13

Conventional State Space Models

The primary purpose of this book is to demonstrate that the innovations form of the state space model provides a simple but flexible approach to forecasting time series. However, for reasons that are not completely clear, the innovations form has been largely over-shadowed in the literature by another version of the state space model that has multiple sources of randomness. We refer to this version as the multi-disturbance or multiple source of error (MSOE) model. The two approaches are compared and contrasted in this chapter. When we are comparing the two frameworks directly, both the finite and infinite start-up assumptions are valid; however, when the two are compared via their ARIMA reduced forms, the infinite start-up assumption will be used. The emphasis will be almost exclusively upon linear state space models, because, as we shall see in Sect. 13.4, the MSOE formulation becomes difficult to manage in the nonlinear case.

In Chap. 2, we introduced the local level and local trend models, together with their seasonal extensions. It will be seen that these innovations, or single source of error (SSOE), models all have their counterparts within a multiple source of error framework. It is often thought that the MSOE provides a better modeling framework than the SSOE because the multiple sources of error appear to allow greater generality. We will show that any MSOE model has an innovations representation, so that this viewpoint cannot be correct.

A general definition of the state space framework is presented in Sect. 13.1. It is seen to encompass both the innovations and the multiple disturbance forms of the state space model. Several important special cases of the MSOE are also given. A general approach to estimation is given in Sect. 13.2. Reduced forms of the MSOE models are examined in Sect. 13.3. The SSOE and MSOE approaches are then compared in Sect. 13.4.

13.1 State Space Models

The overall state of a system in period t is represented by a random vector $\boldsymbol{z}_t$, which incorporates both the observations and the unobservable states. The elements in $\boldsymbol{z}_t$ are arranged so that $\boldsymbol{z}_t = (y_t, \boldsymbol{x}_t)'$, where y_t denotes the observation at time t, which will be recorded over the periods 1 to n, and $\boldsymbol{x}_t$ is the random vector of k unobservable states.

The evolution of the state of the system is governed by the first-order recurrence relationship

$$\boldsymbol{z}_t = \boldsymbol{A}\boldsymbol{z}_{t-1} + \boldsymbol{u}_t, \tag{13.1a}$$

$$\text{where} \quad \boldsymbol{u}_t \sim \text{NID}(\boldsymbol{0}, \boldsymbol{V}_u), \tag{13.1b}$$

$\boldsymbol{A}$ is a fixed matrix and $\boldsymbol{V}_u$ is a positive semi-definite variance matrix. This general format is particularly useful when we consider parameter estimation in Sect. 13.2. When expressed in terms of the observable and unobservable states, (13.1) may be written as

$$y_t = \boldsymbol{w}'\boldsymbol{x}_{t-1} + \varepsilon_t, \tag{13.2a}$$

$$\boldsymbol{x}_t = \boldsymbol{F}\boldsymbol{x}_{t-1} + \boldsymbol{\eta}_t, \tag{13.2b}$$

$$\begin{bmatrix} \varepsilon_t \\ \boldsymbol{\eta}_t \end{bmatrix} \sim \text{NID}\left(\begin{bmatrix} 0 \\ \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} V_\varepsilon & \boldsymbol{V}_{\varepsilon\eta} \\ \boldsymbol{V}_{\eta\varepsilon} & \boldsymbol{V}_\eta \end{bmatrix} \right), \tag{13.2c}$$

where $\boldsymbol{w}$ is a fixed vector and $\boldsymbol{F}$ is a fixed matrix. That is, $\boldsymbol{z}_t = \begin{bmatrix} y_t \\ \boldsymbol{x}_t \end{bmatrix}$, $\boldsymbol{A} = \begin{bmatrix} 0 & \boldsymbol{w}' \\ \boldsymbol{0} & \boldsymbol{F} \end{bmatrix}$, $\boldsymbol{u}_t = \begin{bmatrix} \varepsilon_t \\ \boldsymbol{\eta}_t \end{bmatrix}$, and $\boldsymbol{V}_{\boldsymbol{u}} = \begin{bmatrix} V_\varepsilon & \boldsymbol{V}_{\varepsilon\eta} \\ \boldsymbol{V}_{\eta\varepsilon} & \boldsymbol{V}_\eta \end{bmatrix}$.

As in earlier chapters, (13.2a) and (13.2b) are called the measurement and transition equations respectively. Further, we again assume that the observation y_t depends only on the unobserved states $\boldsymbol{x}_{t-1}$ as they prevailed at the beginning of period t (at time $t-1$).

When $\boldsymbol{\eta}_t = \boldsymbol{g}\varepsilon_t$ (where $\boldsymbol{g}$ is a fixed vector of persistence parameters), the state space model becomes

$$y_t = \boldsymbol{w}'\boldsymbol{x}_{t-1} + \varepsilon_t, \tag{13.3a}$$

$$\boldsymbol{x}_t = \boldsymbol{F}\boldsymbol{x}_{t-1} + \boldsymbol{g}\varepsilon_t, \tag{13.3b}$$

$$\varepsilon_t \sim \text{NID}(0, V_\varepsilon). \tag{13.3c}$$

Equations (13.3) describe the vector form of the innovations model, which was introduced in Sect. 2.5.2. Another form of state space model assumes that $\boldsymbol{V}_{\varepsilon\eta} = \boldsymbol{0}$ and that $\boldsymbol{V}_\eta$ is diagonal. We refer to this model as the *multi-disturbance* or *MSOE* state space model. Both possibilities involve restrictions, but the second form places independence assumptions on the disturbances. When there are k states, this formulation includes $k+1$ unknown variances as parameters, just as the innovations model includes $k+1$ parameters: a

single variance and k persistence parameters. These choices represent the maximum number of parameters that can be built into the models that retain the *estimability* (or *identifiability*) of all parameters.

