \end{aligned} \tag{28}$$

where s_β, S_β and $S_{i,\beta}$ are defined by equation (28). Thus $\hat{\gamma}_i$ contains the estimate of both the shock and the regression parameters associated with the explanatory variables, adjusted for all other effects.

All quantities in (28) are evaluated in a single run of the explanatory variable augmented KFS. The contrast s_i and S_i are calculated using theorem 5.1. The quantities associated with the explanatory variables alone, that is $s_\beta = X_\beta' u$ and $S_\beta = \sigma^{-2}\mathrm{cov}(s_\beta)$ are not dependent on the origin of the intervention. In fact

$$s_\beta = X_\beta' \Sigma^{-1} y = (LX_\beta)' F^{-1} v = \sum_{t=1}^{n} V_t' F_t^{-1} v_t \quad \text{and} \quad S_\beta = \sum_{t=1}^{n} V_t' F_t^{-1} V_t,$$

are calculated in the forward augmented Kalman filter pass. The only remaining quantity is $S_{i,\beta} = D(i)' \Sigma^{-1} X_\beta$. Note that the components of $\Sigma^{-1} X_\beta$ are the smoothations of the columns of X_β, which coincide with the U_t. Using the form of $D(i)$ given by equation (14) yields

$$S_{i,\beta} = X_i' U_i + W_i' \sum_{t=i+1}^{n} T_{j-1,i+1}' Z_j' U_j = X_i' U_i + W_i' R_i,$$

where U_i and R_i are generated as in (27).

When there is no shock, that is $D(i) = 0$, equation (28) generates the null model GLS estimate of the parameter associated with the explanatory variables β. When interventions are included, (28) provides an estimate of β which takes into account the shock effects. Partitioned matrix inversion results applied to (28) yield

$$\hat{\delta}_i = (S_i - S_{i,\beta} S_\beta^{-1} S'_{i,\beta})^{-1}(s_i - S_{i,\beta}\hat{\beta}), \quad \hat{\beta}_i = \hat{\beta} - S_\beta^{-1} S'_{i,\beta}\hat{\delta}_i, \tag{29}$$

where $\hat{\beta} = S_\beta^{-1} s_\beta$ is the null model estimate of the parameter associated with the explanatory variables. Notice that, as before, GLS estimates of the parameters associated with all regression variables are generated with a single null model KFS run.

When $\Sigma = I$ the only meaningful signature is that corresponding to a measurement shock. In this case, equations (29) reduce to

$$\hat{\delta}_i = \{I - X_{i,\beta}(X'_\beta X_\beta)^{-1} X'_{i,\beta}\}^{-1}(y_i - X_{i,\beta}\hat{\beta}), \quad \hat{\beta}_i = \hat{\beta} - (X'_\beta X_\beta)^{-1} X_{i,\beta}\hat{\delta}_i,$$

where $X_{i,\beta}$ is the row block of X_β corresponding to y_i. These expressions coincide with those given for the standard regression case, see for example Atkinson (1985). Treating β as diffuse, the estimate of the shock is the scaled smoothed disturbance, $\hat{\delta}_i = \{I - \sigma^{-2}\text{cov}(\varepsilon_i|y)\}^{-1} E(\varepsilon_i|y)$. This is corollary 5.2 for standard regression models.

Measures of influence of a shock on the parameter estimate $\hat{\beta}$ can be defined from the expressions in (29). For example, a quantity analogous to Cook's distance (Cook 1977) is given by

$$(\hat{\beta} - \hat{\beta}_i)'\{\sigma^{-2}\text{cov}_o(\hat{\beta})\}^{-1}(\hat{\beta} - \hat{\beta}_i) = \hat{\delta}'_i S_{i,\beta} S_\beta^{-1} S'_{i,\beta}\hat{\delta}_i.$$

This statistic is typically scaled by division by $p\hat{\sigma}^2$ where p is the number of explanatory variables and $\hat{\sigma}^2 = n^{-1}(v'F^{-1}v - s'_\beta S_\beta^{-1} s_\beta)$. The estimate $\hat{\sigma}^2$ can be replaced by

$$\hat{\sigma}^2_i = \hat{\sigma}^2 - n^{-1}(s_i - S_{i,\beta}\hat{\beta})'(S_i - S_{i,\beta} S_\beta^{-1} S'_{i,\beta})^{-1}(s_i - S_{i,\beta}\hat{\beta}).$$

which takes into account the effect of the shock. Cook describes several extensions of his method for measuring influence. In the time series case, the appropriate generalisations of these statistics can be be computed using the framework which we have developed.

