State Space Modelling

Kevin Kotzé

Contents

Introduction to state space modelling

- Provides an encompassing framework to time series modelling
- Particularly useful when dealing with structural & dynamic time series, as well as models for TVP
- Unobserved variables reflect the state of the system that evolve through time
- Examples:
  - Macroeconomic analysis & state of business cycle
  - Other unobserved variables incl. expectations, reservation wages, permanent income, etc.
  - Dynamic Stochastic General Equilibrium models
  - Stochastic volatility models
- State of system is determined by an unobserved vector $\{\alpha_1, \ldots, \alpha_n\}$ which is associated with observations $\{y_1, \ldots, y_n\}$
- Relationship between the α_t's and the y_t's is specified by the state space model

Basic State Space Model

- A time series $\{y_1, \ldots, y_n\}$ may be expressed in the additive form

$$y_t = \mu_t + \gamma_t + \varepsilon_t, \quad t = 1, \ldots, T$$

- where:
  - μ_t is a slowly varying component called the *trend*
  - γ_t is a periodic component called the *seasonal*
  - ε_t is an irregular component called the *error*

- To model μ_t and γ_t suppose a scalar series α_t follows a random walk

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

Basic State Space Model

- If we assume that no seasonal is present ($\gamma_t = 0$), and
- $\mu_t = \alpha_t$ (where α_t is a random walk), we can rewrite the previous expressions as

$$\begin{aligned} y_t &= \alpha_t + \varepsilon_t, \qquad \varepsilon_t \sim \text{i.i.d.}\mathcal{N}(0, V_\varepsilon) \\ \alpha_{t+1} &= \alpha_t + \eta_t, \qquad \eta \sim \text{i.i.d.}\mathcal{N}(0, W_\eta) \end{aligned}$$

- Such a model has state space characteristics as we have:
- a **measurement equation** that describes the relation between the observed variables $\{y_1, \ldots, y_n\}$ and the unobserved state variables (α_t)
- a **state equation** that reflects the dynamics of the unobserved state variables $\{\alpha_1, \ldots, \alpha_n\}$

Basic State Space Model

- Objective of state space modelling is to infer properties of α_t's from the observed $\{y_1, \ldots, y_n\}$

- To accomplish this process we can make use of a groovy technique

The Kalman Filter!

General Exposition

- The state space form provides a unified representation of
  - ARIMA models, unobserved components models
  - dynamic time series models, time varying parameters
  - non-parametric regressions, spline regressions
- So this form is probably worth remembering!

Measurement equation

- Consider m state variables that may be subject to distortions and noise (i.e. shocks)
- State variables are contained in an $m \times 1$ vector, α_t
- The N variables that are observed are defined by an $N \times 1$ vector, y_t
- Then we can define the measurement equation as

$$y_t = F_t \alpha_t + S_t \varepsilon_t$$

- where F_t is a fixed matrix of order $N \times m$
- r is the dimension of the measurement equation disturbance vector
- ε_t is a $r \times 1$ vector with zero mean & covariance matrix, V
- S_t is also a fixed matrix of order $N \times r$

State equation

- The state equation could then be described as

$$\alpha_{t+1} = G_t \alpha_t + R_t \eta_t$$

- where G_t and R_t are fixed matrices of order $m \times m$ and $m \times g$
- g refers to the dimension of the state equation disturbance vector
- η_t is a $g \times 1$ vector with zero mean and covariance matrix, W

Disturbances

- Disturbances in the measurement and state equations are assumed to be uncorrelated

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

- And they are also uncorrelated with the initial state vector α_0

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

Unknowns

- The covariance matrix of the error terms may be referred to as Ω
- The coefficient matrix may be referred to as Φ

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

- which represent the unknowns in any standard regression model

Practicalities

- There are several software packages that facilitate the estimation of State Space models
- **R** packages include `StructTS`, `sspir`, `dse`, `dlm`, `dlmodeler`, `rucm`, `MARSS`, `KFAS`, `RWinBugs`, `RStan`, etc.
- Since the unobserved variables take the form of random variables it is usually more convenient to make use of Bayesian estimation techniques

Local Level Model

- Simplest example of state space model
- Level component is allowed to vary over time
- Hence the measurement and state equations are