At first sight the MSOE model appears to be more general than the innovations form because it involves more random disturbances. However, as we will show in Sect. 13.4, any MSOE model may be represented in innovations form so that there is only a need for one primary source of randomness for each observable state. This conclusion, it should be noted, is derived under the assumption that the disturbances have a Gaussian distribution; it may not be true for non-Gaussian state space models. Nevertheless, because most applications rely upon the mean and variance structures of the models, the practical implication is that little, if anything, will be lost by using the SSOE approach. Furthermore, as we will see later in this chapter, the innovations model approach provides several benefits.

In earlier chapters, we have examined the use of the innovations form of the state space framework to model evolving common features such as trends and seasonal patterns. Particularly important cases included the local level, local trend and damped trend, and their seasonal extensions. It will now be shown that each case has a multi-disturbance analogue.

The multi-disturbance versions form what has been called a *structural approach* to time series (Harvey 1989), one that has been widely used in economic studies. The following table shows corresponding standard structural models from the two approaches. We note that although the common symbols ℓ, b and ε are used to represent the level, slope and innovation respectively, their values and meaning differ between the two frameworks. The multiple disturbance versions presented here differ slightly from those of Harvey (1989); a point we explore in the next subsection.

Model	Conventional models	Innovations models
Level	$y_t = \ell_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \eta_t$	$y_t = \ell_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$
Trend	$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \eta_t$ $b_t = b_{t-1} + \xi_t$	$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$
Seasonal	$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \eta_t$ $b_t = b_{t-1} + \xi_t$ $s_t = s_{t-m} + \omega_t$	$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$

13.1.1 Canonical Forms

The MSOE scheme assumes that the various error processes are independent. Thus, in (13.2) we would set $\boldsymbol{V}_{\varepsilon\eta} = \mathbf{0}$. However, most MSOE

formulations (e.g., Harvey 1989; West and Harrison 1997) specify the measurement equation as

$$y_t = \boldsymbol{w}'\boldsymbol{x}_t + \varepsilon_t^*. \tag{13.4}$$

If we substitute the transition equation (13.2b) into this expression we arrive at

$$y_t = \boldsymbol{w}'\boldsymbol{F}\boldsymbol{x}_{t-1} + \boldsymbol{w}'\eta_t + \varepsilon_t^*,$$

so that the errors in the measurement and transition equations are now correlated. In order to make the independence assumption operational, we must choose a specific model, termed the *canonical model* by West and Harrison (1997, Chap. 5). Further, we must recognize that any transformation of the state variables may result in previously uncorrelated errors becoming correlated. The innovations approach provides a simple way out of this dilemma. Because the errors are perfectly correlated, any linear transformation leaves them perfectly correlated. The details are provided in Exercise 13.2. It may be shown that the different forms of the model have no effect on predictions, but the choices mean that individual components such as the local level may have different values; see Exercise 13.3.

13.1.2 Other State Space Models

A number of other formulations have appeared in the literature over the years. Akaike (1974) proposed an innovations model that maps directly into an ARMA$(k, k-1)$ scheme. The details are given in Exercise 13.1. Aoki (1987) also presents an innovations form, but we do not pursue these alternatives further in this book.

13.2 Estimation

The unknown parameters of both innovations and multi-disturbance state space models must be estimated. Because they both conform to the structure described in (13.1), a theory of estimation encompassing both cases is developed in terms of the more general framework. The seed state vector $\boldsymbol{z}_0$ is assumed to be random rather than fixed. Two points need to be made at this stage. The first is that virtually all the current literature on the multi-disturbance model relies upon the Kalman filter to develop the estimates. The second is that this reliance is not necessary as the information filter that is described in Sect. 12.3 is applicable in both frameworks. We proceed by first providing a general framework and then adapting the Kalman filter from Sect. 12.7 to its familiar MSOE form.

Using arguments similar to those in Chap. 12, it may be argued that the likelihood function can be rewritten in a prediction error form. For the moment, the focus is restricted to the case where all the states, both observable and unobservable, are stationary. The observations are represented by

the n-vector $\boldsymbol{y}$. The unknown parameters are collected together into a vector $\boldsymbol{\theta}$. The prediction error decomposition of the likelihood function is (see Schweppe 1965)

$$\mathcal{L}(\boldsymbol{\theta} \mid \boldsymbol{y}) = \prod_{t=1}^{n} (v_{t|t-1})^{-1/2} \exp\left(-\tfrac{1}{2}(y_t - \mu_{t|t-1})^2 \big/ v_{t|t-1}\right),$$

where $\mu_{t|t-1}$ and $v_{t|t-1}$ are the mean and variance of the one-step-ahead prediction distribution. We use $\boldsymbol{\theta}$ to denote the variances of the different error terms in the general model (13.1). The aim is to maximize the likelihood with respect to $\boldsymbol{\theta}$. To implement the maximum likelihood approach, we need a mechanism to generate the updated values $\mu_{t|t-1}$ and $v_{t|t-1}$.

