

State-Space Modelling

by Kevin Kotzé

State-space models provide an encompassing framework to time series modelling. They deal with dynamic time series problems that involve unobserved variables and parameters that describe the evolution in the state of the underlying system. This area of mathematical statistics is relevant to many aspects of econometric research, as we often encounter unobserved variables that may be included in a model: output gaps, business cycles, expectational values of certain variables, permanent income streams, *ex ante* real interest rates, reservation wages, etc. This framework is also used to estimate values for the parameters and unobserved variables in Dynamic Stochastic General Equilibrium models. In addition, this framework is also relevant to those who are interested in financial research and it is used in many variants of stochastic volatility models.

The basic methodology for state-space modelling assumes that the system that we would like to investigate is determined by an unobserved series of vectors, $\{\alpha_1, \dots, \alpha_n\}$, that are associated an observed series of observations, $\{y_1, \dots, y_n\}$. The relationship between the α_t 's and the y_t 's is specified by the state-space model and the purpose of state-space analysis is to infer the relevant properties of the α_t 's from our knowledge of the state-space model and the realisation of the observations $\{y_1, \dots, y_n\}$.¹

1 An Intuitive Example of a State Space Model

A time series is a set of observations, $\{y_1, \dots, y_n\}$, ordered in time that may be expressed in additive form.²

$$y_t = \mu_t + \gamma_t + \varepsilon_t \quad t = 1, \dots, T \quad (1.1)$$

where,

- μ_t is a slowly varying component called the *trend*
- γ_t is a periodic component of fixed period called the *seasonal*
- ε_t is an irregular component called the *error*

To develop suitable models for μ_t and γ_t we may choose to use of a random walk process to describe the scalar series α_t ,³ such that,

$$\alpha_{t+1} = \alpha_t + \eta_t \quad \eta_t \sim \text{i.i.d.} \mathcal{N}(0, W_\eta) \quad (1.2)$$

If $\mu_t = \alpha_t$ (where α_t is a random walk), $\gamma_t = 0$ (no seasonal is present), and all variables are normally distributed then we may rewrite equations (1.1) and (1.2) as,⁴

$$y_t = \alpha_t + \varepsilon_t \quad \varepsilon_t \sim \text{i.i.d.} \mathcal{N}(0, V_\varepsilon) \quad (1.3)$$

$$\alpha_{t+1} = \alpha_t + \eta_t \quad \eta_t \sim \text{i.i.d.} \mathcal{N}(0, W_\eta) \quad (1.4)$$

Such a model has state-space characteristics as we have:

- a **measurement equation** that describes the relation between the observed variables $\{y_1, \dots, y_n\}$ and the unobserved state variables (α_t 's)
- a **state equation** that reflects the dynamics of the unobserved state variables $\{\alpha_1, \dots, \alpha_n\}$

The objective of state-space methodology is to infer the relevant properties of the α_t 's from our knowledge of the observations $\{y_1, \dots, y_n\}$. Although it would be pretty straightforward to find a solution to this problem in a stationary model, the inclusion of the random walk α_t , would imply that the distributions of the random variables y_t and α_t , depend on t . This would make for rather cumbersome multivariate calculations that could be used to find the conditional mean of α_t , as well as the variance and covariances of ε_t and η_t , given $\{y_1, \dots, y_n\}$.

1.1 The Mathematics of State Space Modelling

Although the state-space form is ideally suited to investigations that involve dynamic time series models, where we include unobserved components, it also provides a unified representation for a wide range of ARIMA and time varying regression models.⁵ Indeed, this framework is also flexible enough to encapsulate different specifications of non-parametric and nonlinear spline regression models.

An important feature of these models is that they pay particular attention to the set of m state variables that evolve over time. These state variables may be subject to systematic distortions and an element of noise. The respective state variables could then be contained in an $m \times 1$ vector, which is denoted α_t , while the N observed variables may then be described by an $N \times 1$ vector, y_t . This allows for the derivation of the measurement equation;

$$y_t = F_t \alpha_t + S_t \varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d.} \mathcal{N}(0, V_\varepsilon) \quad (1.5)$$

where F_t and S_t are fixed matrices for the respective coefficients of order $N \times m$ and $N \times r$. In this case, r refers to the dimensions of the disturbance vector in the measurement equation. ε_t is a $r \times 1$ vector with zero mean and covariance matrix, V_ε .⁶

The state equation could then be described as;

$$\alpha_{t+1} = G_t \alpha_t + R_t \eta_t, \quad \eta_t \sim \text{i.i.d.} \mathcal{N}(0, W_\eta) \quad (1.6)$$

where G_t and R_t are fixed coefficient matrices of order $m \times m$ and $m \times g$. In this case g refers to the dimensions of the disturbance vector in the state equation. η_t is a $g \times 1$ vector with zero mean and covariance matrix, W_η .

The disturbances in the measurement and state equations are taken to be uncorrelated over all time periods. They are also assumed to uncorrelated with the initial state vector α_0 , which is to say;

$$\begin{Bmatrix} \varepsilon_t \\ \eta_t \end{Bmatrix} \sim \text{i.i.d.} \mathcal{N} \left[0, \begin{pmatrix} V_\varepsilon & 0 \\ 0 & W_\eta \end{pmatrix} \right]$$

and

$$\mathbb{E} [\alpha_0 \eta_t'] = 0, \quad \mathbb{E} [\alpha_0 \varepsilon_t'] = 0$$

To summarise the parameters further, the coefficient matrices may be included in Φ and the covariance matrices for the error terms could be included in Ω , such that the solution to this problem is found after deriving:

$$\Phi = \begin{Bmatrix} F_t \\ G_t \end{Bmatrix}, \quad \Omega = \begin{Bmatrix} V & 0 \\ 0 & W \end{Bmatrix}.$$

which represent the unknowns in any standard regression model (where we usually make strong assumptions about the covariance matrix that are latter subject to a barrage of tests).

The structure of these matrices are particularly important, as they need to be specified before the model is estimated and they can get a little tricky when you have several equations. For example the model,

$$\begin{aligned}y_t &= \mu_t + \varepsilon_t \\ \mu_{t+1} &= \mu_t + \beta_t + \eta_t \\ \beta_{t+1} &= \beta_t + \zeta_t\end{aligned}$$

may have the coefficient and covariance matrices,

$$\Phi = \begin{Bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{Bmatrix}, \quad \Omega = \begin{Bmatrix} V_\varepsilon & 0 & 0 \\ 0 & W_\eta & 0 \\ 0 & 0 & W_\zeta \end{Bmatrix}.$$

1.2 The practicalities of formulating a State Space model

There are several software packages that have pre-programmed routines that may assist in the formulation and estimation of State Space models. Within the **R** software environment, these include `StructTS`, `sspir`, `dse`, `d1m`, `dlmodeler`, `rucm`, `MARSS`, `KFAS`, `RWinBugs`, `RStan`, etc. Alternatively, you could also use one of the toolboxes that have been developed for **Python**, **Matlab**, **EViews**, **Oxmetrics**, **Ox**, **RATS** or **Gauss**.

In addition to the estimation of the parameters in the model, we would also usually need to estimate values for the unobserved variables in this case, which are used to evaluate the likelihood function. Since the unobserved variables would take the form of random variables it is usually more convenient to make use of Bayesian estimation techniques, where both the parameters and unobserved variables are modelled as random variables.

2 Examples

2.1 The local level model

A simple example of a state-space model is the local level model, where the level component (or intercept term) is allowed to vary over time. It may be formulated by defining the respective *measurement* and *state* equations as,

$$\begin{aligned}y_t &= \mu_t + \varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d.}\mathcal{N}(0, V_\varepsilon) \\ \mu_{t+1} &= \mu_t + \xi_t, \quad \xi_t \sim \text{i.i.d.}\mathcal{N}(0, W_\xi)\end{aligned}$$

This would imply that the matrices for the general form of the model would simplify to,⁷

$$\alpha_t = \mu_t, \quad \eta_t = \xi_t, \quad F_t = G_t = S_t = R_t = 1, \quad W = W_\xi, \quad V = V_\varepsilon,$$

In the case of the local level model, note that the dynamic properties relating to the state of the system at time $t + 1$ are expressed as a function of the state of the system at time t . If $\xi_t = 0$ for $t = 1, \dots, T$ the model would reduce to a traditional linear regression model that may be solved

analytically. However, if ξ_t is modelled as a stochastic process, then we are not able to solve the model analytically and would need to make use of an iterative optimisation procedure (such as maximum likelihood, nonlinear least squares, or MCMC methods).

By way of example, we have applied this model to the South African deflator for the period 1960 to 2014, where the series is represented as the logarithm change in the year-on-year price index. The results are depicted in Figure 1, where we see how the deflator has changed over time. Note how the state equation provides a smoothed version of the observed measure of inflationary pressure, which represents the random walk component of this particular time series.