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

- Or alternatively

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

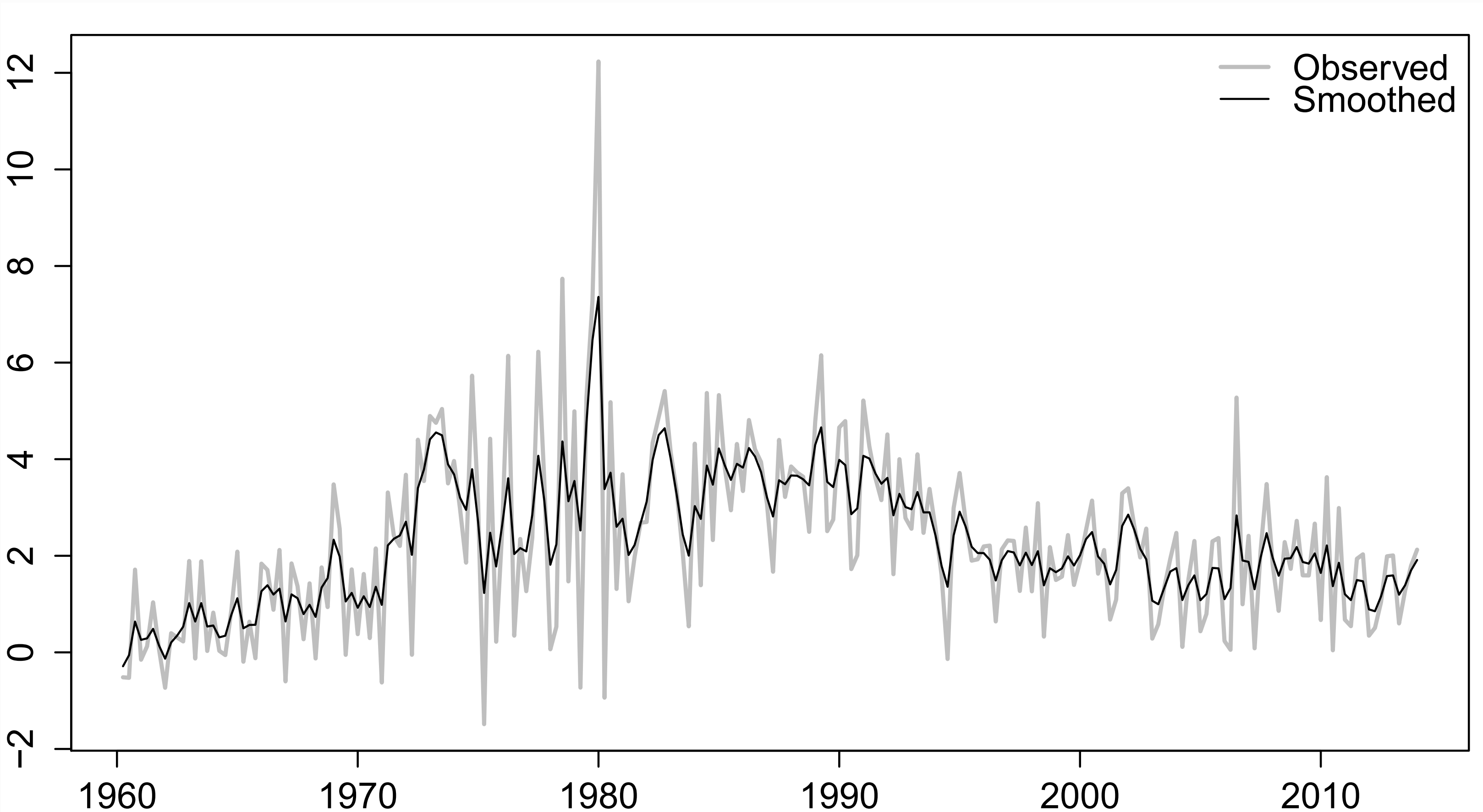

Figure : Local level model - SA deflator (1960Q2-2014Q1)

Local Level Model

- In terms of the statistical results for this model
- negative log-likelihood is 261.5
- $\hat{V}$ is 0.233
- $\hat{W}$ is 0.014
- μ, at the final state at period 2014Q1 is 1.91% per year
- To compare different state space models use Akaike Information Criterion (AIC):

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

- Compensates for the number of parameters in a model
- Smaller (or more negative) values denote better fitting models

Local Level Trend Model

- The local linear trend model has two state equations to include a slope component, υ_t

- It may be derived from the general specification by defining

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

- Or alternatively

$$
\alpha_t = \begin{pmatrix} \mu_t \\ \upsilon_t \end{pmatrix}, \; \eta_t = \begin{pmatrix} \xi_t \\ \zeta_t \end{pmatrix}, \; G_t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \; F_t = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \; S_t = 1,
$$

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

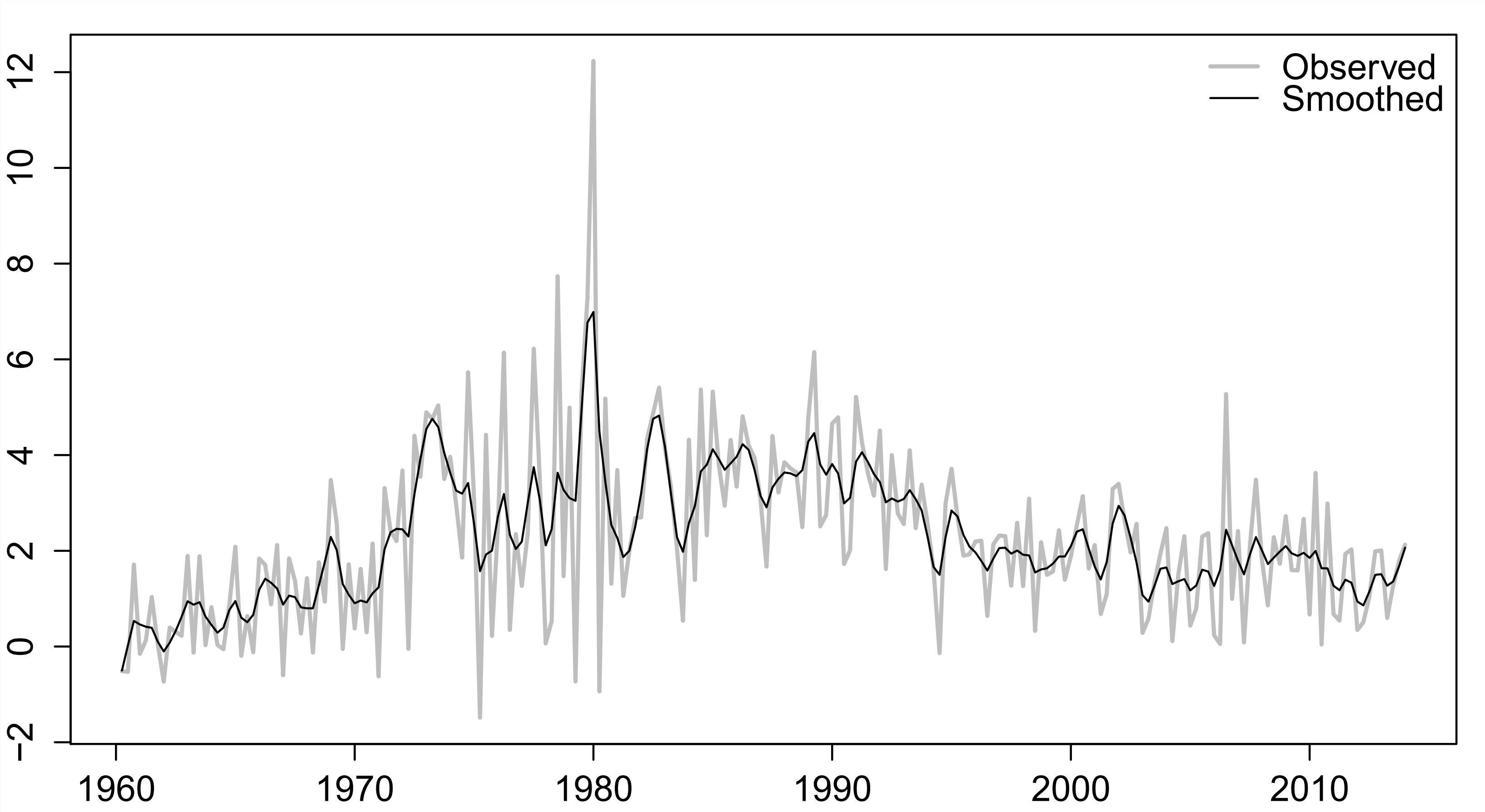

Figure : Local level model - SA deflator (1960Q2-2014Q1)