13.2.1 Kalman Filter

The Kalman filter was considered in Sect. 12.7 for the innovations state space model. The version presented here is based on the more general model (13.1) and so encompasses both the innovations model and the MSOE model as special cases. The argument follows along the same lines as that in Sect. 12.7. We let $\boldsymbol{y}_{1:s} = y_1, \ldots, y_s$ and define

$$\begin{aligned}
\mu_{t|s} &= \mathrm{E}(y_t \mid y_{1:s}), \\
v_{t|s} &= \mathrm{V}(y_t \mid y_{1:s}), \\
\boldsymbol{m}_{t|s} &= \mathrm{E}(\boldsymbol{x} \mid y_{1:s}), \\
\boldsymbol{V}_{t|s} &= \mathrm{V}(\boldsymbol{x} \mid y_{1:s}), \\
\boldsymbol{\zeta}_{t|t-1} &= \mathrm{Cov}(\boldsymbol{x}_t, y_t \mid \boldsymbol{y}_{1:t-1}), \\
\boldsymbol{m}^z_{t|s} &= \mathrm{E}(\boldsymbol{z} \mid y_{1:s}), \\
\text{and} \qquad \boldsymbol{V}^z_{t|s} &= \mathrm{V}(\boldsymbol{z} \mid y_{1:s}) = \begin{bmatrix} v_{t|s} & \boldsymbol{\zeta}'_{t|s} \\ \boldsymbol{\zeta}_{t|s} & \boldsymbol{V}_{t|s} \end{bmatrix}.
\end{aligned}$$

Then, using the notation of (13.1), the Kalman filter is given in part by the equations

$$\boldsymbol{z}_{t|t-1} = \boldsymbol{A}\boldsymbol{z}_{t-1|t-1} + \boldsymbol{u}_t, \tag{13.5a}$$

$$\boldsymbol{m}^z_{t|t-1} = \boldsymbol{A}\boldsymbol{m}^z_{t-1|t-1}, \tag{13.5b}$$

$$\boldsymbol{V}^z_{t|t-1} = \boldsymbol{A}\boldsymbol{V}^z_{t-1|t-1}\boldsymbol{A}' + \boldsymbol{V}_u. \tag{13.5c}$$

It is assumed that the distribution of $\boldsymbol{z}_{t-1|t-1}$ is available from the preceding iteration of the filter after processing $t-1$ observations. The exception is period $t = 1$ where $\boldsymbol{z}_{0|0}$ is described by the steady state distribution. Equations (13.5) are obtained from the general model (13.1). These equations form the *prediction step*, whose application yields the quantities $\mu_{t|t-1}$ and $v_{t|t-1}$

from the top part of (13.5). The remaining part of the Kalman filter is the *revision step*. By an argument similar to that employed in Sect. 12.7.2, we arrive at the relationships:

$$\begin{aligned} & \boldsymbol{m}_{t|t} = \boldsymbol{m}_{t|t-1} + \boldsymbol{k}_t(y_t - \mu_{t|t-1}) \\ \text{and} \quad & \boldsymbol{V}_{t|t} = \boldsymbol{V}_{t|t-1} - v_{t|t-1}\boldsymbol{k}_t\boldsymbol{k}_t', \\ \text{where} \quad & \boldsymbol{k}_t = \boldsymbol{\zeta}_{t|t-1}/v_{t|t-1}. \end{aligned}$$

These expressions provide all the information needed to evaluate the likelihood function for given values of the parameters.

As soon as the assumption of stationarity is dropped, the variances of nonstationary components are infinite and the Kalman filter formulae have no proper limiting form (Ansley and Kohn 1985). Moreover, the likelihood, as defined for stationary time series, is 0 everywhere. The traditional escape from this dilemma is to condition on the first p values of the time series, where p is the number of free nonstationary unobservable components. The density upon which the likelihood is based is then given by

$$p(y_{p+1}, y_{p+2}, \ldots, y_n | \boldsymbol{\theta}, y_1, y_2, \ldots, y_p) = \prod_{t=p+1}^{n} (v_{t|t-1})^{-1/2} \exp\left(-\tfrac{1}{2}(y_t - \mu_{t|t-1})^2 \big/ v_{t|t-1}\right).$$

One approach (Harvey 1989) is based on the assumption that all the unobserved states are nonstationary, so that $p = k$. A set of simultaneous equations is formed by stacking the model equations for the first k periods. The number of unknown unobservable state variables then exactly matches the number of equations and may normally be solved for the unobserved components including the moments of $\boldsymbol{x}_{p|p}$. The Kalman filter is then seeded with the distribution of $\boldsymbol{x}_{p|p}$ in period $p+1$ and used to generate the predictions and associated variance matrices for periods $p+1, p+2, \ldots, n$ needed to evaluate the likelihood function. This approach works in most circumstances, but must be adapted to handle potential complexities such as linear dependencies in the equations, missing values or partially known starting conditions when there is a mix of stationary and nonstationary unobserved state variables. A modern recursive version that allows for these potential complications may be found in de Jong (1991a, b). His algorithm is referred to as an *augmented Kalman filter*.

13.2.2 Convergence of Estimates

As the length of the series t increases, the variance matrix for $\boldsymbol{x}_{t|t}$, defined as $\boldsymbol{V}_{t|t}$ (see (12.18) for this expression in the innovations case) converges to a limiting value, say $\boldsymbol{V}_0$; for a proof, see Anderson and Moore (1979) and Harrison (1997). Harrison's proof applies to the MSOE scheme and does not require

an assumption of Gaussian errors. His approach can be extended using the general form of the Kalman filter outlined in Sect. 12.7. For the MSOE model, this matrix is non-null, but for the innovations model, the limiting value is **0** as shown by Leeds (2000, pp. 78–79). This latter proof is a correction to errors in both the original proof by Caines and Mayne (1970) and a revised proof by the same authors (1971). Thus, it has been shown that, as t increases, the estimates of the state variables in the innovations model will converge in probability to the true values of the unobserved state variables at time t.

Many computer implementations ignore the distinction between states and their estimates. This result suggests that, in sufficiently long series, this practice is justifiable in the innovations framework.

13.3 Reduced Forms

13.3.1 Multi-Disturbance Models

Unobserved components are very useful in the sense that they enable us to specify plausible candidate state space models for the patterns that one may observe in a time series. However, from a strict mathematical perspective, their role is largely redundant. If a time series is stationary, its behavior is essentially determined by its autocorrelation function (ACF). Two state space models may appear to have a different structure because they are based on different states. However, if they yield the same mean, variance and ACF, they are equivalent from a forecasting perspective. Matters are more complicated for nonstationary time series because the unconditional mean and variances are not defined. An appropriate level of differencing may yield a stationary series. In this case, if the same transformations are applied to two models to induce stationarity, and both transformed models have the same mean, variance and ACF, they have the same properties. In the terminology of Chap. 10, the two models have the same minimal state representation.