11 Composite Interventions

So far we have dealt with interventions which can be modeled by the addition of a single shock δ at a specific point of time i. The corresponding shock signatures represent many

of the structural changes observed in practice. However, the behaviour of the signature is constrained by the range of the $Z_t T_{t-1,i+1}$ matrices for $t > i$. The null model is often time invariant so the type of interventions which can be represented using a single state shock are determined by the sequence $\{ZT^k : k = 0, \ldots, n-i-1\}$. Thus, a given null model may not be able to induce the desired shock signature in which case the simple intervention specification is too limiting.

One way around this problem is to build in null model redundancy, that is, write the fitted model in an overelaborate form. This can be used to introduce structure into the matrix T so that the range of ZT^k includes the desired shock signature. To illustrate this, suppose that the null model is a random walk plus measurement noise so that $Z = T = 1$. The shock signatures corresponding to measurement and state shocks are $(\,0 \quad \cdots \quad 0 \quad 1 \quad 0 \quad \cdots \quad 0\,)$ and $(\,0 \quad \cdots \quad 0 \quad 1 \quad \cdots \quad 1\,)$ respectively. Using the single shock methodology we can only test for interventions which are linear combinations of these two patterns. However, the random walk can be written in a form containing common factors, $\alpha_{t+1} = 2\alpha_t - \alpha_{t-1} + \eta_t - \eta_{t-1}$ where η_t is the random walk noise. Setting up the state space representation of this overelaborate version of the null model yields

$$T = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad ZT^k = (\,2k+1 \quad -2k\,)\,.$$

A shock signature $(\,0 \quad \cdots \quad 0 \quad 1 \quad 2 \quad \cdots \quad n-i\,)$ is now also in the range of ZT^k. Hence, the overelaborate representation of the null model allows for testing of slope changes. This approach has obvious uses but is not explored further in this paper.

An alternative approach is to stack the observations yielding a vector time series where each segment of time permits entry of a shock. The corresponding stacked version of the state space model has increased state dimension and more intricate system matrices. The stacked model also bars estimation of shocks whose entry points cross segment boundaries. It is worth developing an approach which does not suffer from these limitations.

The method we adopt is to generalize the representation given by (12) and (13), allowing non-zero X_t and W_t for $t = i, \ldots, i+q$. If i is the origin of the intervention then

$$\begin{aligned} y_t &= X_t(i)\delta + Z_t\alpha_t + G_t\varepsilon_t, \\ \alpha_{t+1} &= W_t(i)\delta + T_t\alpha_t + H_t\varepsilon_t, \end{aligned} \tag{30}$$

where $X_t(i) = 0$ and $W_t(i) = 0$ for $t < i$ and $t > i+q$. The shock signature is then

$$D_t(i) = \begin{cases} 0, & t = 1, \ldots, i-1, \\ X_t(i) + Z_t \sum_{j=i}^{t-1} T_{t-1,j+1} W_j(i), & t = i, \ldots, i+q, \\ Z_t \sum_{j=i}^{i+q} T_{t-1,j+1} W_j(i), & t = i+q+1, \ldots, n. \end{cases}$$

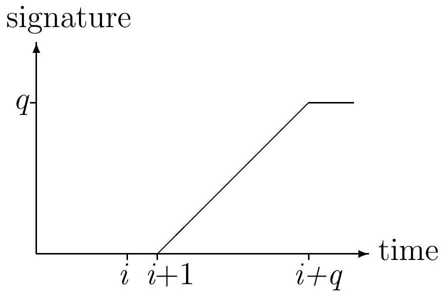

Figure 3: The signature for a temporary slope change

The case $q = 0$ is the simple intervention case.