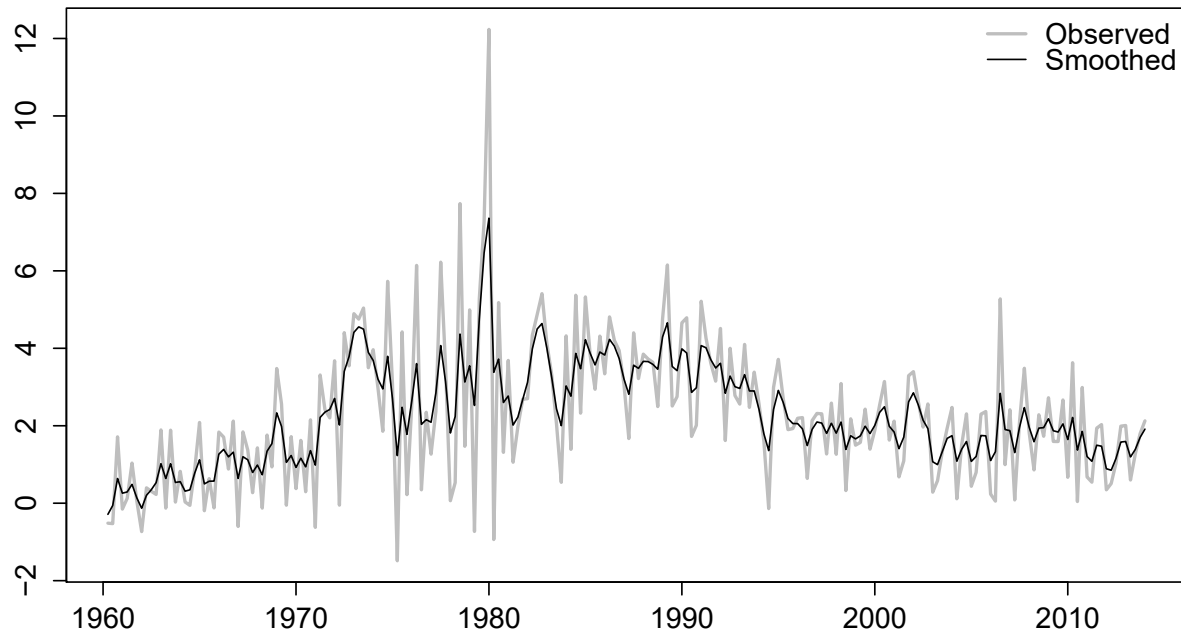


Figure 1: **Local level model - SA deflator (1960Q2-2014Q1)**

In terms of the in-sample statistics that are used to describe the goodness-of-fit of this model, the value of the negative log-likelihood is 261.5, the variance of the irregular component (\hat{V}) is 0.233 and the variance of the level (\hat{W}) is 0.014. The value for the level, μ , at the final state at period 2014Q1 is 1.91%. The residuals for the measurement equation are provided in Figure 2, where we note that they would appear to represent white noise.⁸

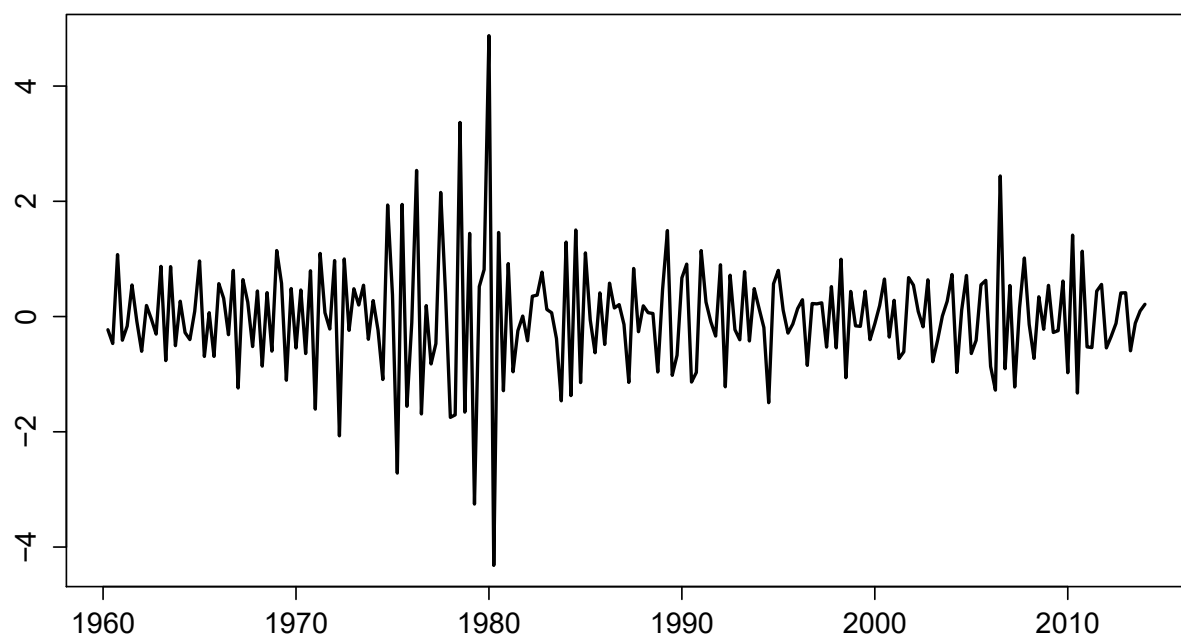


Figure 2: **Local level model - Measurement equation residuals**

To compare different state-space models we often make use of the information criterion, such as that of the AIC, which in this case may be formulated as:

$$AIC = [-2 \log \ell + 2(k)] = 526.9$$

where $\log \ell$ is the value of the negative log-likelihood function that is to be maximised and k is the number of parameters that are to be estimated. When applying this procedure to compare models, smaller values denote better fitting models (or when the result is negative, we prefer models that have values that are closer to $-\infty$). This technique is extremely helpful when comparing the in-sample fit of different models, as it compensates for the number of parameters in the model specification.

2.2 The local level trend model

The local level trend model is formulated with the aid of two state equations that include an additional slope component, v_t , to the specification of the local level model. It may be derived as follows

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d.} \mathcal{N}(0, V_\varepsilon) \\ \mu_{t+1} &= \mu_t + v_t + \xi_t, \quad \xi_t \sim \text{i.i.d.} \mathcal{N}(0, W_\xi) \\ v_{t+1} &= v_t + \zeta_t, \quad \zeta_t \sim \text{i.i.d.} \mathcal{N}(0, W_\zeta) \end{aligned}$$

Or alternatively,

$$\begin{aligned} \alpha_t &= \begin{pmatrix} \mu_t \\ v_t \end{pmatrix}, \quad \eta_t = \begin{pmatrix} \xi_t \\ \zeta_t \end{pmatrix}, \quad G_t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad F_t = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad S_t = 1, \dots \\ W &= \begin{pmatrix} W_\xi & 0 \\ 0 & W_\zeta \end{pmatrix}, \quad R_t = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad V = V_\varepsilon \end{aligned}$$

Note that in this case, v_t is the slope of the trend component, which differs to the slope of the regression line. Hence, this parameter is not the same as the coefficient in the classic regression model.

When all state disturbances ξ_t and ζ_t are set to zero it is easy to see that;

$$y_t = \mu_1 + v_1 g_t + \varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d.} \mathcal{N}(0, V_\varepsilon)$$

where g_t is a variable that represents time (i.e. $g_t = t$). This example would therefore represent a traditional linear regression model, where we regress an observed variable on time and a constant.⁹

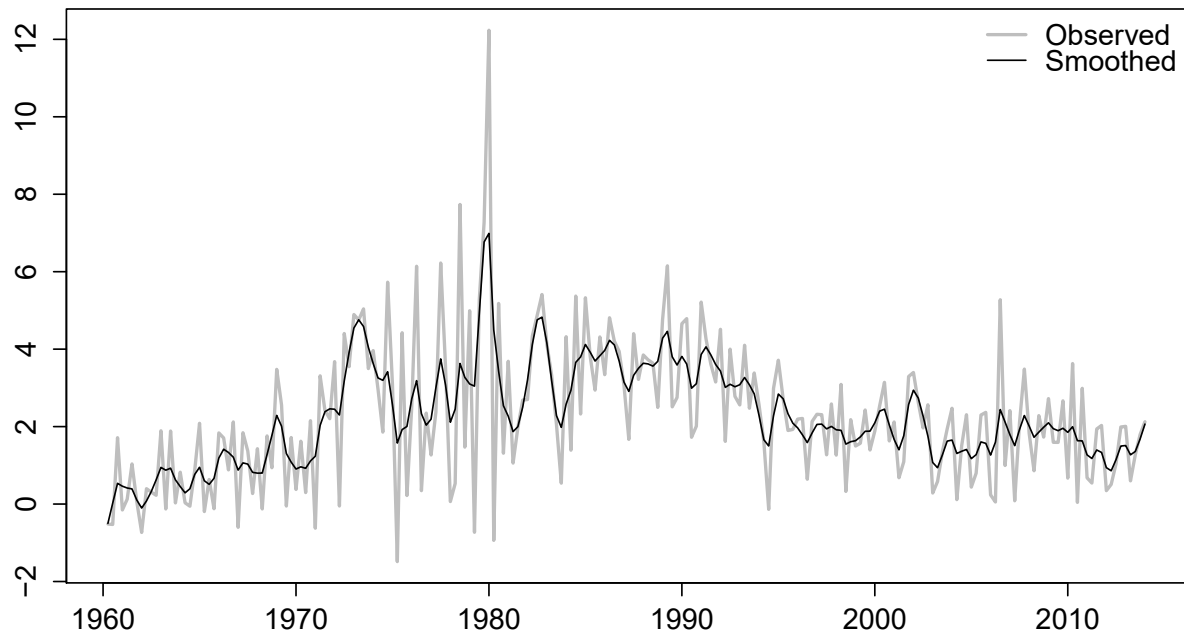


Figure 3: **Local level trend model - SA deflator (1960Q2-2014Q1)**

When we make use of this specification to model the South African deflator data, where both the level and the slope may vary over time, then we observe the results that are contained in Figure 3. The results for the stochastic slope could then be graphed separately in Figure 4. In this case the slope component does not appear to contain a great deal of information and is difficult to interpret.