The Wold representation theorem states that any linear stationary time series can be expressed as a moving average process and that this representation is unique. These moving average representations may involve infinite series and a more parsimonious structure is often achieved by converting to autoregressive moving average (ARMA) processes (Box et al. 1994).

The ARIMA representation is the reduced form corresponding to the *minimal dimension* representation of the state space model.

Common multi-disturbance state space models and their reduced forms are shown in Table 13.1. The right hand sides of the reduced forms are multi-disturbance moving average processes. However, the Granger–Newbold theorem (Granger and Newbold 1986) asserts that

(a) The sum of *uncorrelated* moving average processes is itself a moving average process
(b) The covariance function of the sum is the sum of the component covariance functions

Table 13.1. Reduced forms of multi-disturbance state space models.

Multi-disturbance model	Reduced form
Level	
$y_t = \ell_{t-1} + \varepsilon_t$	$\Delta y_t = \Delta\varepsilon_t + \eta_{t-1}$
$\ell_t = \ell_{t-1} + \eta_t$	
Trend	
$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$	$\Delta^2 y_t = \zeta_{t-1} + \Delta\eta_{t-1} + \Delta^2\varepsilon_t$
$\ell_t = \ell_{t-1} + b_{t-1} + \eta_t$	
$b_t = b_{t-1} + \zeta_t$	
Seasonal	
$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$	$\Delta^2\Delta_m y_t = \Delta_m\zeta_{t-1} + \Delta\Delta_m\eta_{t-1} + \Delta^2\omega_{t-m}$
$\ell_t = \ell_{t-1} + b_{t-1} + \eta_t$	$+ \Delta^2\Delta_m\varepsilon_t$
$b_t = b_{t-1} + \zeta_t$	
$s_t = s_{t-m} + \omega_t$	

In other words, any multiple disturbance moving average process has an equivalent traditional innovations moving average representation. The reduced forms, in terms of the multiple error terms, are shown in Table 13.1. Note that in the table, the difference operators are defined as $\Delta x_t = x_t - x_{t-1}$ and $\Delta_m x_t = x_t - x_{t-m}$. The reduced forms for the innovations models follow immediately when we replace the error terms from the transition equations by the appropriate linear functions of the single source of error.

The reduced forms may be obtained as an equation solving exercise. In any period, the model consists of $k+1$ equations. Stacking the model $k+1$ equations over the k periods $(t-1), \ldots, (t-k)$ gives $k(k+1)$ equations involving the $k+1$ state vectors $\boldsymbol{x}_{t-1}, \ldots, \boldsymbol{x}_{t-k-1}$; these equations also involve $y_{t-1}, \ldots, y_{t-k-1}$ and the disturbances. Because each state vector contains k elements, the number of *state* variables exactly matches the number of equations. Ignoring the possibility of linear dependence for the moment, the stacked equations can be solved for the state variables in terms of the lagged y values and the disturbances. The solution for $\boldsymbol{x}_{t-1}$ found in this manner can be substituted into the measurement equation to yield an expression that no longer depends on the state variables. It is the required reduced form.

In deriving the reduced form, ultimately only the solution for $\boldsymbol{x}_{t-1}$ is required. It is possible to adapt the above procedure to avoid finding the solutions for $\boldsymbol{x}_{t-2}, \ldots, \boldsymbol{x}_{t-k-1}$. The procedure can be understood by placing the stacked equations in tableau form; the tableau is then supplemented by the equation for y_t, which is placed in the final row. We then apply Gaussian elimination to eliminate the state variables from the measurement equation. The approach is illustrated in Example 13.1 for a local level model ($k = 1$).

Example 13.1: Local level model

The relevant equations for the local level may be stacked in detached form as follows:

y_t	y_{t-1}	ℓ_{t-1}	ℓ_{t-2}	ε_t	ε_{t-1}	η_{t-1}
0	0	−1	1	0	0	1
0	1	0	1	0	1	0
1	0	1	0	1	0	0

The aim is to eliminate all the state variables from the final row. The process begins by eliminating the state ℓ_{t-1} by adding the first row to the final row to give

y_t	y_{t-1}	ℓ_{t-1}	ℓ_{t-2}	ε_t	ε_{t-1}	η_{t-1}
0	0	−1	1	0	0	1
0	1	0	1	0	1	0
1	0	0	1	1	0	1

Now ℓ_{t-2} appears in the final row. It is eliminated by subtracting the second row to give:

y_t	y_{t-1}	ℓ_{t-1}	ℓ_{t-2}	ε_t	ε_{t-1}	η_{t-1}
0	0	−1	1	0	0	1
0	1	0	1	0	1	0
1	−1	0	0	1	−1	1

The reduced form is shown in the final row of the third tableau. It was obtained without explicitly solving for the local levels.

It may be observed that only the bottom row was changed by the transformation process in this example. This is not true in general. The rows before the final row may not always have a triangular structure in the columns corresponding to the state variables. Then below-diagonal elements must be eliminated using conventional pivoting operations associated with Gaussian elimination. It is also sensible to undertake simple pivots to ensure that all diagonal elements equal one. Only then should the elements in the final row of the tableau corresponding to state variables be eliminated, as illustrated in the example, to yield the reduced form.

The tableaux associated with this method explode in size when models with more states are considered, and so the derivation of the reduced form is more difficult to illustrate in the available space. Nevertheless, the method is readily coded and is quite tractable when implemented on a computer.