We illustrate the use of composite interventions by considering the local linear trend model given by equations (24). Temporary changes in the structure of this type of series can be represented using just two shocks; one to initiate the change and the other q periods later to turn it off. For example, we can model a slope change which only affects q periods by applying a shock to the appropriate component of the state vector at $t = i$ and then an equivalent reverse shock at $t = i + q$. Taking $X_t(i) = 0$ and $W_t(i) = 0$ except for $W_i(i) = (\,0 \quad 1\,)'$ and $W_{i+q}(i) = (\,0 \quad -1\,)'$, we can write down the intervention signature,

$$D_t(i) \;=\; \begin{cases} 0, & t = 1, \ldots, i, \\ t - i - 1, & t = i + 1, \ldots, i + q, \\ q, & t = i + q + 1, \ldots, n. \end{cases}$$

The intervention signature indicates a linear increase in the level for periods $t = i$ through to $t = i + q$. After this the series reverts to its normal dynamics but starting from a level $q\delta$ higher than it would have been in the absence of an intervention.

The estimation and testing of shock effects under the more general model is based on the following theorem.

Theorem 11.1 *For a composite intervention the intervention contrast s_i is given by*

$$s_i = \sum_{j=i}^{i+q} \{X_j(i)' u_j + W_j(i)' r_j\},$$

where u_j and r_j are generated by the KFS applied to the null model. Further, $S_i = \sigma^{-2}\mathrm{cov}(s_i)$ can be generated using the following recursion. Set $S_i \leftarrow 0$ and $C \leftarrow 0$ and for $j = i+q, \ldots, i$,

compute recursively

$$
\begin{array}{rcl}
S_i & \leftarrow & S_i + X_j(i)' F_j^{-1} X_j(i) + Q_j(i)' N_j Q_j(i) + Q_j(i)' C + C' Q_j(i), \\
C & \leftarrow & Z_j' F_j^{-1} X_j(i) + L_j' \{ N_j Q_j(i) + C \},
\end{array}
$$

where $Q_j(i) = W_j(i) - K_j X_j(i)$.

To illustrate the method, consider a state shock at $t = i$ with an equivalent reverse shock at $t = i + q$, that is, $W_i(i) = I$ and $W_{i+q}(i) = -I$. Applying theorem 11.1 yields

$$s_i = s_{i,i} + s_{i,i+q} = r_i - r_{i+q},$$

and thus

$$S_i = \sigma^{-2} \mathrm{cov}(r_i - r_{i+q}) = N_i + N_{i+q} - L_{i+q,i+1}' N_{i+q} - N_{i+q} L_{i+q,i+1}.$$

Theorem 11.1 also covers the case where different components of δ enter the system at different times $t = i, \ldots, i + q$. For example, δ may have four components each entering once at successive times.

12 Conclusion

The introduction of shocks is a powerful tool for analysing departures from a fitted null model. Many forms of aberant behaviour can be modeled efficiently by shocks to the transition equation of a state space representation. We establish that alternative model statistics take simple forms when viewed as functions of the smoothations. Test statistics, for any number of interventions, can be generated using the output of a single null model KFS run. The theoretical methods in this paper can be extended to several areas some of which are outlined briefly below.

Atkinson, Koopman and Shephard (1995) use score statistics to detect structural changes. They use the augmented filter to generate adjustments to the smoother output. Derivatives of the alternative model likelihood function are generated by applying the method described by Koopman and Shephard (1992) using the adjusted smoothed disturbances and their covariance matrices. The approach adopted by Atkinson et al. requires running the augmented KFS for each intervention type and every origin. This can be avoided by applying the results in section 8 to adjust the smoother output for the presence of a shock.

Many authors (Harvey & Phillips 1979, Ansley & Kohn 1985, De Jong 1988b, De Jong 1991) describe the problem of filtering a series with diffuse intial conditions. They put forward a number of related algorithms which involve elaborate adjustments to the filtering process. Diffuse initial conditions can be modeled by the introduction of a shock at $t = 0$. Running the KFS, using an arbitrary initialisation, will provide an updated estimate of the starting conditions. This step can be incorporated into the optimization algorithm. Thus, on convergence of the maximisation process, the estimate of the initial conditions will be computed along with the maximum likelihood estimates of the hyperparameters.

Appendix

Proof of theorem 4.1

In order to prove this theorem we need the following lemma.