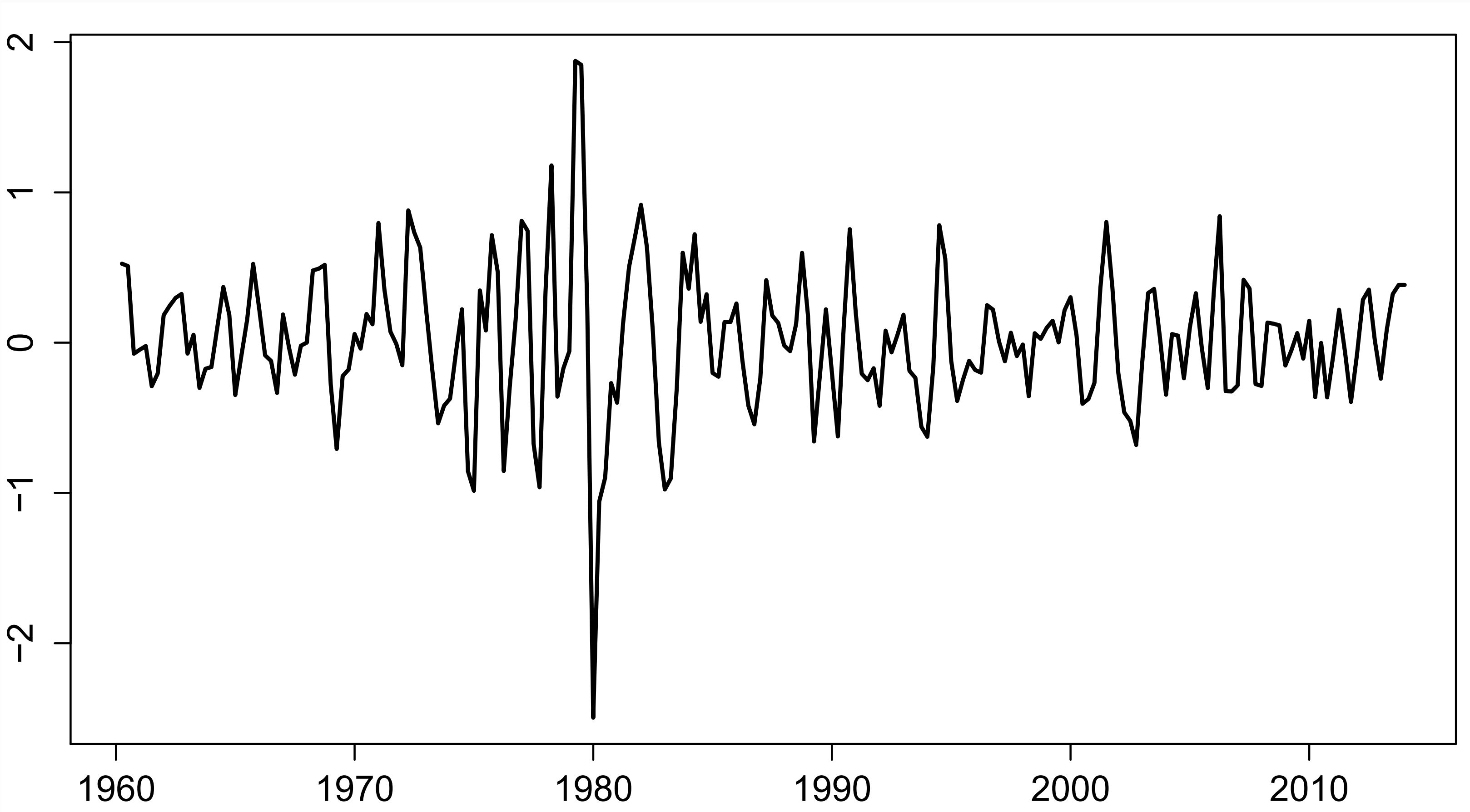

Figure : Local level trend model - Stochastic slope

Local Level Trend with Seasonal

- Time series are often influenced by seasonal characteristics incorporated in the state space framework

$$
\begin{aligned}
y_t &= \mu_t + \gamma_{1,t} + \varepsilon_t, & \varepsilon_t &\sim \text{i. i. d.}\mathcal{N}(0, V_\varepsilon) \\
\mu_{t+1} &= \mu_t + \upsilon_t + \xi_t, & \xi_t &\sim \text{i. i. d.}\mathcal{N}(0, W_\xi) \\
\upsilon_{t+1} &= \upsilon_t + \zeta_t, & \zeta_t &\sim \text{i. i. d.}\mathcal{N}(0, W_\zeta) \\
\gamma_{1,t+1} &= -\gamma_{1,t} - \gamma_{2,t} - \gamma_{3,t} + \omega_t, & \omega_t &\sim \text{i. i. d.}\mathcal{N}(0, W_\omega) \\
\gamma_{2,t+1} &= \gamma_{1,t}, \\
\gamma_{3,t+1} &= \gamma_{2,t}
\end{aligned}
$$

- where the γ terms refer to the seasonal components
- disturbance ω allows for the seasonal to change over time
- require several state equations (i.e. frequency -1)
- the last two equations show how γ_1, γ_2 and γ_3 are related over time

Local Level Trend with Seasonal

Or alternatively,

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

$$F_t = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix},\ S_t = 1,\ W = \begin{pmatrix} W_\xi & 0 & 0 \\ 0 & W_\zeta & 0 \\ 0 & 0 & W_\omega \end{pmatrix},\ R_t = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$V = V_\varepsilon$