When a zero pivot is encountered, the elements in the pivot column below the pivot are searched for a non-zero value. The row containing this non-zero value is swapped with the pivot row so that resulting new pivot element in the new pivot row is non-zero. Then the reduction algorithm continues as normal.

When no non-zero element lies below a zero pivot element, the required row swap is not possible. In this case, there is not a unique solution for some of the state vectors in terms of the observations, and so the states are not identifiable. This occurs when the model is not of minimal dimension (see Chap. 10).

Identifiability of the state variables is not always necessary in order to derive a unique reduced form. In the ETS(A,A,A) model, the level and seasonal indexes are not identified, yet the reduced form is unique because the linear combination of the state vectors in the measurement equation is unique. The above reduction method can be adapted to handle such cases. As shown in Chap. 10, the solution typically involves the elimination of common factors from the two sides of the reduced model to achieve a canonical form.

In general, the reduced form may be written as

$$y_t = \sum_{i=1}^{k} \phi_i y_{t-i} - \sum_{j=1}^{k} \theta_j \varepsilon_{t-j} + \varepsilon_t, \tag{13.6}$$

where ε_t is formed from all the disturbances associated with period t via the Granger–Newbold theorem. The autoregressive elements may involve unit roots, which can be separated out as in Chap. 10. As is evident from Table 13.1, the moving average component can be expressed as the sum of uncorrelated moving average components:

$$y_t = \sum_{i=1}^{k} \phi_i y_{t-i} + \sum_{j=0}^{k} \boldsymbol{\xi}_j' \boldsymbol{\eta}_{t-j}, \tag{13.7}$$

where $\boldsymbol{\eta}_t$ represents the independent errors in the state equations, as in (13.2b) with coefficients $\boldsymbol{\xi}_j$. The individual moving average components have autocovariance functions that can be combined to provide the autocovariance function for ε_t.

13.3.2 Innovations Models

The triangularization method described in the previous section for finding the reduced form of a multi-disturbance state space model is readily adapted to the innovations state space model. The independence assumption of the disturbances is nowhere used in the algorithm, and so it applies in exactly the same way when the disturbances are correlated. In the particular case where

the disturbances are perfectly correlated, the reduced form of an innovations model can be obtained from the reduced form of a multi-disturbance model using the substitution $\boldsymbol{\eta}_t = \boldsymbol{g}\varepsilon_t$. For example, in the case of a local trend model, the substitutions $\eta_t = \alpha\varepsilon_t$ and $\xi_t = \beta\varepsilon_t$ yield the innovations reduced form $\Delta y_t = -\theta_2\varepsilon_{t-2} - \theta_1\varepsilon_{t-1} + \varepsilon_t$ where $\theta_1 = 2 - \alpha - \beta$ and $\theta_2 = \alpha - 1$.

13.4 Comparison of State Space Models

Multi-disturbance state space models encompass two special cases: the MSOE model where the disturbances are uncorrelated and the innovations form where they are perfectly correlated. It is often thought that the first assumption is less restrictive than the second; the argument is that the MSOE model has many sources of randomness, and should therefore be more flexible than the innovations form.

Paradoxically, the opposite is true. Anderson and Moore (1979) appear to be the first to have asserted, for discrete time contexts at least, that *any* multi-disturbance linear state space model has an equivalent innovations form. Their claim was remarkably general: it encompassed non-invariant as well as invariant state space models. They provided strong evidence that this has to be true by recognizing that the Kalman filter for any multi-disturbance state space model is always expressed in terms of the one-step-ahead prediction error, and that this implies the existence of an innovations model with the same gains. Their proof is opaque and possibly incomplete, so we do not pursue it further.

Hannan and Deistler (1988) proved the conjecture for stationary time series. They relied on transfer functions (i.e., polynomial functions of the lag operator) for their proof. They did not cover nonstationary time series. However, for those nonstationary time series that can be differenced to create a stationary time series, the same basic theory may be applied.

The general result can be stated as follows and the proof is derived by identifying various results presented earlier in the book.

Theorem 13.1. *The following statements hold for linear time series with invariant coefficients and Gaussian disturbances:*

A. Any MSOE model may be represented as an ARIMA model
B. Any innovations model may be represented as an ARIMA model
C. Any ARIMA model may be represented as an innovations model
D. Not all ARIMA models are representable by an MSOE model

Proof. This proof is somewhat informal and proceeds by drawing together results presented earlier in the book:

A. This property was discussed in Chap. 11 and in Sect. 13.3. The property holds provided that (13.7) corresponds to a minimal state space model. That is, the AR component $\phi(L) = 0$ has roots on or outside the unit

circle, the MA component is invertible and any common factors in the two polynomials have been eliminated.

B. This property was demonstrated in Sect. 11.3. The same requirements on the polynomials apply.

C. This property was demonstrated in Sect. 11.5, but we should recall that the innovations model will correspond specifically to an exponential smoothing form only when the polynomial $\theta(L) = 0$ has real roots.

D. This negative statement can be demonstrated by means of counter-examples. Two models are equivalent in this framework if their (differenced) reduced forms have the same autocorrelation function (ACF). We define the autocorrelations by $\rho_j = \gamma_j/\gamma_0$, where $j = 1, 2, \ldots$ and $\gamma_j = \mathrm{Cov}(y_t, y_{t-j})$.

The ACF of the MSOE reduced form depends on the system variances; for the innovations model it is determined by the persistence parameters.

The ACFs for the local level and local trend models are given in Table 13.2; all autocovariances not listed in the table are zero. Examination of the expressions in the table reveals that for the local level model $-0.5 < \rho_1 \leq 0$ for MSOE with the limiting value corresponding to $\sigma_\eta^2 = 0$. The ARIMA scheme has $-0.5 < \rho_1 < 0.5$, with the limits corresponding to $|\theta_1| = 1$. Thus, an ARIMA(0,1,1) model with $\theta_1 < 0$ does not have an MSOE counterpart.