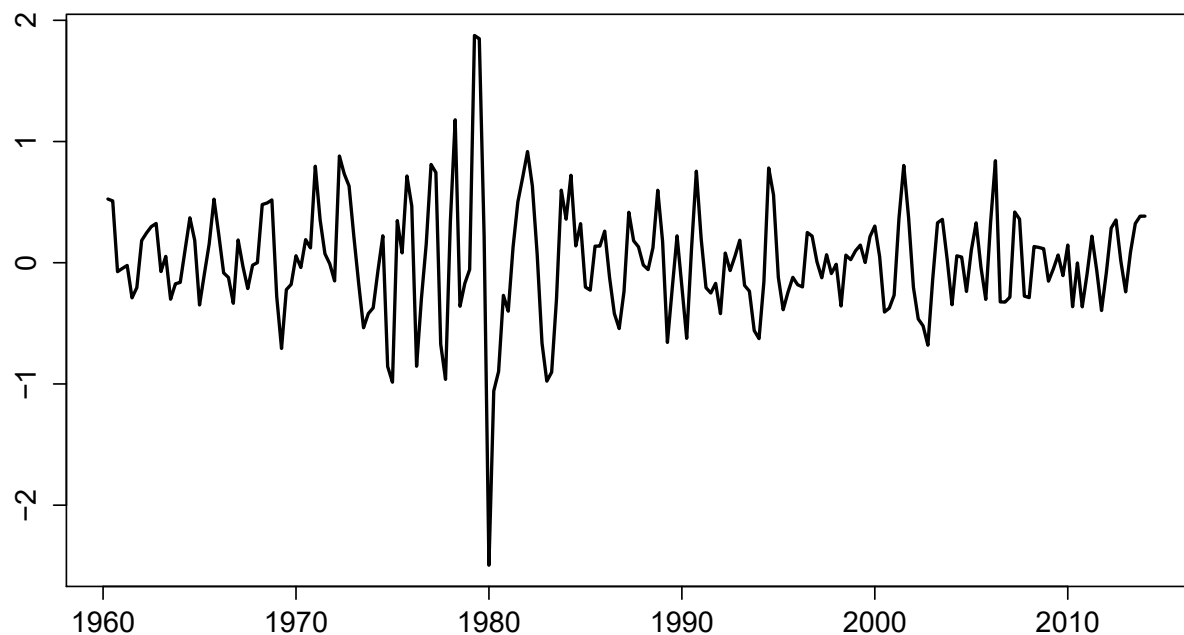


Figure 4: **Local level trend model - Stochastic slope**

The statistical results for this model are similar to those that were presented previously, where we note that the value of the negative log-likelihood is 341.6, the variance of the irregular component (\hat{V}_{ε}) is 0.23, the variance of the level (\hat{W}_{ξ}) is 0.04, and the variance of the trend (\hat{W}_{ζ}) is 0.002. The value for the level, μ , at the final state during period 2014Q112 is 2.07% and the value for the AIC is 689.15. This is greater than that which was produced for the local level model, which would suggest that the relative fit of the local level trend model is not as good.

2.3 The local level trend model with a seasonal

When modelling the behaviour of a time series, one should be conscious of the possibility that the data is influenced by seasonal characteristics. For quarterly data, such a seasonal component could be modelled with the framework

$$y_t = \mu_t + \gamma_{1,t} + \varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d.} \mathcal{N}(0, V_\varepsilon) \quad (1)$$

$$\mu_{t+1} = \mu_t + v_t + \xi_t, \quad \xi_t \sim \text{i.i.d.} \mathcal{N}(0, W_\xi) \quad (2)$$

$$v_{t+1} = v_t + \zeta_t, \quad \zeta_t \sim \text{i.i.d.} \mathcal{N}(0, W_\zeta) \quad (3)$$

$$\gamma_{1,t+1} = -\gamma_{1,t} - \gamma_{2,t} - \gamma_{3,t} + \omega_t, \quad \omega_t \sim \text{i.i.d.} \mathcal{N}(0, W_\omega) \quad (2.1)$$

$$\gamma_{2,t+1} = \gamma_{1,t}, \quad (2.2)$$

$$\gamma_{3,t+1} = \gamma_{2,t} \quad (2.3)$$

Or alternatively,¹⁰

$$\alpha_t = \begin{pmatrix} \mu_t \\ v_t \\ \gamma_{1,t} \\ \gamma_{2,t} \\ \gamma_{3,t} \end{pmatrix}, \quad \eta_t = \begin{pmatrix} \xi_t \\ \zeta_t \\ \omega_t \end{pmatrix}, \quad G_t = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \quad F_t = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

$$S_t = 1, \quad W = \begin{pmatrix} W_\xi & 0 & 0 \\ 0 & W_\zeta & 0 \\ 0 & 0 & W_\omega \end{pmatrix}, \quad R_t = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad V = V_\varepsilon$$

where the γ 's refer to the seasonal components and the disturbance ω allows for the seasonal to change over time. When adding a seasonal component to a state-space model we usually require several state equations (i.e. frequency -1). In the above example, the last two equations, which represent identities, suggest that γ_1 follows γ_2 and γ_2 follows γ_3 .

When we allow both the level and the slope to vary, and include a seasonal variable for quarterly data (which would imply that we would need to included three different γ components to describe the South African deflator), we are able to produce the results that are contained in Figure 5. The separate graph for the seasonal component is the depicted in Figure 6, which includes a stochastic term.

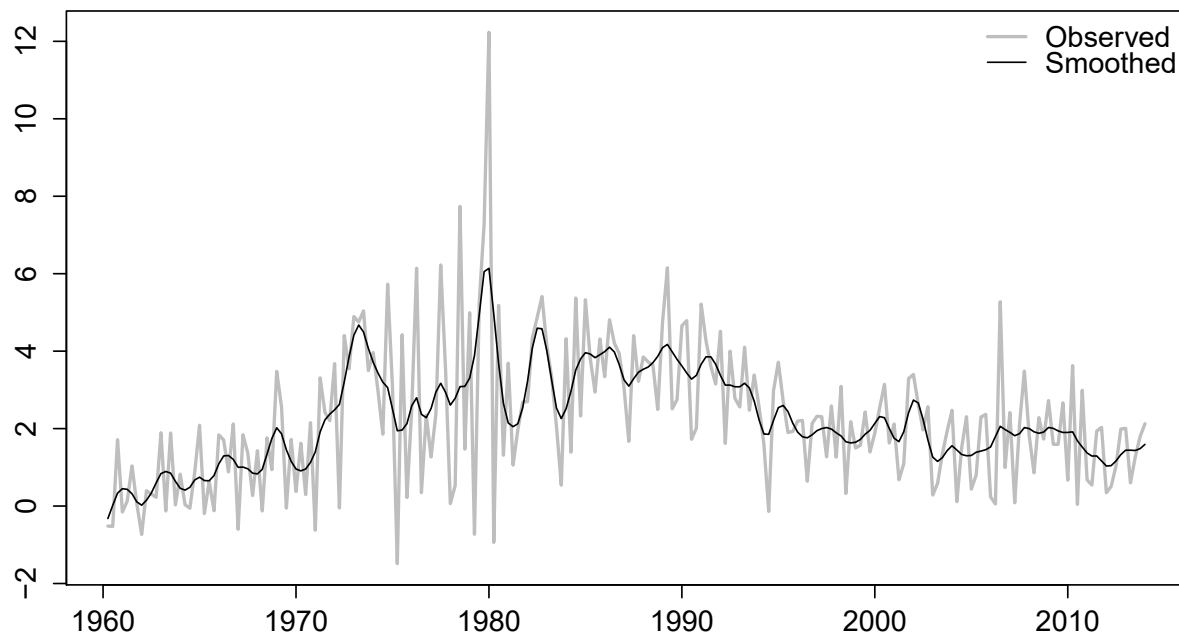


Figure 5: **Local level trend model with seasonal - SA deflator (1960Q1-2014Q1)**

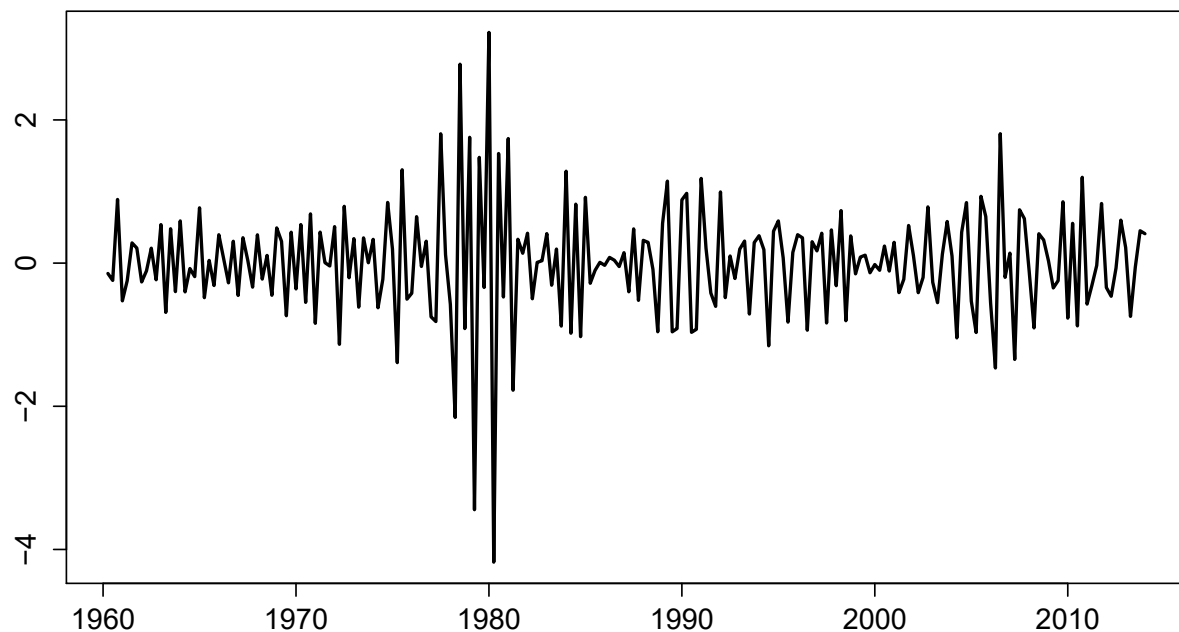


Figure 6: **Local level trend model with seasonal - Seasonal component**

The graph for the seasonal component suggests that there has not been greater seasonal variation around 1980. Once again, the statistical results for this model are similar to those above where we note that the value of the negative log-likelihood is 371.43.¹¹ Once again, the *AIC* is greater than the local level model. Therefore, this model fails to improve upon the fit of the first model.

2.4 The local level model with an intervention variable

In certain cases we may be interested in an assessment of the impact of a structural change on a particular time series. In a state-space framework, such effects can be described by adding intervention variables to any of the above models. Structural changes may result in a *level shift*, where the value of the level of the time series suddenly exhibits a permanent change at that point in time where the structural change takes place. Alternatively, it may be that a *slope shift* is experienced,

where the value that is attached to the slope coefficient experiences a permanent change after the structural break. A third possibility is that of a *pulse* effect, where the value of the level suddenly changes and then returns to the previous level (i.e. prior to the structural change).