Lemma *Suppose* $\mathrm{cov}(y) = \sigma^2\Sigma$. *Let* Jy *be any selection of the components of* y *with* Ky *denoting the remaining components. The jackknife residual and conditional covariance are given by*

$$\begin{aligned} \{\sigma^{-2}\mathrm{cov}(Jy|Ky)\}^{-1}\{Jy - E(Jy|Ky)\} &= J\Sigma^{-1}\{y - E(y)\}, & (31)\\ \{\sigma^{-2}\mathrm{cov}(Jy|Ky)\}^{-1} &= J\Sigma^{-1}J'. & (32) \end{aligned}$$

Proof: Put $y_1 = Ky$ and $y_2 = Jy$ and, without loss of generality, assume that $y = (\, y_1' \quad y_2' \,)'$. Give Σ a partition conformal to the partition of y. By the inverse of a partitioned matrix

$$\Sigma^{-1} = \begin{pmatrix} \Sigma_1 & \Sigma_{12} \\ \Sigma_{21} & \Sigma_2 \end{pmatrix}^{-1} = \begin{pmatrix} \Sigma_1^{-1} + \Sigma_1^{-1}\Sigma_{12}\Sigma_{2|1}^{-1}\Sigma_{21}\Sigma_1^{-1} & -\Sigma_1^{-1}\Sigma_{12}\Sigma_{2|1}^{-1} \\ -\Sigma_{2|1}^{-1}\Sigma_{21}\Sigma_1^{-1} & \Sigma_{2|1}^{-1} \end{pmatrix},$$

where

$$\Sigma_{2|1} = \Sigma_2 - \Sigma_{21}\Sigma_1^{-1}\Sigma_{12} = \sigma^{-2}\mathrm{cov}(y_2|y_1),$$

which establishes (32). By definition

$$E(y_2|y_1) = E(y_2) + \Sigma_{21}\Sigma_1^{-1}\{y_1 - E(y_1)\},$$

and hence the lower block of $\Sigma_1^{-1}\{y_1 - E(y_1)\}$ is given by

$$\Sigma_{2|1}^{-1}[\{y_2 - E(y_2)\} - \Sigma_{21}\Sigma_1^{-1}\{y_1 - E(y_1)\}] = \{\sigma^{-2}\mathrm{cov}(y_2|y_1)\}^{-1}\{y_2 - E(y_2|y_1)\},$$

which establishes (31). $\square$

Returning to the proof of theorem 4.1, let v be the stack of innovations, u be the stack of smoothations and $\text{cov}(y) = \sigma^2\Sigma$. The innovations are given by $v = Ly$, where L is a lower triangular matrix with identity matrices on the diagonal. Thus

$$\text{cov}(v) = \text{cov}(Ly) = \sigma^2 L\Sigma L' = \sigma^2\text{diag}(F_1, \ldots, F_n) = \sigma^2 F.$$

The Kalman filter equations (9) for v_t and a_{t+1} can be written as

$$\begin{pmatrix} v_t \\ a_{t+1} \end{pmatrix} = \begin{pmatrix} I & -Z_t \\ K_t & L_t \end{pmatrix} \begin{pmatrix} y_t \\ a_t \end{pmatrix}.$$

Similarly, combining the equations (11) for u_t and r_{t-1} yields

$$\begin{pmatrix} -u_t \\ r_{t-1} \end{pmatrix} = \begin{pmatrix} I & -Z_t \\ K_t & L_t \end{pmatrix}' \begin{pmatrix} -F_t^{-1}v_t \\ r_t \end{pmatrix},$$

which works through the data, in reverse order, with $r_n = 0$ and $-F_t^{-1}v_t$ as input. If $v = Ly$ then

$$u = L'F^{-1}v = \Sigma^{-1}y, \quad \text{cov}(u) = \sigma^2 L'F^{-1}L = \sigma^2\Sigma^{-1}. \tag{33}$$

By definition, $\text{cov}(u_t) = \sigma^2 M_t$ so, from (33), it follows that M_t is the t^{th} diagonal block of Σ^{-1}. The final part of the theorem follows from the lemma given above.

Proof of theorem 5.1

We establish the first part of this theorem in section 5, $s_i = X_i'u_i + W_i'r_i$. From this it follows that

$$s_i = X_i'(F_i^{-1}v_i - K_i'r_i) + W_i'r_i = X_i'F_i^{-1}v_i + (W_i - K_iX_i)'r_i = X_i'F_i^{-1}v_i + Q_i'r_i.$$

Now r_i is linear in the future innovations and v_i and r_i are uncorrelated. The covariance matrices of v_i and r_i are respectively F_i and N_i. Thus, $S_i = \sigma^{-2}\text{cov}(s_i)$ is as asserted.