Likewise, for the local trend model, we have $-0.667 < \rho_1 \leq 0$ for the MSOE, but $-0.707 < \rho_1 < 0.707$ for the ARIMA scheme, so that some ARIMA(0,2,2) models do not possess an MSOE form. The derivation is left as an exercise.

A counter-claim to these examples could be that the parameter spaces may be extended by allowing correlation among the disturbances. We explore this conjecture below. □

We may use the entries in Table 13.2 to explore the relationships between the MSOE and innovations models. The general point may be illustrated using the local level model. The first-order autocorrelation is always negative for the MSOE version. It may be either positive or negative in the innovations model. When the autocorrelation is negative, it is always possible to find a

Table 13.2. Reduced forms of common state space models.

Model	Multiple error	Innovations
Local Level	$\gamma_0 = \sigma_\eta^2 + 2\sigma_\varepsilon^2$ $\gamma_1 = -\sigma_\varepsilon^2$	$\gamma_0 = [(\alpha-1)^2 + 1]\sigma_\varepsilon^2$ $\gamma_1 = (\alpha-1)\sigma_\varepsilon^2$
Local Trend	$\gamma_0 = (\sigma_\xi^2 + 2\sigma_\eta^2 + 6\sigma_\varepsilon^2)$ $\gamma_1 = -(\sigma_\eta^2 + 4\sigma_\varepsilon^2)$ $\gamma_2 = \sigma_\varepsilon^2$	$\gamma_0 = \left[(\alpha+\beta-2)^2 + (1-\alpha)^2 + 1\right]\sigma_\varepsilon^2$ $\gamma_1 = -(2-\alpha-\beta)(2-\alpha)\sigma_\varepsilon^2$ $\gamma_2 = (1-\alpha)\sigma_\varepsilon^2$

corresponding value for α by equating the two expressions and solving the quadratic in α to obtain

$$\alpha = -\frac{q}{2} + \sqrt{\left(1 + \frac{q}{2}\right)^2 - 1}, \tag{13.8}$$

where q is the so-called signal-to-noise ratio defined by $q = \sigma_\eta^2/\sigma_\varepsilon^2$.

This analysis serves to illustrate a further point, relating to the relative ease of use of the MSOE and innovations models. As we saw in Sect. 11.3, the relationships between the parameters in the ARIMA and innovations models are linear. Those between the MSOE model and the other two are quadratic, making it more difficult to establish relationships between the sets of parameters.

13.4.1 Size of the Parameter Space

In order to explore the size of the parameter space under different assumptions about the correlations among the error terms in the measurement and transition equations, we revert to a consideration of the general form given in (13.2). The general argument is due to Leeds (2000, pp. 50–56), and the details are given in the Appendix. The argument given there shows that when there are J transition equations, we must consider 2^J possible solutions, and only one of these solutions will satisfy the forecastability conditions. To be specific, we demonstrate the argument for the local trend model, although it applies quite generally.

We first put the problem into an appropriate framework that enables us to apply the linear fractional programming approach described in the Appendix. The notation we now use is specific to this subsection and purely for convenience in the present discussion. The local trend model has three error terms in the general case, and we may write the variance matrix for $(\varepsilon_t, \eta_t, \xi_t)$ as

$$\begin{bmatrix} v_0^2 & \rho_1 v_0 v_1 & \rho_2 v_0 v_2 \\ \rho_1 v_0 v_1 & v_1^2 & \rho_3 v_1 v_2 \\ \rho_2 v_0 v_2 & \rho_3 v_1 v_2 & v_2^2 \end{bmatrix}.$$

Extending the result in Table 13.2, the general form of the lag one autocorrelation for the twice-differenced series is

$$\frac{-(4v_0^2 + 4\rho_1 v_0 v_1 + 2\rho_2 v_0 v_2 + v_1^2 + \rho_3 v_1 v_2)}{(6v_0^2 + 6\rho_1 v_0 v_1 + 2\rho_2 v_0 v_2 + 2v_1^2 + 2\rho_3 v_1 v_2 + v_2^2)}.$$

A comparable expression may be obtained for the autocorrelation at lag two; see Exercise 13.4. All higher-order autocorrelations are zero. We may determine the maximum size of the parameter space by finding the smallest and largest possible values for each autocorrelation, provided the extremes are achieved for the same choices of the correlations.

If we fix the value of $\boldsymbol{v} = (v_0, v_1, v_2)'$, the numerator and denominator of the autocorrelation are linear in the correlations, and we may maximize (minimize) the value of the expression using linear fractional programming. The details are given in the Appendix. We find that the same extreme solutions apply whatever the value of $\boldsymbol{v}$, and so we conclude that the size of the parameter space is maximized when the errors are perfectly correlated. However, an innovations model with J transition equations still has 2^J possible solutions, and we now proceed to select a unique solution from this set.

The local trend model has $J = 2$, and the transition equations have the error terms $(g_1\varepsilon_t, g_2\varepsilon_t)$. From Table 13.1 we may write the right hand side of the reduced form equation for the local trend model as

$$\varepsilon_t - (2 - g_1 - g_2)\varepsilon_{t-1} - (g_1 - 1)\varepsilon_{t-2} \equiv (1 - \theta_1 L - \theta_2 L^2)\varepsilon_t.$$

The forecastability conditions may be written as:

$$|\theta_2| < 1, \quad 1 - \theta_1 - \theta_2 > 0, \quad 1 + \theta_1 - \theta_2 > 0.$$

These conditions reduce to the requirements that $(g_1 > 0, g_2 > 0)$, which establishes the uniqueness of the solution. The reader is asked to verify these conditions in Exercise 13.5.