To determine the impact of a structural change in a local level model, one could estimate;

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, \quad \varepsilon_t \sim \text{i. i. d. } \mathcal{N}(0, V_\varepsilon) \\ \mu_{t+1} &= \mu_t + \lambda_t w_t + \xi_t, \quad \xi_t \sim \text{i. i. d. } \mathcal{N}(0, W_\xi) \\ \lambda_{t+1} &= \lambda_t \end{aligned}$$

where w_t is a dummy variable.¹² To investigate whether a level shift has occurred we would set the dummy to zero before the proposed structural change, and at one thereafter. If we wanted to determine whether this change was the result of a change in slope then we could include a slope coefficient in state equation for the intervention variable as we did previously. Alternatively, we could consider whether this break had a pulse effect, which would involve setting the dummy variable to 1 only at the point of the proposed structural change. We could then compare the information criteria from the respective models to see which of the models fits the data best.

To consider if there is a structural break in the level, using the above model, then we would consider whether the coefficient estimate for λ_t is significantly different to zero. Note that in the above example, we assume that λ_t is a fixed regression coefficient that will take on a single value. In this sense, one could employ the methodology that is presented in Chow (1960) for this analysis. Similarly, if we wanted to test for a structural break in the slope, then we would need to include the dummy and accompanying coefficient in the slope equation (rather than in the level).

2.5 The local level model with an explanatory variable

To describe the effect of explanatory variables on a time series within a state-space framework, we add the explanatory variables to the measurement equation of the model, such that the general representation produces;

$$\begin{aligned} y_t &= \mu_t + \beta_t x_t + \varepsilon_t, \quad \varepsilon_t \sim \text{i. i. d. } \mathcal{N}(0, V_\varepsilon) \\ \mu_{t+1} &= \mu_t + \xi_t, \quad \xi_t \sim \text{i. i. d. } \mathcal{N}(0, W_\xi) \\ \beta_{t+1} &= \beta_t + \tau_t, \quad \tau_t \sim \text{i. i. d. } \mathcal{N}(0, W_\tau) \end{aligned}$$

Should we wish to do so, we could then include an additional state equation to allow for a time-varying parameter. In this case, the coefficient β is allowed to vary over time. Note that if we removed the stochastic components, τ and ξ , then the above model would represent a traditional linear regression model.

In the following example, we make use of the dataset from Durbin and Watson (1951), which includes per capita consumption of spirits in the UK, per capita income, and the relative price of spirits on an annual basis from 1870 to 1938. In this example we allow for a stochastic level that describes changes in tastes and habits that cannot be modelled explicitly. We also allow the β parameters to be time-varying, so that the regression coefficient would reflect current behaviour. The results of this model are shown in Figure 7.

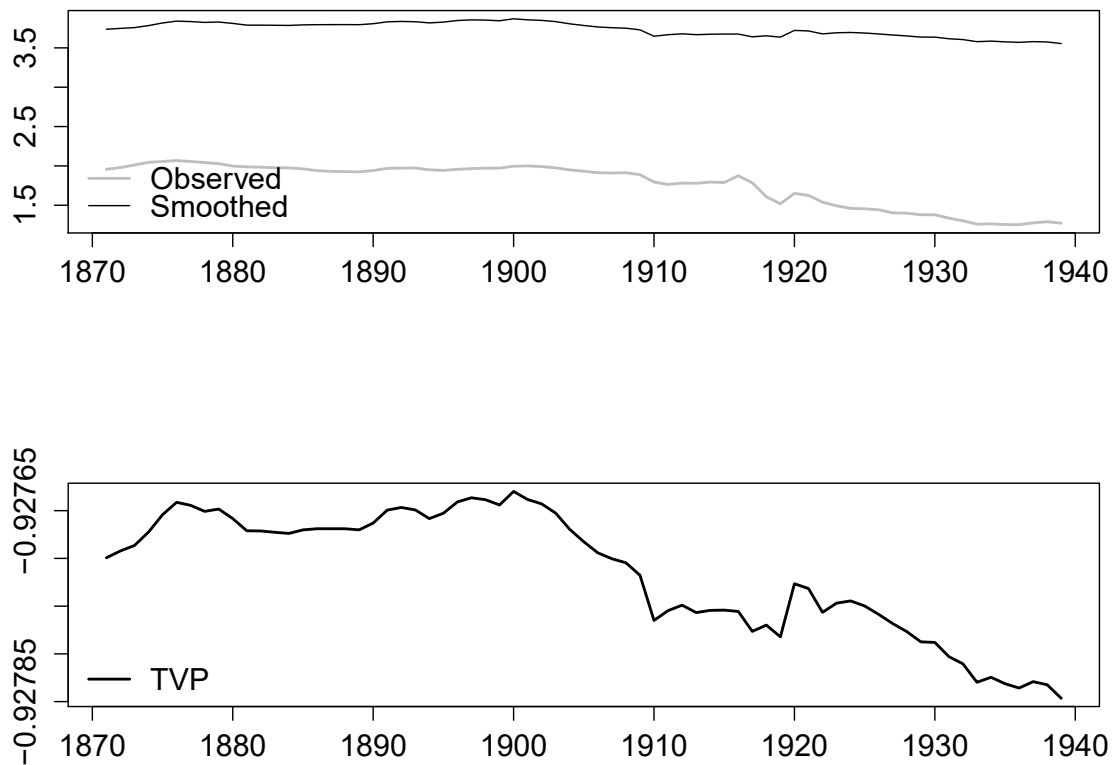


Figure 7: **Local level trend model with explanatory variable (1870-1938)**

The output for the coefficient that is associated with the variable is provided in the second panel and may be interpreted in a similar way to that of an ordinary regression coefficient. Hence, the average for the β coefficient, which is associated with price sensitivity (and has a value of around -0.9277), indicates that a one percent increase in price leads to a fall in spirit consumption of -0.9277 (on average). It is worth noting that as the value of this coefficient increases (in absolute terms) over time, it would suggest that consumers have become slightly more price sensitive.

To see what is happening to the level (which may be used to describe tastes or preferences), we note that there has been a gradual decline over time. Hence, it could suggest that tastes would be moved away from hard spirits.

As is the case with all the other models, one could then test the residuals to ensure that they represent white noise.

2.6 Confidence Intervals

In the state-space framework the stochastic state components are associated with estimation errors. These stochastic errors have unique variances, which allows for the construction of confidence intervals around each of the state components. These may be used to measure the uncertainty that is associated with each of the estimated state processes. Letting $W = 2$ we can produce confidence intervals using the standard formula,

$$\mu_t \pm 2\sqrt{W}$$

An example of these confidence intervals, when applied to the local level model for the South African deflator, is provided in Figure 8.

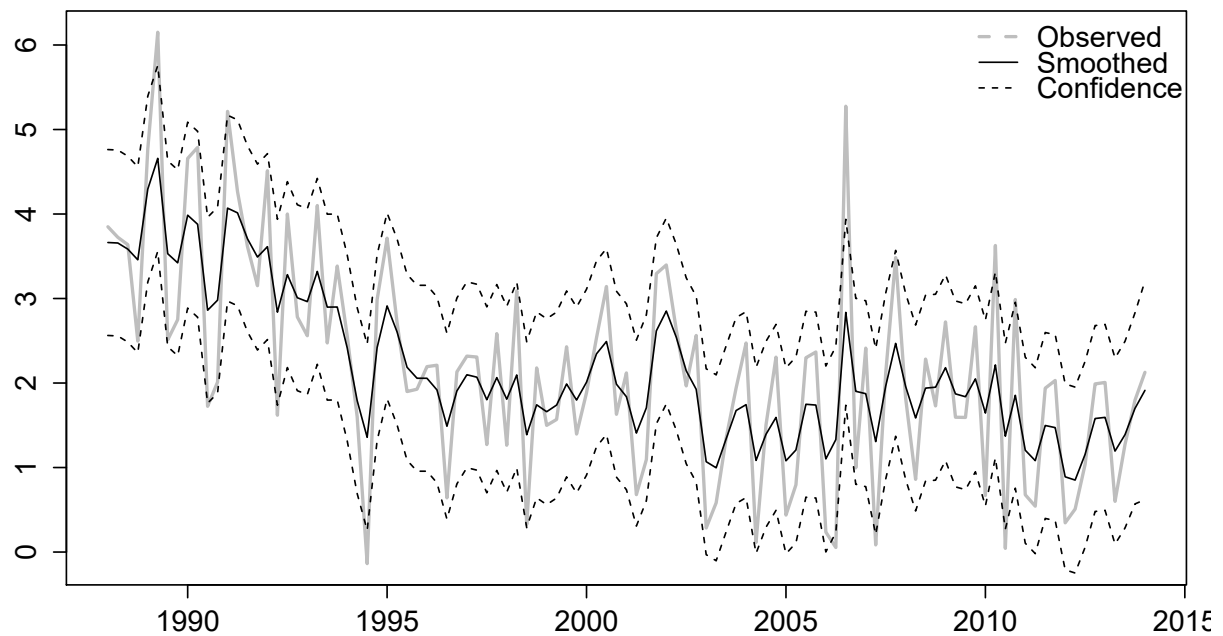


Figure 8: **Local level model with confidence intervals - SA deflator (1960Q1-2014Q1)**

3 The Kalman filter

The values for the state components can be estimated with an iterative filter, such as the one that has been proposed by Kalman (1960) and Kalman and Bucy (1961). This procedure would usually involve both *filtering*, which is the result of a forward pass of the data, and *smoothing*, which is applied to the output of the Kalman filter.¹³

3.1 Filtering

For a given model and set of observed variables, $\{y_1, \dots, y_n\}$, the Kalman filter produces successive one-step ahead predictions conditional on the past and concurrent observations. Once we have point estimates for the one-step ahead predictions for the filtered-state, we can then also calculate the variance of this filtered-state variable. The value of the one-step ahead filtered-state is ultimately dependant upon the uncertainty that is associated with the errors for both the measurement and state equations.