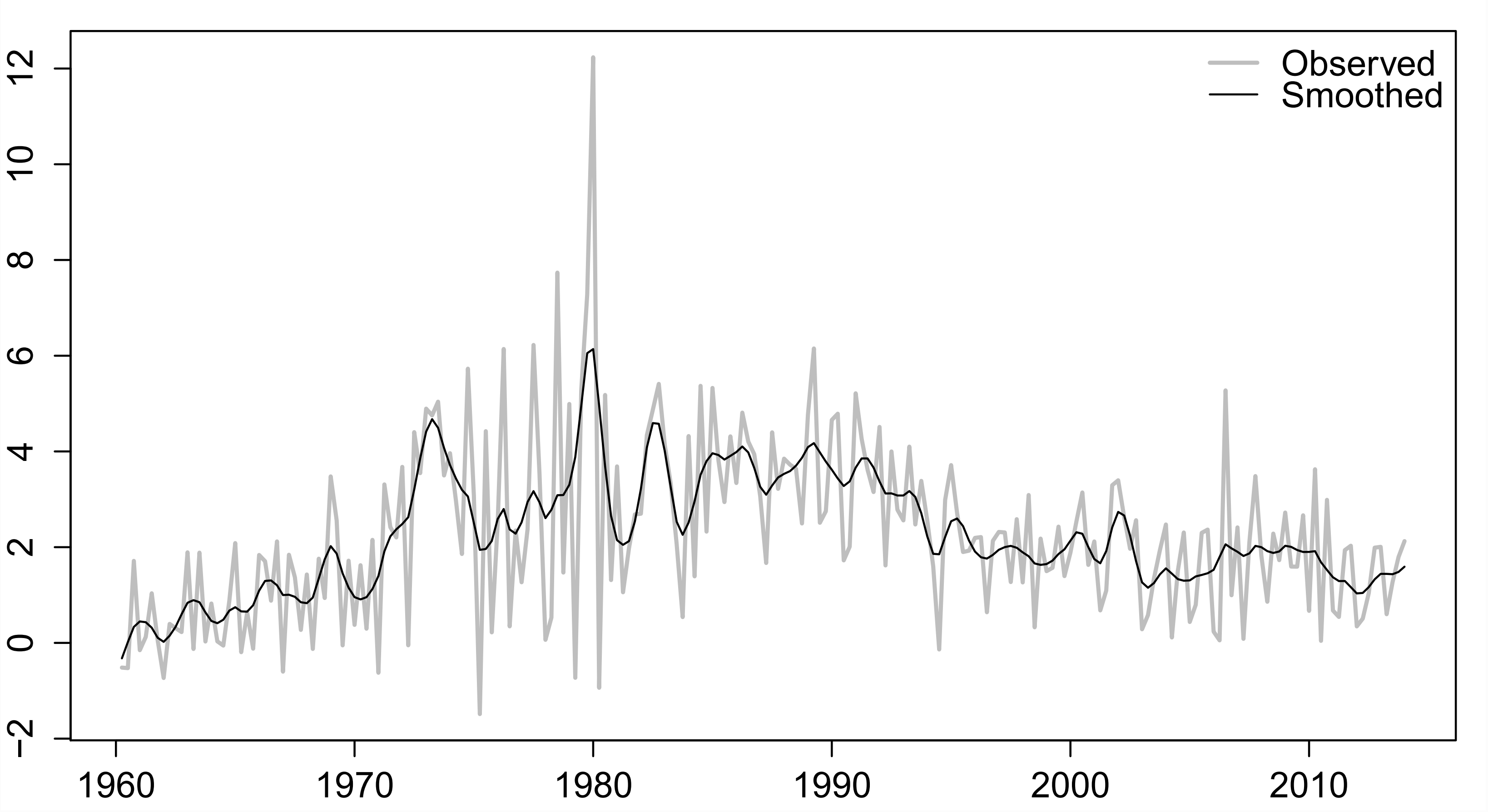

Figure : Local level model with seasonal- SA deflator (1960Q2-2014Q1)

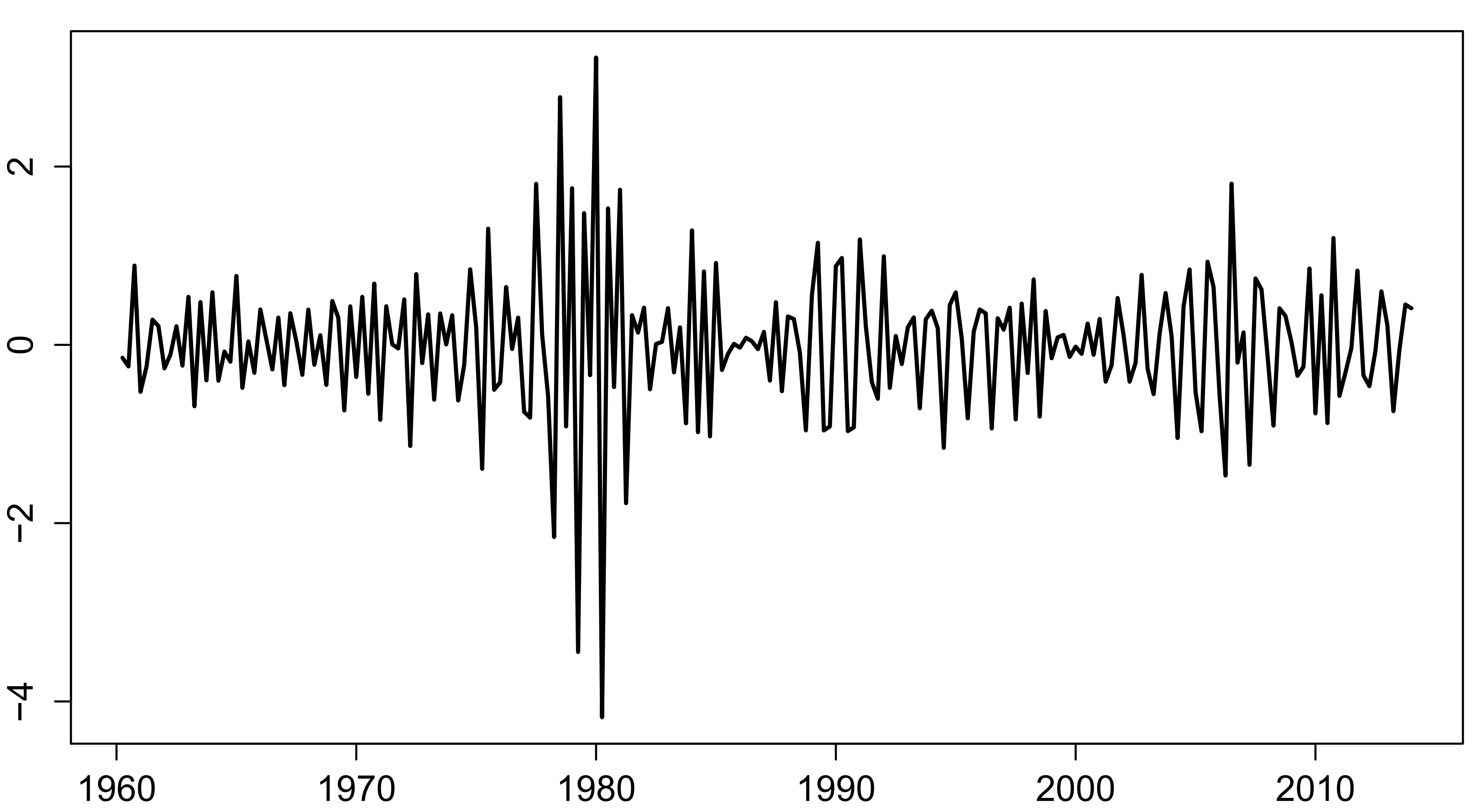

Figure : Local level trend model with seasonal - Stochastic seasonal

Local level model with intervention

- Assess the impact of a structural change on a particular time series over time
- Performed by adding intervention variables to any of the above models
- Includes *level shift*, where the value of the level of the time series exhibits a permanent change
- Or *slope shift*, where the value that is attached to the slope experiences a permanent change
- Or *pulse* effect, where the value of the level suddenly changes and then returns to previous levels