Proof of corollary 5.2

If $X_i = G_i$ and $W_i = H_i$ then

$$s_i = G_iu_i + H_iR_i = G_iF_i^{-1}v_i + J_i'r_i.$$

The theorem then follows directly from De Jong (1996).

Proof of theorem 6.1

Put $A = (\, X_i' \quad W_i' \,)$ and $x = (\, u_i' \quad r_i' \,)'$ then A is of fixed column dimension and

$$\rho_i^2 = (Ax)'\{\sigma^{-2}\text{cov}(Ax)\}^{-1}(Ax).$$

Define A^- as a generalised inverse of A. If A has full column rank then $A^-A = I$ and

$$\begin{aligned}
\rho_i^2 &= x'A'\{A\sigma^{-2}\text{cov}(x)A'\}^{-1}Ax \\
&= x'(A^-A)'\{\sigma^{-2}\text{cov}(x)\}^{-1}(A^-A)x \\
&= x'\{\sigma^{-2}\text{cov}(x)\}x \\
&= (\, u_i' \quad r_i' \,)\,[\sigma^{-2}\text{cov}\{(\, u_i' \quad r_i' \,)'\}]^{-1}\,(\, u_i' \quad r_i' \,)',
\end{aligned}$$

where

$$\text{cov}\{(\, u_i' \quad r_i' \,)'\} = \begin{pmatrix} M_i & -K_i'N_i \\ -N_iK_i & N_i \end{pmatrix}.$$

Applying standard partitioned matrix inversion results yields

$$\rho_i^2 = v_i'F_i^{-1}v_i + r_i'N_i^{-1}r_i,$$

Thus, the maximum is as stated in (17) when A has full column rank. If A has rank 1 then

$$\rho_i^2 = (Ax)^2/\{\sigma^{-2}A\text{cov}(x)A'\}$$

which attains the maximum given by (17) at

$$A = x'\{\sigma^{-2}\text{cov}(x)\}^{-1} = (\, u_i' \quad r_i' \,)\,[\sigma^{-2}\text{cov}\{(\, u_i' \quad r_i' \,)'\}]^{-1} = (\, v_i' \quad v_i'K_i' + r_i'N_i^{-1} \,)',$$

(see for example Rao (1973)). For A with rank between 1 and full column rank the maximum must be equal to (17) since it must lie between the rank 1 and full column rank maxima.

Proof of theorem 6.2

We can exploit the fact that the v_ts are uncorrelated and $r_t = \sum_{k=t+1}^{n} L_{k-1,t+1}'Z_k'F_k^{-1}v_k$ to derive a simple expression for $\text{cov}(s_i, s_j)$. For $i < j$,

$$\begin{aligned}
\text{cov}(s_i, s_j) &= \text{cov}(X_i'F_i^{-1}v_i + Q_i'r_i, X_j'F_j^{-1}v_j + Q_j'r_j) \\
&= Q_i'\text{cov}(r_i, v_j)F_j^{-1}X_j + Q_i'\text{cov}(r_i, r_j)Q_j \\
&= Q_i'L_{j-1,i+1}'Z_j'F_j^{-1}\text{cov}(v_j, v_j)F_j^{-1}X_j + Q_i'L_{j-1,i+1}'L_j'\text{cov}(r_j, r_j)Q_j \\
&= \sigma^2 Q_i'L_{j-1,i+1}'(Z_j'F_j^{-1}X_j + L_j'N_jQ_j),
\end{aligned}$$

Noting that $\text{cov}(\hat{\delta}_i, \hat{\delta}_j) = S_i^{-1}\text{cov}(s_i, s_j)S_j^{-1}$, yields the required result.

Proof of theorem 8.1

First note

$$E_a(\gamma|y,\delta) = E_a\{\gamma|\delta, y - E_a(y|\delta)\} = E_a(\gamma|\delta) + C\{y - E_a(y|\delta)\}, \tag{34}$$

where, since $E_a(y|\delta) = D(i)\delta$,

$$C = \text{cov}_a\{\gamma, y - D(i)\delta\}[\text{cov}_a\{y - D(i)\delta\}]^{-1} = \text{cov}_o(\gamma, y)\{\text{cov}_o(y)\}^{-1} = \sigma^{-2}\text{cov}_o(\gamma, \Sigma^{-1}y).$$