Given the general framework for the state-space model, as presented in equations (1.5) and (1.6), the estimated state value that is provided by the Kalman filter at point $t+1$ is denoted $\alpha_{t+1} = \mathbb{E}_t[\alpha_{t+1}|y_t]$. In addition, we could also calculate the estimated variance of this filtered-state variable, which is denoted $Q_{t+1} = \text{var}[\alpha_{t+1}|y_t]$. The general formula for the recursive Kalman filter may then be represented as follows,

$$\alpha_{t+1} = \alpha_t + K_t(y_t - F'_t \alpha_t)$$

where for the local level model this would reduce to,

$$\mu_{t+1} = \mu_t + K_t(y_t - \mu_t)$$

Given the previous expressions for the local level model, the one-step ahead *innovation errors* are derived from the measurement equation and would be expressed as, $\varepsilon_t = y_t - \mu_t$, where they are regarded as innovations, since they present new information from the observed variable. Note that the uncertainty surrounding these errors is derived from the variance of ε_t , which we denote V_ε .

Similarly, the errors for the state equation in a local level model would be expressed as, $\xi_t = \mu_{t+1} - \mu_t$. In this case the uncertainty that is associated with these errors is derived from the variance of ξ_t , which we denote W_ξ .

The variable K_t is called the Kalman gain and it refers to the simultaneous compromise between the uncertainty that relates to the two pieces of information that we have at each point in time. When the uncertainty of the state, based on past observations of the state variable, is relatively large then the variance for the error in the state equation will be large. In such cases, it would be ideal if K_t were to tend towards a value of one, as this would allow for new information that is obtained from y_t to have a large impact on the next value of the state. Similarly, if the difference between the observed variable y_t and the estimated state variable μ_t is highly volatile (as would be the case when y_t includes a number of outliers), then it would be ideal if K_t were to tend towards zero. Therefore, an appropriate statistic for the value of K_t may be derived as follows:

$$K_t = \frac{Q_t}{Q_t + V_{\varepsilon_t}}$$

where Q_t refers to the one-step ahead filtered-state estimation-error, or more commonly the *prediction error* variance, and is largely related to the variance that is due to the errors from the state equation, W_ξ . Note that future values for Q_t , or the contribution of uncertainty from the state equation to the total amount of uncertainty, will be affected by the previous value for the Kalman gain, since a large value for K_t would ensure that the observed variable, y_t , will also have more influence over the future variability of the state variable. Therefore, we make use of the expression $Q_{t+1} = (1 - K_t)Q_t + W_\xi$ to derive recursive values for Q_t .

To consider the operations of a Kalman Filter in further detail we can work with hypothetical data for the observed process, $y_t = \{6.07, 6.09, 5.89, 5.83, 6.00, 6.03\}$. Since we only have a single state equation we could use α_t and μ_t interchangeably. In what follows, we will use μ_t , however as long as we have made the correct specification of the matrices in the general representation of the model, the following explanation would also apply to those cases where we make use of several state equations.

To initialise the filter we need to provide starting values. In this case, we assume that the starting value for μ_0 is equal to the mean of \bar{y}_t . In addition, we also assume that the initial value for Q_t is 2, while the values for W_ξ and V_ε will equal 1. In almost all practical cases, the variance for the errors would need to be estimated and the only reason for imposing this simplification is to provide clarity to the exposition. For further details on the derivation of appropriate starting values, see De Jong (1991) and Chapter 2 of Durbin and Koopman (2012). In this particular case, the initial value for the Kalman gain is, $K_0 = 2/3$ and to work out the value for μ_{t+1} with the aid of the Kalman filter we have:

$$\begin{aligned}\mu_{t+1} &= \mu_t + K_t(y_t - \mu_t) \\ &= 5.985 + 2/3(6.07 - 5.985) \\ &= 6.041667\end{aligned}$$

We could then work out the subsequent value for the Kalman gain, after calculating $Q_{t+1} = (1 - K_t)Q_t + W_\xi = 1.6$,

$$K_{t+1} = \frac{Q_{t+1}}{Q_{t+1} + V_{\varepsilon_t}} = 0.625$$

This allows for the derivation of the future value of the state-vector

$$\begin{aligned}\mu_{t+2} &= \mu_{t+1} + K_{t+1}(y_{t+1} - \mu_{t+1}) \\ &= 6.041667 + 0.625(6.09 - 6.041667) \\ &= 6.071875\end{aligned}$$

And so the procedure continues until we have all the *filtered* values for state equation. The results of this hypothetical exercise have been displayed in Figure 9. Note that since the update of the filtered state at point $t + 1$ is based on values from time point t , it always projects forward by one observation and in this case it would appear as if the *filtered* value in period $t + 1$ seems to provide a more accurate description of state of the system in period t .

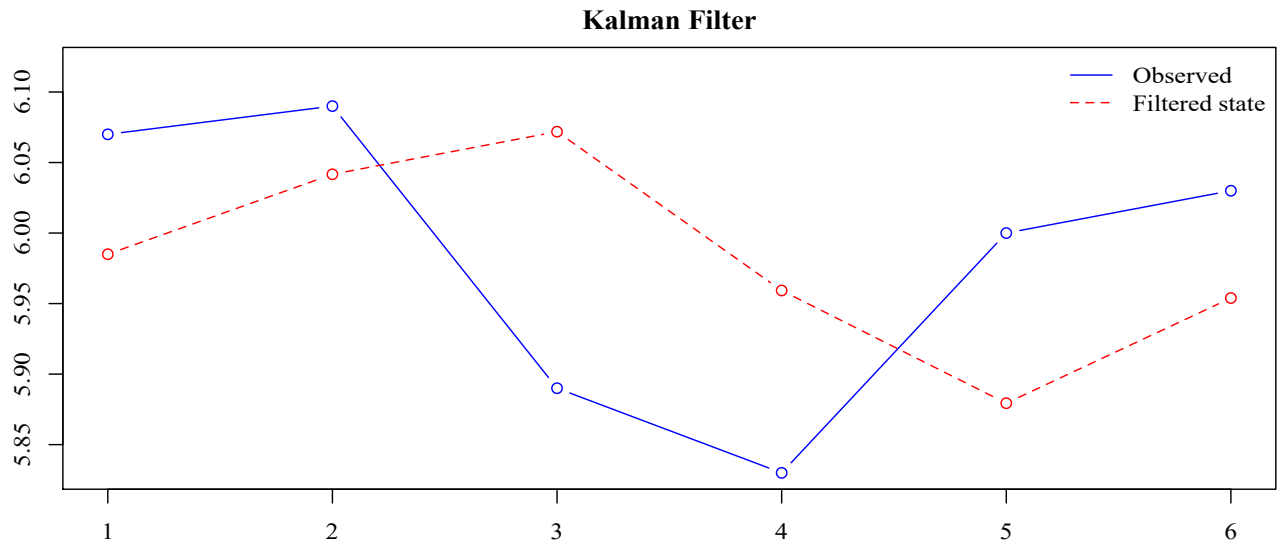


Figure 9: **Kalman filter and observed variable**

Over time the values of Q_t and $V_{\varepsilon,t}$ would usually converge towards a constant value, even when they are estimated, which implies that K_t would also converge on a constant value. An example of the values that may be produced for the Kalman gain and prediction errors are provide in Figure 10.

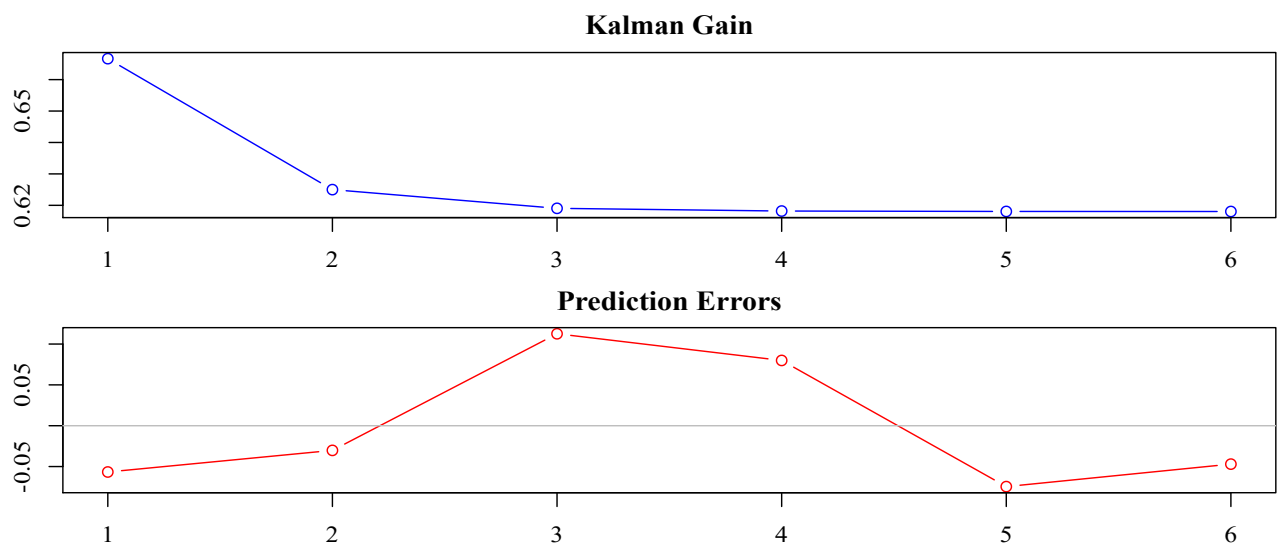


Figure 10: **Kalman gain and prediction errors**

3.2 Smoothing

The Kalman filter is a recursive algorithm that evaluates the respective one-step ahead estimates, where the values of α_{t+1} are largely related to the observation from a previous period, y_t . The general idea behind the smoothing algorithm is to ensure that projected values of the state are more closely matched with observation that was realised during the same period of time (i.e. y_t). Many potential algorithms exist to achieve this objective, and the interested reader is referred to Kitagawa and Gersch (1996) for an extensive treatment of these techniques.