Local level model with intervention

- To determine the impact of a structural change on a local level model one could estimate:

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

Including explanatory variable

- It's easy to add the explanatory variables to the measurement equation of the model
- Just add an additional state equation for each additional explanatory variable while maintaining the unobserved components

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

Including explanatory variables

- Consider per capita consumption of spirits in the UK
- Per capita income, and the relative price of spirits from 1870 to 1938
- Allow for a stochastic level that describes changes in tastes and habits
- Allow for β parameters to be time-varying

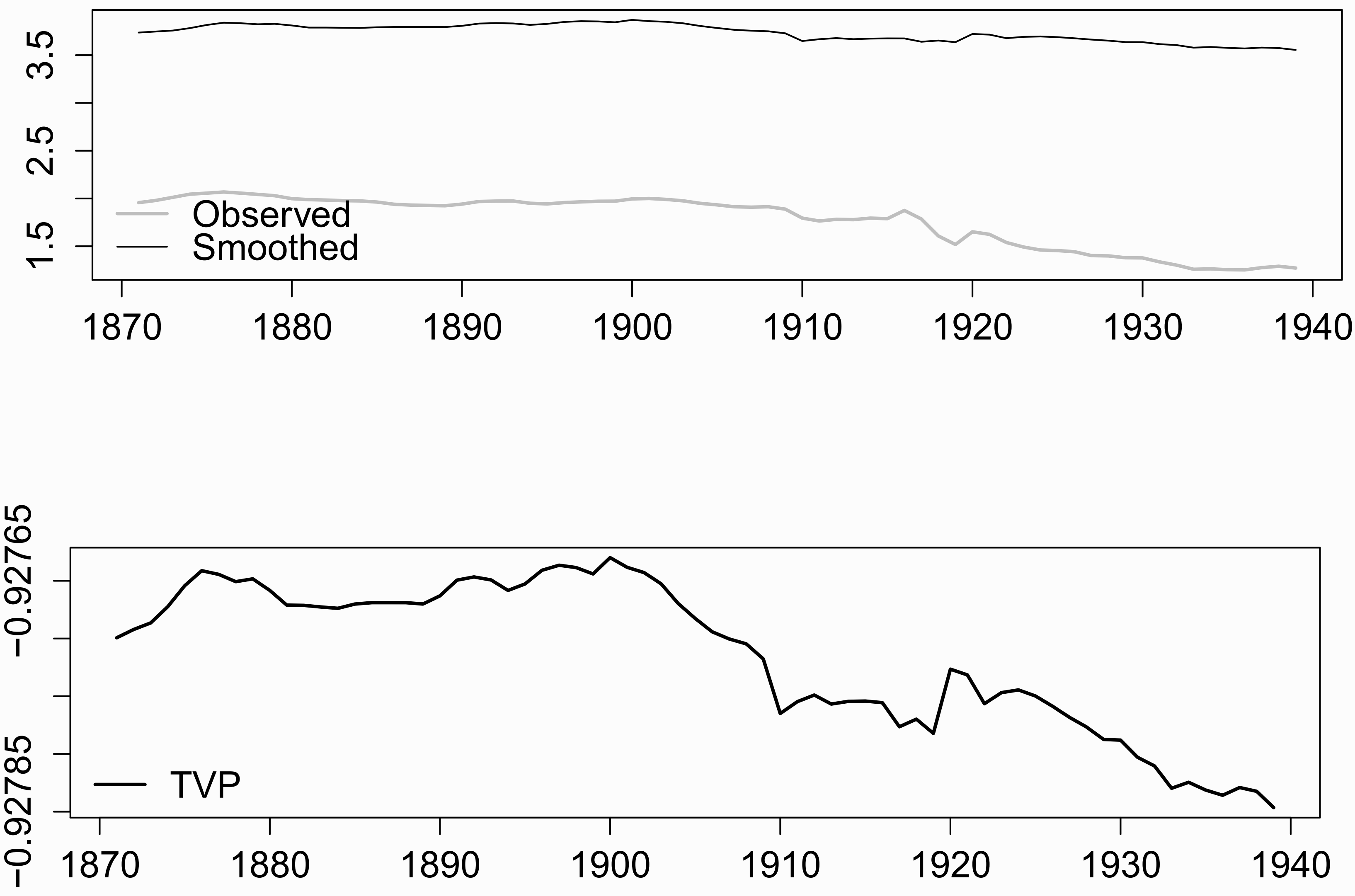

Figure : Local level model with explanatory variables

Including explanatory variables

- The output for the regressive variables are interpreted as ordinary regression coefficients
- $\bar{\beta} = -0.927$ indicates that a one percent increase in price leads to a fall in spirit consumption of -0.9 (on average)
- Decline in coefficient indicates consumers have become more price sensitive (but only by a small amount)
- The decline in the level suggests that change in taste occurred, away from the hard stuff

Confidence Intervals

- State equation variables are estimated - provides standard errors
- Allows for the construction of confidence intervals for each of the state components
- Thus allowing for the an evaluation of uncertainty at each point in time