From the form of the alternative hypotheses $E_a(\gamma|\delta) = E_o(\gamma) + \Gamma\delta$ and $E_a(y|\delta) = E_o(y) + D(i)\delta$. Substituting these into equation (34) yields

$$\begin{aligned} E_a(\gamma|y,\delta) &= E_o(\gamma) + \Gamma\delta + C\{y - E_o(y) - D(i)\delta\} \\ &= [E_o(\gamma) + C\{y - E_o(y)\}] + \{\Gamma - CD(i)\delta\} \\ &= E_o(\gamma|y) + (\Gamma - \Lambda_i)\delta, \end{aligned}$$

since

$$CD(i) = \sigma^{-2}\text{cov}_o(\gamma, \Sigma^{-1}y)D(i) = \sigma^{-2}\text{cov}_o(\gamma, s_i) = \Lambda_i.$$

We know that $E_a(\gamma|y) = E_a\{E_a(\gamma|y,\delta)|y\}$, so

$$E_a(\gamma|y) = E_a\{E_o(\gamma|y) + (\Gamma - \Lambda_i)\delta|y\} = E_o(\gamma|y) + (\Gamma - \Lambda_i)\hat{\delta}_i,$$

and

$$\text{cov}_a(\gamma|y) = \text{cov}_a(\gamma|y,\delta) + \text{cov}_a\{E_a(\gamma|y,\delta)|y\} = \text{cov}_o(\gamma|y) + (\Gamma - \Lambda_i)\text{cov}_o(\delta|y)(\Gamma - \Lambda_i)',$$

as required.

Proof of proposition 8.1

Basic identities used through out the proof are

$$\begin{aligned} s_i = X_i'F_i^{-1}v_i + Q_i'r_i, &\qquad Z_iP_iL_i' = -G_iJ_i', \\ v_i = Z_i(\alpha_i - a_i) + G_i\varepsilon_i, &\qquad \alpha_{i+1} - a_{i+1} = L_i(\alpha_i - a_i) + J_i\varepsilon_i. \end{aligned}$$

These can be established by direct manipulation of filter and smoother quantities. Thus, in general

$$\text{cov}(\gamma, s_i) = \text{cov}(\gamma, v_i)F_i^{-1}X_i + \text{cov}(\gamma, r_i)Q_i,$$

and in particular

$$\begin{aligned}\text{cov}(\alpha_i, v_i) &= \text{cov}(\alpha_i, \alpha_i - a_i)Z_i' = \sigma^2 P_i Z_i', \\ \text{cov}(\varepsilon_i, v_i) &= \sigma^2 G_i'.\end{aligned}$$

Further

$$\begin{aligned}\text{cov}(\alpha_i, r_i) &= \text{cov}\{E(\alpha_i|y), r_i\} = P_i \text{cov}(r_{i-1}, r_i) = \sigma^2 P_i L_i' N_i, \\ \text{cov}(\varepsilon_i, r_i) &= \text{cov}(E(\varepsilon_i|y), r_i) = J_i' \text{cov}(r_i) = \sigma^2 J_i' N_i.\end{aligned}$$

The expressions for $\text{cov}(\alpha_i, s_i)$ and $\text{cov}(\varepsilon_i, s_i)$ follow directly from the above. Next

$$\text{cov}(Z_i\alpha_i, s_i) = \sigma^2 Z_i P_i R_i = \sigma^2\{(F_i - G_i G_i')F_i^{-1}X_i - G_i J_i' N_i Q_i\},$$

which simplifies to the expression given in the statement of the proposition. When $t < i$,

$$\begin{aligned}\text{cov}(\alpha_t, v_i) &= \text{cov}(\alpha_t, \alpha_i - a_i)Z_i' = \text{cov}(\alpha_t, \alpha_{i-1} - a_{i-1})L_{i-1}Z_i' = \sigma^2 P_t L_t'(L_{i-1}\dots L_{t+1})'Z_i', \\ \text{cov}(\varepsilon_t, v_i) &= \text{cov}(\varepsilon_t, \alpha_i - a_i)Z_i' = \sigma^2 J_t'(L_{i-1}\dots L_{t+1})'Z_i',\end{aligned}$$