13.4.2 Seasonal Models

In order to compare the seasonal models we make use of the autocovariance generating function (ACGF) for the differenced series. Consider an ARIMA model written in moving average form with the error variance equal to 1 (without loss of generality) and the auxiliary polynomial

$$\theta(z) = 1 - \theta_1 z - \theta_2 z^2 - \cdots - \theta_q z^q - \cdots . \tag{13.9}$$

The ACGF is then defined as:

$$C(z) = \theta(z)\theta(z^{-1}). \tag{13.10}$$

The coefficient of z^j is the autocovariance at lag j. Thus $\gamma_0 = 1 + \theta_1^2 + \theta_2^2 + \cdots$, $\gamma_1 = -\theta_1 + \theta_1\theta_2 + \cdots$, and so on. The general forms for the seasonal models are cumbersome, and it is convenient to summarize them in somewhat different ways. Thus, for the innovations model, using the canonical reduced form given in Example 11.6, (13.9) becomes

$$\begin{aligned}\theta(z) = 1 &- (1 - \alpha - \beta)z + \beta(z^2 + \cdots + z^{m-1}) - (1 - \beta - \gamma)z^m \\ &+ (1 - \alpha - \gamma)z^{m+1}.\end{aligned}$$

For the MSOE model, it is easier to specify the autocovariances directly. Again using the canonical reduced form, we arrive at the expressions:

$$\begin{aligned}
\gamma_0 &= (m\sigma_\xi^2 + 2\sigma_\eta^2 + 2\sigma_\omega^2 + 6\sigma_\varepsilon^2),\\
\gamma_1 &= (m-1)\sigma_\xi^2 - (\sigma_\omega^2 + 2\sigma_\varepsilon^2),\\
\gamma_j &= (m-j)\sigma_\xi^2, \quad j = 1, 2, \ldots, m-2,\\
\gamma_{m-1} &= \sigma_\xi^2 + \sigma_\varepsilon^2,\\
\gamma_m &= -(\sigma_\eta^2 + 2\sigma_\varepsilon^2),\\
\gamma_{m+1} &= \sigma_\varepsilon^2.
\end{aligned}$$

For these two seasonal models, any attempt to equate autocovariances of the same order leads to more equations than unknowns. No solution exists that matches the autocovariances, other than the degenerate form with $\sigma_\omega^2 = 0$. Thus, the two models are not equivalent. Interestingly, McKenzie (1976) derived an ARIMA representation of this additive Holt-Winters scheme. Careful reading of his paper reveals that he used an innovations form of the state space model to obtain the result. The covariance expressions just derived do not allow a simple mapping from the state space parameters to the ARIMA coefficients. More generally, because the autocovariances are typically quadratic in the moving average parameters, it is only in special cases that explicit solutions are available for the mapping from the MSOE model to its ARIMA reduced form. There can be multiple solutions to such equations, but the requirement of invertibility ensures that there is at most one acceptable solution.

13.4.3 Nonlinear Models

We saw in Chap. 4 that it was possible to specify nonlinear and heteroscedastic schemes using the innovations form, and that the resulting (albeit approximate) Gaussian likelihood was readily obtained, as seen in Chap. 5. Comparable models may be specified in the MSOE framework, but computational difficulties immediately arise. The probability density function now involves terms for each of the unobserved errors and there is no simple way to integrate these out to obtain the likelihood for the unknown parameters. We could make use of Markov Chain Monte Carlo (MCMC) methods, but Gaussian likelihood remains an approximation and adding an *extra* layer of simulations adds to the computational burden.

13.5 Smoothing and Filtering

Harvey and Koopman (2000) showed that the MSOE scheme leads to optimal symmetric two-sided smoothers. These were defined for an infinite series, although applications will clearly involve truncation after a finite number

of terms. This smoother corresponds to the Wiener-Kolmogorov (WK) filter. They also noted that when the components are correlated, as in the innovations formulation, the resulting signal extraction filter is asymmetric. Indeed, the perfect correlation among the components of the innovations model led to our observation in Sect. 12.6 that the two-sided filter does not improve the estimates of the state variables asymptotically. However, the WK filter remains available, once we recognize that its role is to smooth the series, not to estimate the state variables as such. Because any innovations model may be expressed in ARIMA form, an appropriate WK filter may be developed within that framework.

The following example illustrates how an appropriate WK smoother can be constructed.

Example 13.2: Local level model

Consider the local level model, written as the reduced ARIMA(0,1,1):

$$(1-L)y_t = [1-(1-\alpha)L]\varepsilon_t.$$

The (doubly infinite) WK filter is given by:

$$\ell_{S,t} = \frac{\alpha^2 y_t}{[1-(1-\alpha)L][1-(1-\alpha)L^{-1}]} = \frac{\alpha}{2-\alpha}\sum_{j=-\infty}^{\infty}(1-\alpha)^{|j|}y_{t-j}.$$

This smoother also corresponds to the two-sided Beveridge–Nelson (BN) filter given by Proietti and Harvey (2000), although it should be noted that the filter is admissible only for $0 < \alpha < 1$. The WK and BN filters often do not have the same form.

As pointed out by Gijbels et al. (1999), when exponential smoothing is interpreted as a kernel estimate, simple exponential smoothing is the natural forecast and the filter given above is the natural smoother.

The approach just described provides a smoothed estimator for the mean of the process, and we now turn to consider the individual components. Key elements in the analysis of economic time series are the creation of the deseasonalized series and the creation of a smoothed trend. Bell (1984) and Burridge and Wallis (1988) extended the WK filter to nonstationary series to enable the extraction of unobserved components.

One way to develop a WK filter for the components of an innovations process is to generate the corresponding ARIMA model and then apply a canonical decomposition, such as that developed by Hillmer and Tiao (1982). However, if we recall from Sect. 13.2.2 that the estimates of the state variables converge to their true values, a much simpler approach

is possible. We may construct the seasonally adjusted or detrended series directly, and then smooth the remaining components. This is illustrated in the next example.