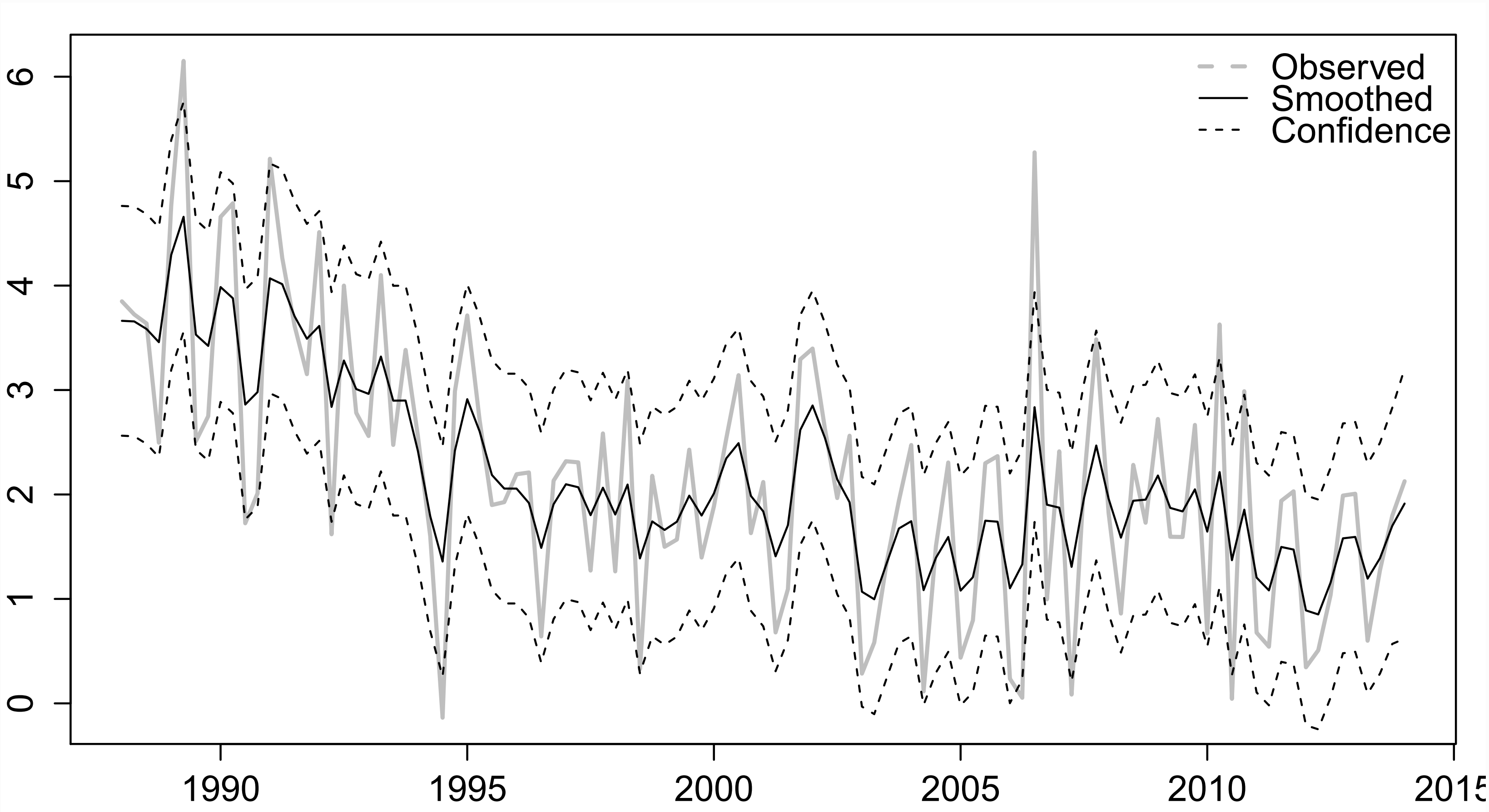

Figure : Local level trend model with confidence intervals - SA deflator (1960Q2-2014Q1)

Kalman Filter and Smoother

- The estimation of the state vectors are carried out by performing:
  - a *forward pass* - use of a recursive Kalman Filter may be used to find the filtered state
  - the *backward pass* - state and disturbance smoothers are applied to the output of the Kalman Filter

- Kalman filter provides the optimal values of that state at point t considering present and past observations:
  - provides estimates for the filtered state
  - filtered state error variances

- Disturbance smoother derives estimates that relate the current state vector to the current observation:
  - provides estimates for the smoothed state
  - provides the smoothed state estimation error variances
  - smoothed irregular components

Kalman Filter

- Given the state-space model

$$
\begin{aligned}
y_t &= F_t \alpha_t + S_t \varepsilon_t \\
\alpha_{t+1} &= G_t \alpha_t + R_t \eta_t
\end{aligned}
$$

- The expression for the Kalman filtered state is:

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

- For the local level model this given as:

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

Kalman Filter

- One-step ahead *innovation errors* are derived from the measurement equation and would be expressed as, $\varepsilon_t = y_t - mu_t$
- Variance of ε_t is denoted V_ε
- Errors for the state equation in a local level model would be expressed as, $\xi_t = \mu_{t+1} - \mu_t$
- Variance of ξ_t, which we denote W_ξ
- 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

Kalman Filter

- 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 *prediction error* variance
- Future values for Q_t will be affected by the previous value for the Kalman gain, since K_t determines the degree to which the observed variable, y_t, influences 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

Kalman Filter

- To consider the operations of a Kalman Filter we use hypothetical data for the observed process, $y_t = \{6.07, 6.09, 5.89, 5.83, 6.00, 6.03\}$
- To initialise the filter we need to provide starting values
  - assume the starting value for μ_0 is equal to the mean of $\bar{y}_t$
  - initial value for Q_t is 2 while the values for W_ξ and V_ϵ will equal 1
- The initial value for the Kalmain 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:

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

Kalman Filter

- Could then work out the subsequent value for the Kalman gain
- First calculate $Q_{t+1} = (1 - K_t)Q_t + W_\xi = 1.\dot{6}$ before