One example of a smoother that may be used in the hypothetical example that we have discussed, would be to start at the last observation of the time series, n and move towards to the first observation with the aid of a similar recursive procedure. Such a smoothing algorithm could be specified, in a similar way to that of the original Kalman filter, where for the local level model, we could specify a smoother as follows:

$$\alpha_t^s = \alpha_t + J_t(\alpha_{t+1}^s - \alpha_t)$$

where α_t^s is the smoothed estimate and α_t is the filtered estimate. The value for J_t would then be determined by $(1 - 1/Q_t)$ which is equal to K_t/Q_t . In the above example, we would use the Kalman filter to generate a value for $\mu_{t+6}^s = 6.0009$, which would serve as the starting value for the smoothing algorithm. Given that we have already calculate all the values for Q_t during the forward pass, we could calculate all the values for J , where in this case $J_{t+5} = 0.38197$. This would imply that:

$$\begin{aligned}\mu_{t+5}^s &= \mu_{t+5} + J_{t+5}(\mu_{t+6}^s - \mu_{t+5}) \\ &= 5.9539 + 0.38197(6.0009 - 5.9539) \\ &= 5.97188\end{aligned}$$

Where we can then generate values for μ_{t+4}^s with the use of μ_{t+5}^s , until we have all the smoothed values for this process. This would provide the result that is provided in Figure 11. Note how the smoothed values for the state now appear to be time consistent with the observed values for the time series.

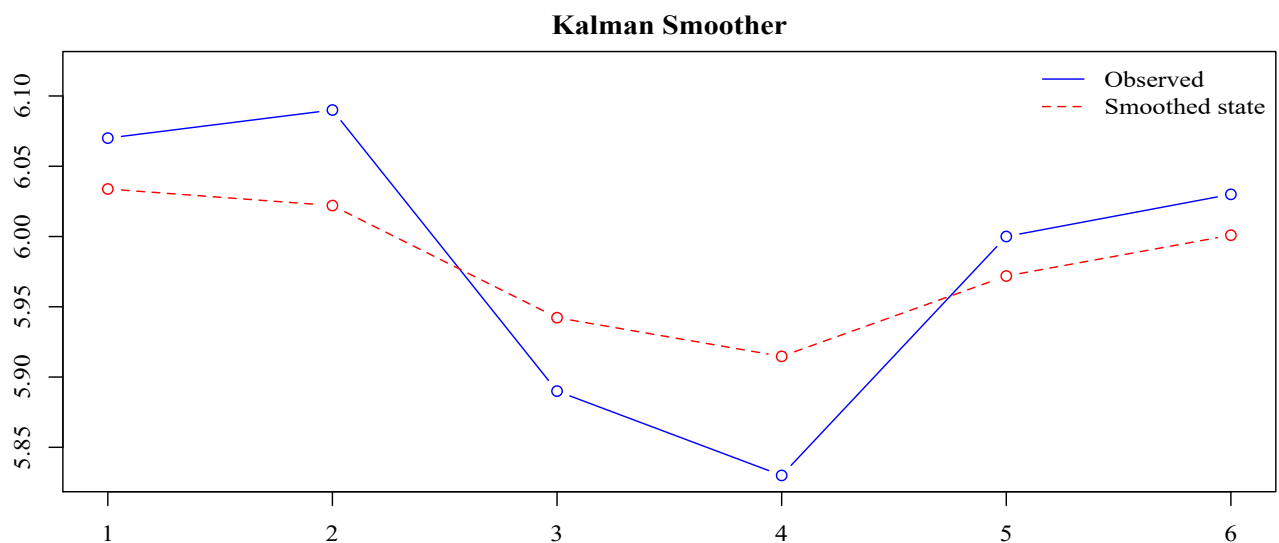


Figure 11: Kalman smoother and observed variable

3.3 Example - Kalman filter in local level model

In the following example, we seek to estimate a state-space model for a simulated random-walk process that has 50 observations, where we make use of the assumption that the errors are distributed $\xi_t = \text{i.i.d.}\mathcal{N}[0, 1]$ and $\varepsilon_t = \text{i.i.d.}\mathcal{N}[0, 1]$. In addition, we make use of starting values, $\alpha_0 = \mathcal{N}[0, 1]$, for the state variable. After generating values for the filtered and smoothed state variable, we are then able to inspect the values for the Kalman gain and the prediction errors. These are provided in Figure 11. Note that the Kalman gain quickly converges to a constant value. In addition, we also note that the predicted errors that have been made would appear to display behaviour that is consistent with a white noise process. These values are subjected to further tests in the following subsection.

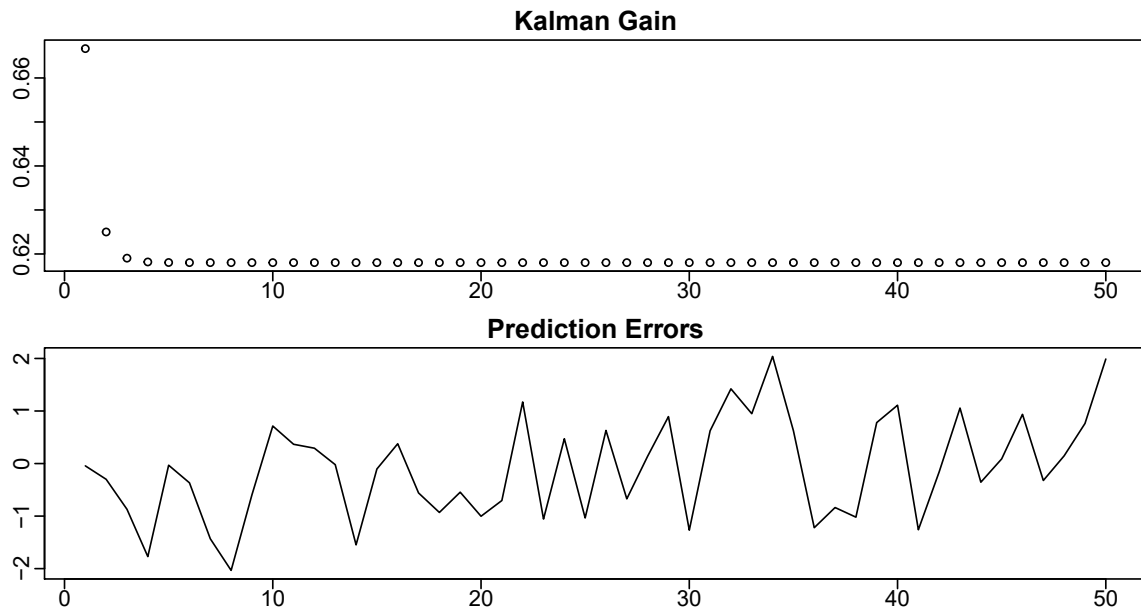


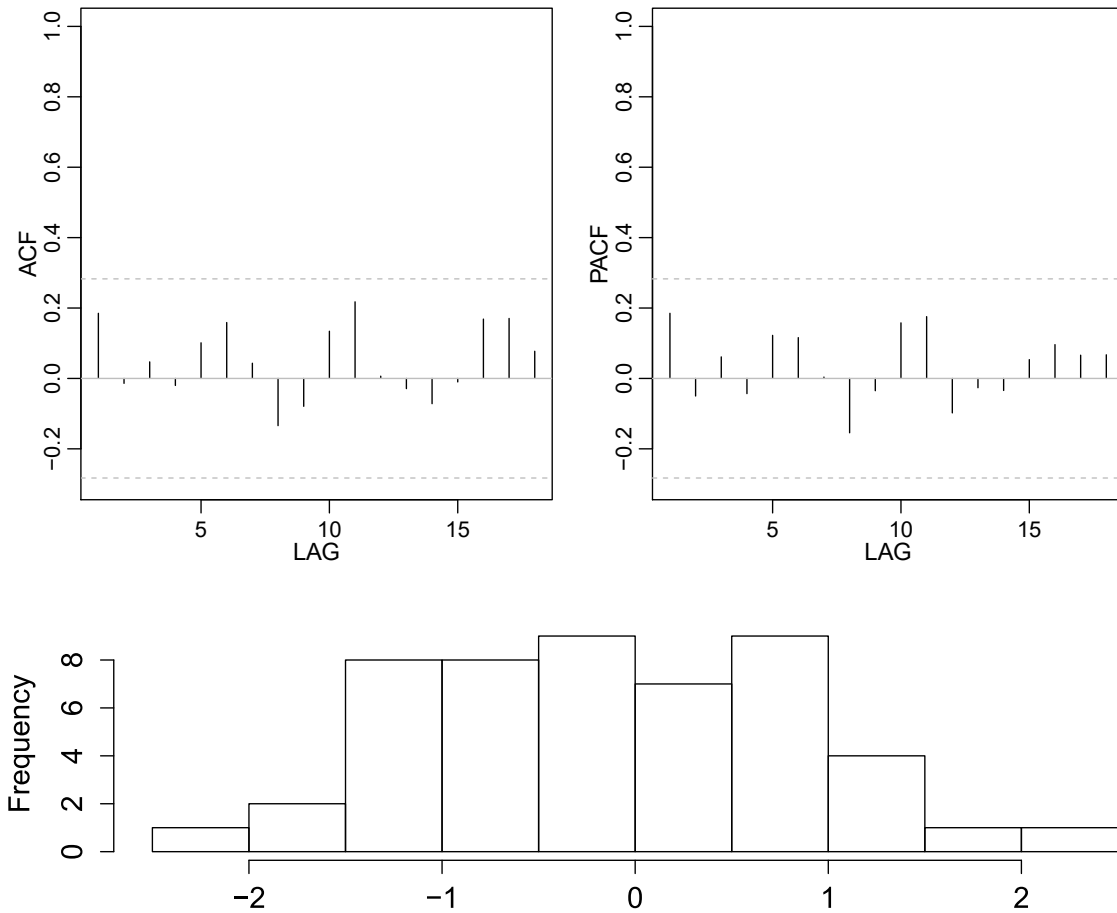
Figure 12: **Kalman gain and one-step ahead prediction errors**

3.4 Diagnostic tests

Residuals should be independent, homoskedastic, and normally distributed. To investigate whether they satisfy these properties we consider the behaviour of the standardised prediction errors, which are defined as;

$$e_t = \frac{\xi_t}{\sqrt{Q_t}}$$

These standardised prediction errors are displayed in the attaching graph (13), which seeks to describe the inflationary gap (i.e. the difference between the core deflator and actual inflationary pressure) with a local level trend model with seasonal and intervention components, and an explanatory variable (output gap).