Example 13.3: Seasonal levels model

Consider the following innovations model, which should also include the appropriate normalization as described in Chap. 8:

$$\begin{aligned} y_t &= \ell_{t-1} + s_{t-m} + \varepsilon_t, \\ \ell_t &= \ell_{t-1} + \alpha\varepsilon_t, \\ s_t &= s_{t-m} + \gamma\varepsilon_t. \end{aligned}$$

We may generate the approximately detrended series as:

$$z_{1t} = y_t - \ell_{t|n} \approx s_{t-m} + \varepsilon_t.$$

It follows from Example 13.2 that the smoothed seasonal components may be computed as:

$$s_{S,t} \approx \frac{\gamma^2 z_{1t}}{[1-(1-\gamma)L^m][1-(1-\gamma)L^{-m}]} = \frac{\gamma}{2-\gamma} \sum_{j=-\infty}^{\infty} (1-\gamma)^{|j|} z_{1,t-jm}.$$

In turn, the seasonal components lead to the deseasonalized series:

$$z_{2t} = y_t - s_t \approx \ell_t + \varepsilon_t.$$

The smoothed trend is then given by:

$$\ell_{S,t} \approx \frac{\alpha^2 z_{2t}}{[1-(1-\alpha)L][1-(1-\alpha)L^{-1}]} = \frac{\alpha}{2-\alpha} \sum_{j=-\infty}^{\infty} (1-\alpha)^{|j|} z_{1,t-j}.$$

We may iterate between the seasonal and trend components until convergence is obtained, although the differences may be expected to be small provided the series is of reasonable length.

In summary, we observe that while the primary motivation for using the innovations approach is that it is more directly beneficial for forecasting (the focus of this text), smoothing and filtering operations may also be performed within the innovations framework.

13.6 Exercises

Exercise 13.1. Consider a state space model in the general form of (13.3) with

$$\boldsymbol{w}' = (1, 0, \ldots, 0), \quad \boldsymbol{g}' = (1, \psi_1, \ldots, \psi_{k-1}) \text{ and } \boldsymbol{F} = \begin{bmatrix} 0 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \ldots & 1 \\ \phi_k & \phi_{k-1} & \cdots & \cdots & \phi_1 \end{bmatrix}.$$

The state vector is defined as $\boldsymbol{x}_t = (y_t, y_{t+1|t}, \ldots, y_{t+k-1|t})$ and ε_t is deleted from the measurement equation. Show that this model reduces to an ARMA$(k, k-1)$ model.

Exercise 13.2. Consider the innovations model with measurement equation (13.4) used in place of (13.1). Show that the form of the model given by (13.2) still applies, with revised coefficients $\boldsymbol{w}_1 = \boldsymbol{F}'\boldsymbol{w}$ and $\boldsymbol{g}_1 = \frac{\boldsymbol{g}}{1+\boldsymbol{w}'\boldsymbol{g}}$.

Exercise 13.3. Show that the reduced forms of the two MSOE schemes given in Sect. 13.1.1 result in the same ARIMA reduced forms.

Exercise 13.4. Show that the general form of the lag 2 autocovariance for the local trend model (in the notation of Sect. 13.4.1) is

$$v_0^2 + \rho_1 v_0 v_1 + \rho_2 v_0 v_2.$$

Hence show that the first and second order autocorrelations have the same set of conditions for extreme values.

Exercise 13.5. Show that the conditions for forecastability discussed in Sect. 13.4.1 lead to a unique local trend model with a maximal parameter space.

Appendix: Maximizing the Size of the Parameter Space

In a seminal paper on Linear-Fractional Programming (LFP), Charnes and Cooper (1962) showed that the LFP optimization problem

$$\max_u \frac{\sum_j A_j u_j}{\sum_j B_j u_j}, \qquad \text{subject to} \quad 0 \le u_j \le c_j \quad \text{for all } j,$$

may be reformulated as a linear program of the form:

$$\max_u \sum_j A_j u_j \qquad \text{subject to} \quad \sum_j B_j u_j = c \quad \text{and} \quad 0 \le u_j \le c_j \quad \text{for all } j.$$

In our application, the denominator is always a strictly positive variance term and the $\{u_j\}$ represent either the positive or negative parts of correlation coefficients, so that $c_j = 1$ for all j.

When there are J transition equations and one measurement equation, the joint error distribution involves $K = J(J+1)/2$ correlation coefficients. The LFP optimization is subject to $2K$ constraints and K non-negativity constraints on the correlations plus one equality constraint. By inspection, we can see that $K-1$ of the correlations must each take on one of the three values $(-1, 0, +1)$; the remaining correlation is then determined by the equality constraint. We now proceed to incorporate additional features of the particular problem to arrive at a unique solution:

- A simple reparameterization of the problem (replacing each correlation ρ by $\rho^* = \rho + 1$) serves to demonstrate that the zero values are internal solutions and can be ignored.
- We now have that $K-1$ of the correlations are ± 1; it follows that the remaining correlation must be ± 1.
- We can now return to the state space formulation, because the correlations are generated by the $J+1$ terms $(\varepsilon_t, g_1\varepsilon_t, g_2\varepsilon_t, \ldots, g_J\varepsilon_t)$. The J coefficients g_j give rise to the 2^J possible solutions after setting the coefficient in the measurement equation equal to $+1$, without loss of generality.
- Finally, we may demonstrate that the only solution to satisfy the forecastability conditions is that with all $g_j > 0$. The argument for the local trend model is illustrated in Sect. 13.4.1.

In principle, other formulations may provide parameter spaces of equal size for specific cases, but there is no loss in restricting attention to the innovations models.