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

- And

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

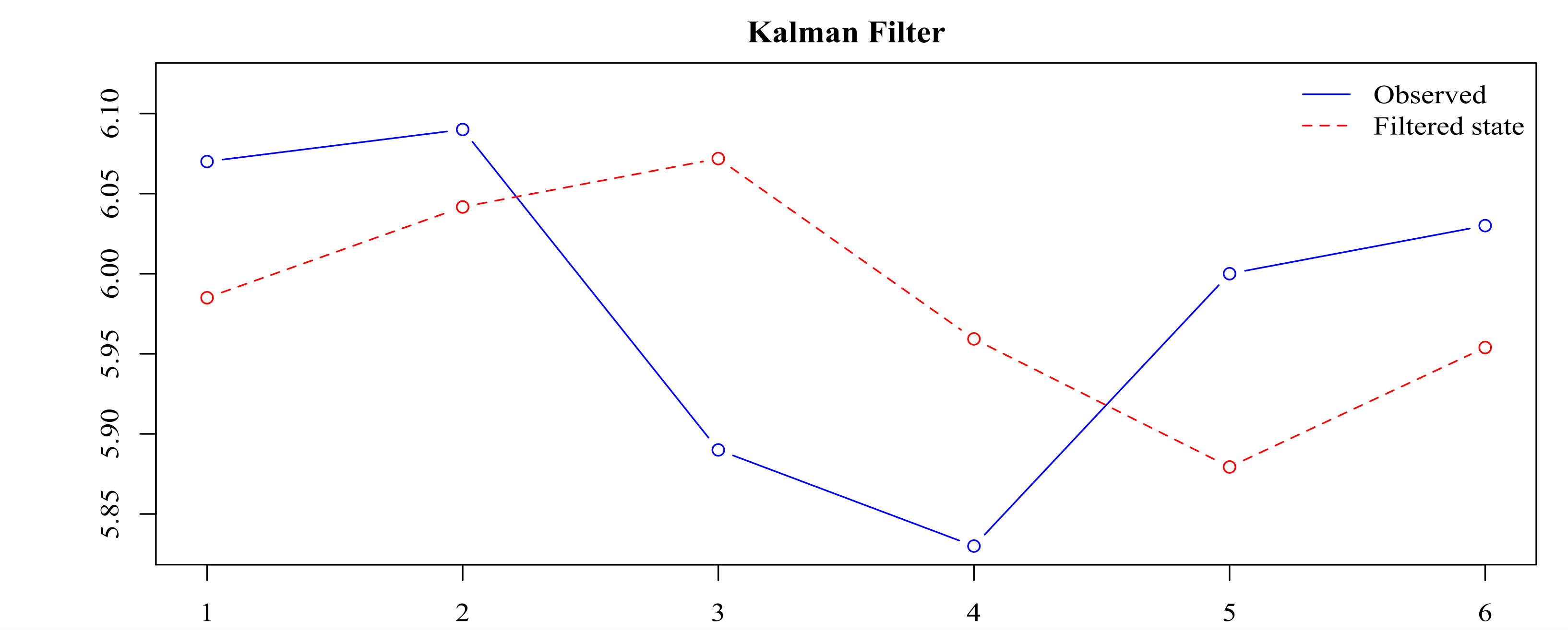

Figure : Kalman filter and observed variable

Kalman Gain

- Over time the values of Q_t and $V_{\varepsilon,t}$ would usually converge towards a constant value, even when they are estimated
- This implies that K_t would also converge on a constant value

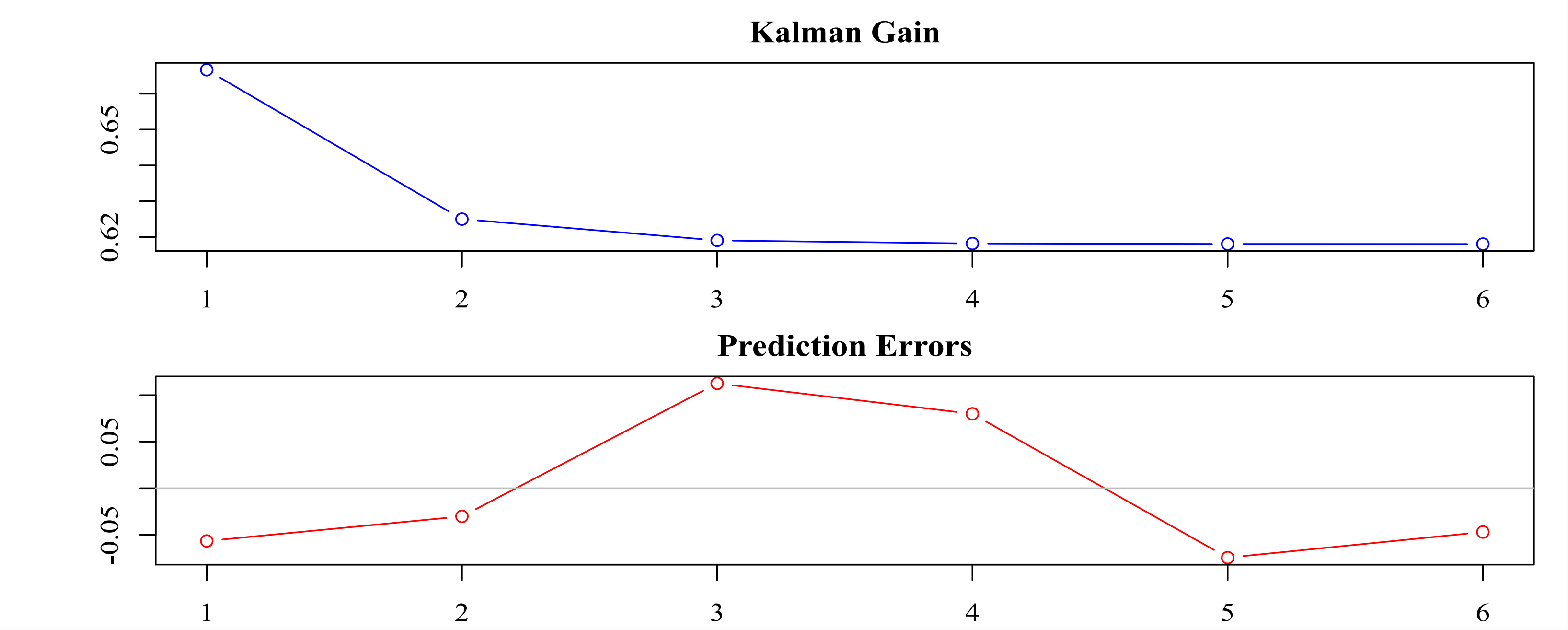

Figure : Kalman gain and prediction errors

Kalman Smoother

- The Kalman filter uses values from y_t and α_t to estimate values for α_{t+1}
- Hence the Kalman filter values appear to project y_t by one period
- To shift these values back a period we use a smoother
- Similar procedure to that of Kalman filter, but we start at the end of the sample and work to the first observation
- Could use the specification:

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

- Where α_t^s is the smoothed estimate and α_t is the filtered estimate
- Value for J_t would then be determined by $(1 - 1/Q_t)$ which is equal to K_t/Q_t

Kalman Smoother

- To initialise the smoother we would use the Kalman filter to generate a value for $\mu^s_{t+6} = 6.0009$
- This serves as the starting value for the smoothing algorithm
- Given that we have calculated all the values for Q_t during the forward pass, we could calculate all the values for J
- This would imply that:

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

- Where we can then generate values for μ^s_{t+4} with the use of μ^s_{t+5}, until we have all the smoothed values for this process