Figure 13: **Diagnostic tests on prediction errors**

To test for **independence** we make use of the Box-Ljung statistic, where the residual autocorrelation from lag k ,

$$r_k = \frac{\sum_{t=1}^{T-k} (e_t - \bar{e})(e_{t+k} - \bar{e})}{\sum_{t=1}^T (e_t - \bar{e})^2}$$

and \bar{e} is the mean of the n residuals. The Box-Ljung statistic may then be expressed as,

$$Q(k) = T(T+2) \sum_{l=1}^k \frac{r_l^2}{T-l}$$

for lags $l = 1, \dots, k$. This value is then compared to a χ^2 distribution with $(k - w + 1)$ degrees of freedom (where w is the number of hyperparameters or disturbance variances). If the calculated value is less than the critical value at some level of significance (eg. 5%), $Q(k) < \chi_{(k-w+1;0.05)}^2$, it implies that the null of independence is not rejected and there is no reason to assume that the residuals are serially correlated.

The second assumption of **homoskedasticity** of the residuals may be tested by comparing the variance of the residuals in the first third of the series with the variance of the residuals of the last third of the series. Hence, the following statistic,

$$H(h) = \frac{\sum_{t=T-h+1}^T e_t^2}{\sum_{t=d+1}^{d+h} e_t^2}$$

where d is the number of diffuse initial state values (there is usually one element for each state equation) and h is the nearest integer to $(T - d)/3$. This value is then compared to an F -distribution with (h, h) degrees of freedom. When using a 5% level of significance the critical values for a two-tailed test correspond to the upper and lower 2.5% of the F -distribution. When $H(h) > 1$ then we test $H(h) < F(h, h; 0.025)$, when $H(h) < 1$ then we test $1/H(h) < F(h, h; 0.025)$. If $1 < H(h) < F(h, h; 0.025)$, the null of equal variances is not rejected and there is no reason to assume a departure from homoskedasticity of the residuals.

Normality of the residuals can be tested with the following statistic which considers the skewness and kurtosis of the residual distribution,

$$N = T \left(\frac{S^2}{6} + \frac{(K - 3)^2}{24} \right)$$

with the skewness, S , and the kurtosis, K , being defined as,

$$S = \frac{\frac{1}{T} \sum_{t=1}^T (e_t - \bar{e})^3}{\sqrt{\left(\frac{1}{T} \sum_{t=1}^T (e_t - \bar{e})^2 \right)^3}}$$

$$K = \frac{\frac{1}{T} \sum_{t=1}^T (e_t - \bar{e})^4}{\left(\frac{1}{T} \sum_{t=1}^T (e_t - \bar{e})^2 \right)^2}$$

If $N < \chi_{(2;0.05)}^2$ the null hypothesis of normality is not rejected, and there is no reason to assume that the residuals are not normally distributed.

The further diagnostic tool considers an inspection of the auxiliary residuals, which are obtained by dividing the smoothed disturbances with the square root of their corresponding variances,

$$\frac{\hat{\varepsilon}_t}{\sqrt{\text{var}(\hat{\varepsilon}_t)}} \text{ and } \frac{\hat{\eta}_t}{\sqrt{\text{var}(\hat{\eta}_t)}}$$

This results in the standardised smoothed disturbances. Inspection of the standardised smoothed observation disturbances allows for the detection of outliers whilst the detection of the standardised smoothed state disturbances allows for the detection of structural shifts.

If an outlier is detected then check for measurement error and where necessary include a pulse intervention (dummy) variable. For structural breaks in the level include a level shift intervention variable. The inclusion of such variables should always be based on some underlying theory concerning the cause of the structural break or outlier.

3.5 Forecasting

To compute forecasts of a time series we simply continue with the Kalman filter. When we arrive at the end of the sample, the update of the filter state at time point $t = n$ equals,

$$\alpha_n = \alpha_{n-1} + K_{n-1}(y_{n-1} - F'_{n-1}\alpha_{n-1})$$

The last observation, y_n , can then be used to update the filtered state at time point $t = n + 1$ as follows,

$$\alpha_{n+1} = \alpha_{n-1} + K_{n-1}(y_n - F'_n\alpha_n)$$

From $n + 1$ onwards the filtered state no longer changes and by letting $\bar{\alpha}_{n+1} = \alpha_{n+1}$ the forecasts simply become $\bar{\alpha}_{n+1+j} = \bar{\alpha}_{n+f}$, where F refers to the number of time points for the forecast (i.e. the lead time or forecasting horizon). Such forecasts are useful since they provide information on future developments based on the past, and in addition, they may also be used for out-of-sample testing to determine whether the series behaves according to our expectation during future periods. For such exercises we are often less conservative with confidence limits which may be decreased to 90% (where, $\alpha_t \pm 0.64\sqrt{Q_t}$) or even 85%. Such an exercise is presented in Figure 14, where we can see that the confidence interval quickly increases with time.

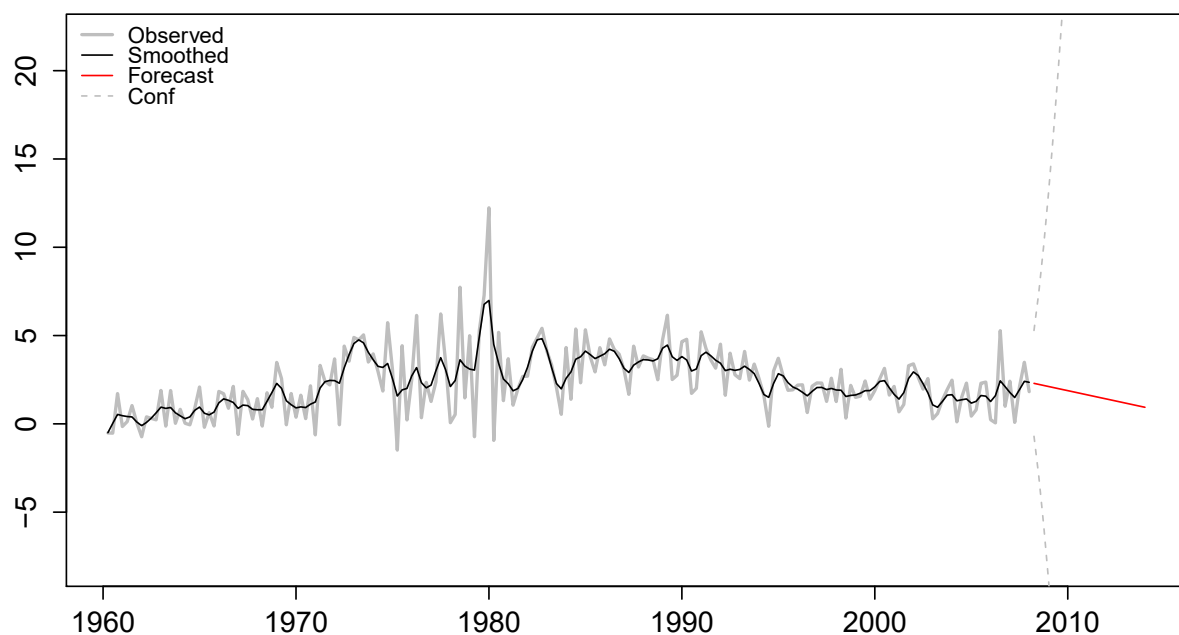


Figure 14: **Filtered forecasts - SA deflator (1960Q1-2014Q1)**

3.6 Missing observations

Missing observations are easily dealt with and are treated as if they were forecasts. In estimating the filtered state the values of the prediction errors and the Kalman gain K_t are set to zero whenever the value of an observation is missing.

4 State Space and Box-Jenkins methods for Time Series Analysis

This section considers the relative merits of state-space and Box-Jenkins methods. It starts off with a brief recap of central elements of the Box-Jenkins approach.

4.1 Stationary processes and related concepts

A stochastic process, μ_t , is weakly stationary if its mean, variance and autocovariance are constant. In addition, a stochastic process is called a *purely random* process if it consists of random variables, η_t , that are mutually independent and identically distributed. Such a process is also stationary since it has constant mean and variance, and all the autocorrelations are zero.

In contrast, a process is a *random walk* if,

$$\mu_{t+1} = \mu_1 + \sum_{j=1}^t \eta_j = \mu_t + \eta_t$$

It follows that the first difference of a random walk yields a stationary random process.

$$\Delta\mu_{t+1} = \mu_t - \mu_{t-1} = \eta_{t-1}$$

Where η_t is a purely random process with mean zero and variance σ^2 . Then a process, μ_t is called a moving average process of order q , [MA(q)], if

$$\mu_t = \beta_0\eta_t + \beta_1\eta_{t-1} + \beta_2\eta_{t-2} + \dots + \beta_q\eta_{t-q}$$

For example a first order MA(1) process may be given as,

$$\mu_t = \eta_t + 0.5\eta_{t-1}$$

A pure MA(q) process is always a stationary process. For example, where $\beta_0 = 1$

$$\frac{\beta_1}{1 + \beta_1^2} = \frac{0.5}{1 + 0.5^2} = 0.4$$

When η_t is a purely random process with mean zero and variance σ^2 . Then the process, μ_t is called a moving average process of order p , [AR(p)], if

$$\mu_t = \alpha_1\mu_{t-1} + \alpha_2\mu_{t-2} + \dots + \alpha_p\mu_{t-p} + \eta_t$$

In this case, μ_t is regressed on past values of itself. An example of an AR(1) process would be,

$$\mu_t = 0.5\mu_{t-1} + \eta_t$$

Such an AR(1) process is stationary since the coefficient is within the unit circle.