and

$$\begin{aligned}\text{cov}(\alpha_t, r_i) &= \text{cov}(\alpha_t, v_{i+1})F_{i+1}^{-1}Z_{i+1} + \text{cov}(\alpha_t, r_{i+1})L_{i+1} = \sigma^2 P_t L_t'(L_{i-1}\dots L_{t+1})'Z_i', \\ \text{cov}(\varepsilon_t, r_i) &= \text{cov}(\varepsilon_t, v_{i+1})F_{i+1}^{-1}Z_{i+1} + \text{cov}(\varepsilon_t, r_{i+1})L_{i+1} = \sigma^2 J_t'(L_{i-1}\dots L_{t+1})'L_i'N_i,\end{aligned}$$

implying that $\text{cov}(\alpha_t, s_i)$ and $\text{cov}(\varepsilon_t, s_i)$ are as asserted for $t < i$. The expression for $\text{cov}(Z_t\alpha_t, s_i)$ follows from $Z_t\text{cov}(\alpha_t, s_i)$ and the identity $Z_t P_t L_t' = -G_t J_t'$. Finally consider the case $t > i$. We have

$$\text{cov}(\alpha_{t+1}, s_i) = T_t \text{cov}(\alpha_t, s_i) + H_t \text{cov}(\varepsilon_t, s_i),$$

which is a recursive formula with $\text{cov}(\varepsilon_i, s_i)$ as above. In particular

$$\text{cov}(\alpha_{i+1}, s_i) = \sigma^2(T_i P_i R_i + H_i C_i) = \sigma^2(K_i X_i + P_{i+1} N_i Q_i).$$

Futher for $t > i$

$$\begin{aligned}\text{cov}(\varepsilon_t, s_i) &= \text{cov}(\varepsilon_t, r_i)Q_i = \text{cov}(\varepsilon_t, r_{i+1})L_{i+1}Q_i = \text{cov}(\varepsilon_t, r_t)(L_{t-1}\dots L_{i+1})Q_i \\ &= \{\text{cov}(\varepsilon_t, v_t)F_t^{-1}Z_t + \text{cov}(\varepsilon_t, r_t)L_t\}(L_{t-1}\dots L_{i+1})Q_i \\ &= \sigma^2(G_t' F_t^{-1} Z_t + J_t' N_t L_t)(L_{t-1}\dots L_{i+1})Q_i\end{aligned}$$

as stated in the proposition.

Proof of theorem 11.1

The expression for s_i is proved by noting that the intervention signature is given by $D(i) = \sum_{j=i}^{i+q} D(i,j)$, where $D(i,j)$ is the shock signature arising from the shock entry at $t = j$, and $s_i = \sum_{j=i}^{i+q} s_{i,j}$ where

$$s_{i,j} = D(i,j)'u = X_j(i)'u_j + W_j(i)'r_j = X_j(i)'F_j^{-1}v_j + Q_j(i)'r_j.$$

The proof of the recursion to compute S_i follows by inspection of the covariances between the $s_{i,j}$.

References

Ansley, C. F. & Kohn, R. (1985). Estimation, filtering and smoothing in state space models with incompletely specified initial conditions, *Annals of Statistics* **13**: 1286–1316.

Ansley, C. F. & Kohn, R. (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*, Clarendon Press, Oxford.

Atkinson, A. C., Koopman, S. J. & Shephard, N. (1997). Detecting shocks: outliers and breaks in time series, *Journal of Econometrics* **80**: 387–422.

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

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

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

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

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

Cook, R. D. (1977). Detection of influential observation in linear regression, *Technometrics* **19**: 15–18.

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

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

De Jong, P. (1988b). The likelihood for a state space model, *Biometrika* **75**: 165–169.

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. (1997). Fixed interval smoothing. Submitted to International Statistical Review.

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

Harrison, J. & West, M. (1991). Dynamic linear model diagnostics, *Biometrika* **78**: 797–808.

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

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

Harvey, A. C. & Phillips, G. D. A. (1979). Maximum likelihood estimation of regression models with autoregressive-moving average disturbances, *Biometrika* **66**: 49–58.

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

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

Koopman, S. J. & Shephard, N. (1992). Exact score for time series models in state space form, *Biometrika* **79**: 823–826.

LeFrançois, B. (1991). Detecting over-influential observations in time series, *Biometrika* **78**: 91–99.

McCulloch, R. E. & Tsay, R. S. (1993). Bayesian inference and prediction for mean and variance shifts in autoregressive time series, *Journal of the American Statistical Association* **88**: 968–978.

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

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

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.

Whittle, P. (1984). *Prediction and Regulation by Linear Least Square Methods*, 2nd edn, Blackwell, Oxford.