The combination of autoregressive and moving average processes produces a autoregressive moving average model, ARMA(p, q), with p AR terms and q MA terms. It may be expressed as,

$$\mu_t = \alpha_1\mu_{t-1} + \alpha_2\mu_{t-2} + \dots + \alpha_p\mu_{t-p} + \eta_t + \beta_1\eta_{t-1} + \beta_2\eta_{t-2} + \dots + \beta_q\eta_{t-q}$$

An example of such a model could take the following form, which is stationary,

$$\mu_t = 0.5\mu_{t-1} + \eta_t + 0.5\eta_{t-1}$$

4.2 Non-stationary ARIMA models

A typical Box-Jenkins approach to time series proceeds as follows. For a time series that includes some non-stationary features (due to trend and/or seasonal effects), the observed time series is transformed into a stationary series using time and lag functions (often through differencing). This may involve removing the trend by taking the first difference of an observed series, y_t , to create a new series y_t^* ,

$$y_t^* = \Delta y_t = y_t - y_{t-1}$$

Alternatively the seasonal with periodicity s may be removed by differencing,

$$y_t^* = \Delta_s y_t = y_t - y_{t-s}$$

Or in certain cases it may be necessary to remove both the trend and the seasonal, which would involve,

$$y_t^* = \Delta \Delta_s y_t = (y_t - y_{t-s}) - (y_{t-1} - y_{t-s-1})$$

In cases where the variable is still not stationary, the differencing procedure can be continued by taking the second difference,

$$y_t^* = \Delta^2 \Delta_s^2 y_t,$$

After sufficient differencing the appropriate AR(p), MA(q) or ARMA(p,q) the model that can best account for the differenced time series needs to be identified, where the residuals of the best model should follow a random process.

4.3 Unobserved components and ARIMA

To consider the similarities between state-space and ARIMA models, recall that the local level model has the form,

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t \\ \mu_t &= \mu_{t-1} + \eta_t \end{aligned} \tag{4.1}$$

Where the first difference of y_t yields,

$$\Delta y_t = y_t - y_{t-1} = \mu_t - \mu_{t-1} + \varepsilon_t - \varepsilon_{t-1} \tag{4.2}$$

Since (4.1) implies that,

$$\mu_t - \mu_{t-1} = \eta_t \tag{4.3}$$

We can rewrite (4.2) as,

$$\Delta y_t = y_t - y_{t-1} = \eta_{t-1} + \varepsilon_t - \varepsilon_{t-1}$$

Which is stationary and has the same correlogram as an MA(1) process. This implies that the local level model is an ARIMA(0,1,1). Similarly we can also show that a local level trend model can be represented as a ARIMA(0,2,2). A comprehensive overview of the equivalencies between state-space and ARIMA models is provided in Harvey (1989). Finally it should be noted that ARIMA models could also be put into a state-space form and fitted by state-space methods.

4.4 State-space versus ARIMA approaches

Despite the similarities between the two approaches they differ in a number of important respects. The most important of these is that the state-space approach seeks to explicitly model the non-stationarity in terms of the trend and the seasonal components, while the Box-Jenkins approach treats these as nuisance components that need to be removed prior to any analysis. Hence the state-space approach seeks to provide an explicit structural framework for the simultaneous decomposition of a time series into the respective dynamic components, while the Box-Jenkins approach is primarily concerned with short-run dynamics and short-run forecasts.

Furthermore, since we are never really sure as to whether a series is stationary or non-stationary (and this is particularly so for most economic and financial time series) there may be problems with the application of the Box-Jenkins approach. When making use of state-space methods, stationarity is not required and it is often easier to deal with missing data, time-varying regression coefficients and multivariate extensions when adopting this approach.

5 Conclusion

Models with unobserved components are frequently encountered in economics and finance. These processes may be modelled with the aid of a state-space representation that provides a parsimonious way of modelling dynamic systems. These models avoid the need to use *ad hoc* proxies for the unobservable variables, which can result in biased and inconsistent parameter estimates. Researchers are able to employ the Kalman filter for the identification and extraction of the unobserved components in the model. This iterative technique has been used in many examples, in those cases where the state variables exhibit linear properties.

Then finally, it is worth noting that Bayesian techniques are frequently used to estimate the parameters in these models as one is then able to assume that all the parameters and unobserved components are modelled as random variables. In addition, these estimation techniques also lend themselves to instances where a nonlinear filter or smoother may need to be employed.¹⁴

6 References

- Anderson, B. D. O., and J. B. Moore. 1979. *Optimal Filtering*. New York: Englewood Cliffs: Prentice-Hall.
- Carter, C. K., and R. Kohn. 1994. "On Gibbs Sampling for State Space Models." *Biometrika* 81: 541–53.
- Chow, G. C. 1960. "Tests of Equality Between Sets of Coefficients in Two Linear Regressions." *Econometrica* 28: 591–605.
- Commandeur, Jacques J. F., and Siem Jan Koopman. 2007. *An Introduction to State Space Time Series Analysis*. Oxford: Oxford University Press.
- De Jong, P. 1991. "The Diffuse Kalman Filter." *Annals of Statistics* 19: 1073–83.
- Durbin, James, and Siem Jan Koopman. 2001. *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.
- . 2012. *Time Series Analysis by State Space Methods*. Second. Oxford: Oxford University Press.

- Durbin, J., and G. Watson. 1951. "Testing for Serial Correlation in Least Squares Regression Ii." *Biometrika* 38: 159–78.
- Hamilton, James D. 1994. *Time Series Analysis*. Princeton: Princeton University Press.
- Harvey, Andrew C. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.
- . 1993. *Time Series Models*. Cambridge, Mass: MIT Press.
- Kalman, R. E. 1960. "A New Approach to Linear Filtering and Prediction Problems." *Journal of Basic Engineering, Transactions ASMA, Series D* 82: 35–45.
- Kalman, R. E., and R. S. Bucy. 1961. "New Results in Linear Filtering and Prediction Theory." *Journal of Basic Engineering, Transactions ASMA, Series D* 83: 95–108.
- Kim, Chang-Jin, and Charles R. Nelson. 1998. *State-Space Models with Regime-Switching: Classical and Gibbs-Sampling Approaches with Applications*. Cambridge, Mass: MIT Press.
- Kitagawa, Genshiro, and Will Gersch. 1996. *Smoothness Priors Analysis of Time Series*. Vol. 116. Springer Verlag.
- Petris, Giovanni, Sonia Petrone, and Patrizia Campagnoli. 2009. *Dynamic Linear Models with R*. Edited by Robert Gentleman, Kurt Hornik, and Giovanni Parmigiani. New York: Springer.
- Pole, P. J., and M. West. 1988. *Nonnormal and Nonlinear Dynamic Bayesian Modeling*. Edited by J. C. Spall. Vol. Bayesian Analysis of Time Series and Dynamic Linear Models. New York: Marcel Dekker.
- Prado, Raquel, and Mike West. 2010. *Time Series - Modeling, Computation, and Inference*. Boca-Raton, Florida: Chapman & Hall.
- Shumway, R., and D. Stoffer. 2010. *Time Series Analysis and Its Applications*. New York: Springer-Verlag.
- Stroud, J. R., P. Muller, and N. G. Polson. 2003. "Nonlinear State-Space Models with State Dependent Variance." *Journal of the American Statistical Association* 98: 377–86.
- Tsay, R. S., and R. Chen. 2018. *Nonlinear Time Series Analysis*. Hoboken (NJ): John Wiley & Sons.
- West, M., and J. Harrison. 1997. *Bayesian Forecasting and Dynamic Models*. Edited by Second. New York: Springer-Verlag.

-
1. The interested reader is referred to a number of excellent alternative expositions that include Durbin and Koopman (2012), Commandeur and Koopman (2007), Harvey (1989), Harvey (1993), Hamilton (1994), Kim and Nelson (1998), Shumway and Stoffer (2010), and Tsay and Chen (2018).↵
 2. In many economic applications the components combine multiplicatively, however, by working with logarithmic values we are able to reduce the multiplicative model to the form that is represented by equation (1.1).↵
 3. Of course many other processes could be used to describe the evolution of α , i.e. an AR(2), ARMA(2,1), etc.↵

4. This linear Gaussian state-space model is commonly referred to as the local level model.↵
5. See, Harvey (1989) for more on this.↵
6. In what follows, V_ε is used to describe the vector of terms for the variance of the stochastic error in the measurement equation; and W_η is used to describe the vector of terms for the variance of the respective stochastic errors in the state equations.↵
7. Although specifying the matrices in the general form would seem like a pointless (and thankless) task, it should be borne in mind that when specifying more complex models, the software that you need to use would usually require that the input takes this form.↵
8. One could test this formally with the aid of autocorrelation functions and Q -statistics.↵
9. To confirm this specification of the model, consider in period $t = 1$; $y_1 = \mu_1 + \varepsilon_1$, where $\mu_2 = \mu_1 + v_1$ and $v_2 = v_1$. Then in period $t = 2$; $y_2 = \mu_2 + \varepsilon_2 = \mu_1 + v_1 + \varepsilon_2$, where $\mu_3 = \mu_2 + v_2 = \mu_1 + 2v_1$ and $v_3 = v_1$. Then finally in $t = 3$; $y_3 = \mu_3 + \varepsilon_3 = \mu_1 + 2v_1 + \varepsilon_3$, where $\mu_4 = \mu_3 + v_3 = \mu_1 + 3v_1$ and $v_4 = v_1$.↵
10. In the above model, equations (2.1) to (2.3) refer to the construction of a matrix of dummy variables. To see how this is constructed note that we could choose to make use of the fact that $\gamma_{2,t} = \gamma_{1,t-1}$ and $\gamma_{3,t} = \gamma_{1,t-2}$. This would imply that equation (2.1) would refer to the passage of time.↵
11. In this particular case the Hessian is singular as the variance of the stochastic term in the seasonal is zero. This prevents us from reporting on the variance of the error terms and we should conclude that the seasonal should not be included.↵
12. i.e. w_t is a binary variable that takes on values of 0 or 1.↵
13. For alternative explanations see, Anderson and Moore (1979), Harvey (1989), or Durbin and Koopman (2001).↵
14. The interested reader is referred to Carter and Kohn (1994), Pole and West (1988), West and Harrison (1997), Stroud, Muller, and Polson (2003), Petris, Petrone, and Campagnoli (2009), Tsay and Chen (2018) and Prado and West (2010).↵