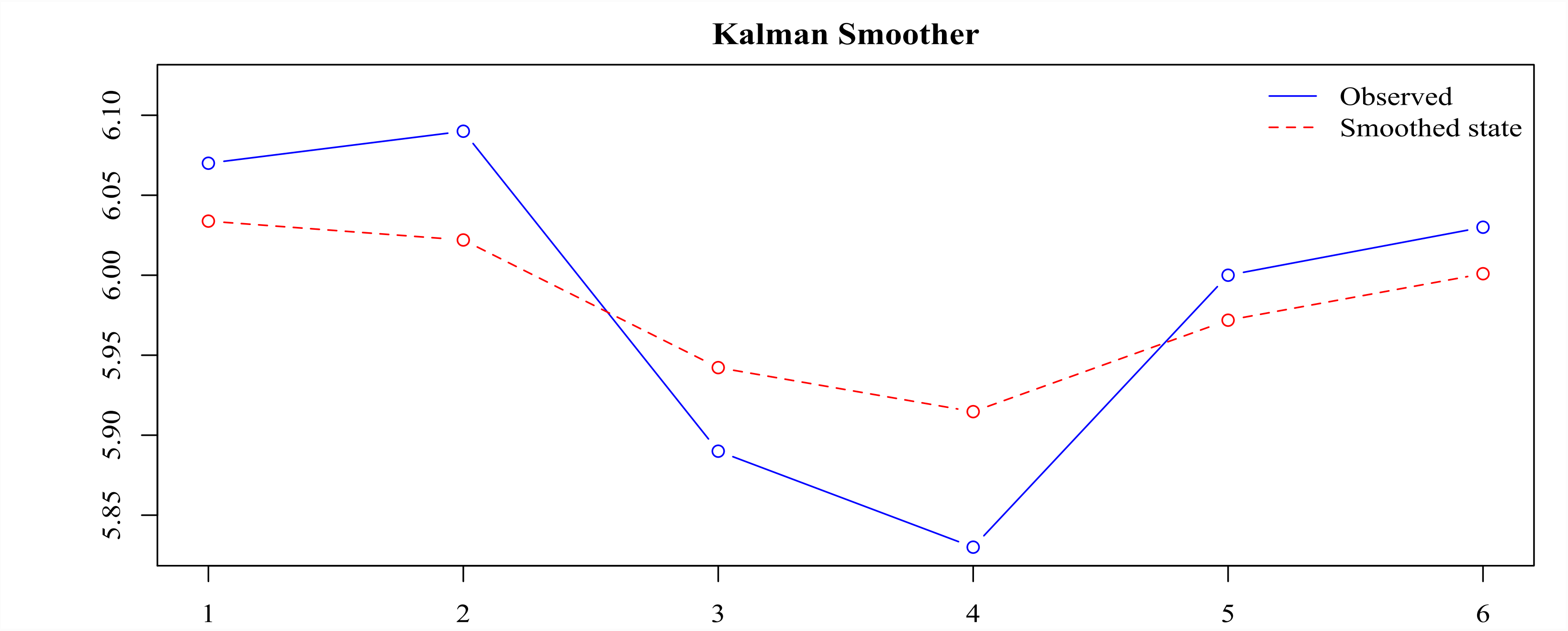

Figure : Kalman smoother and observed variable

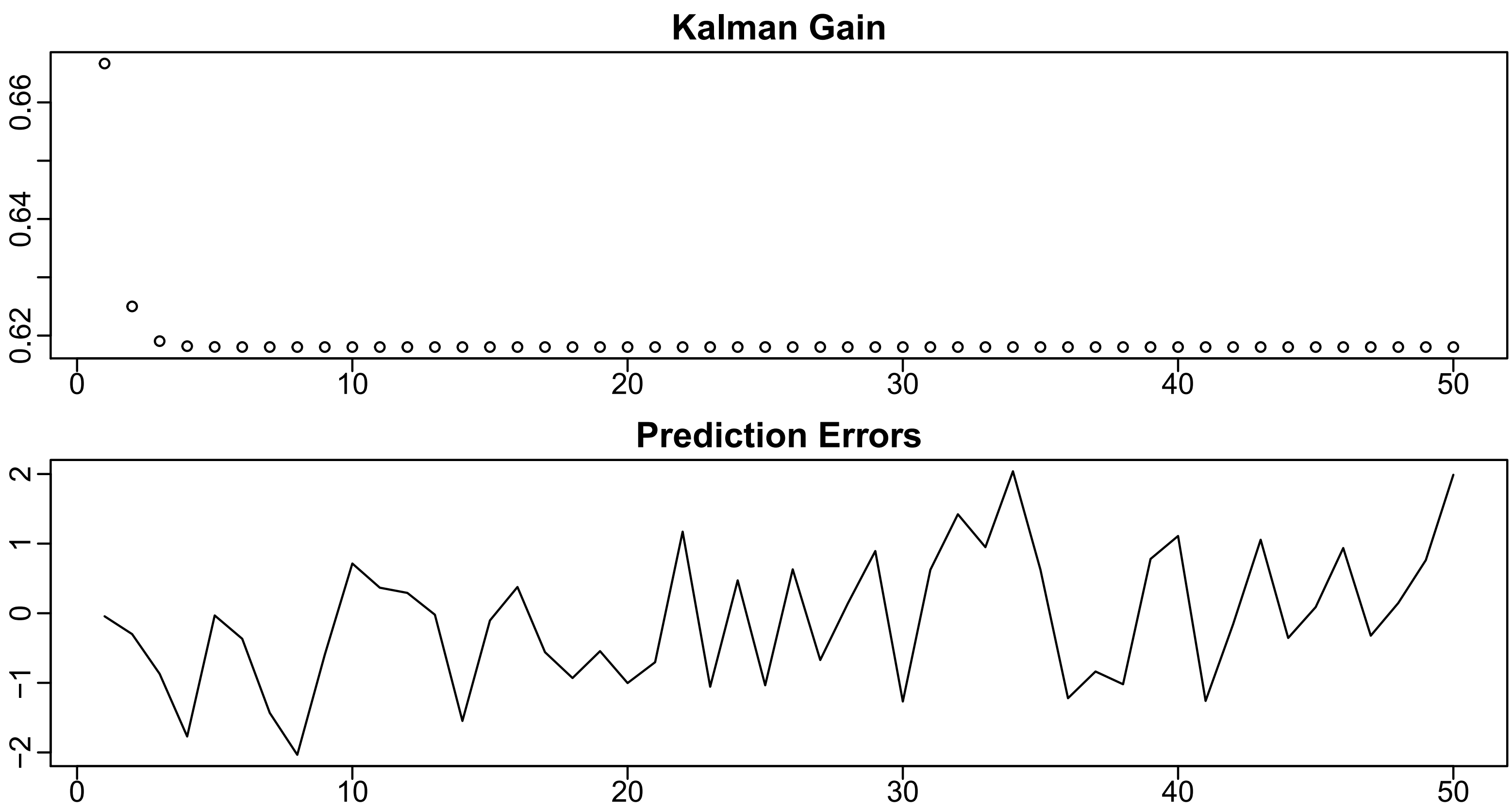

Figure : Kalman gain and one-step ahead prediction errors

Diagnostic evaluation

- Residuals should be:
  - independent
  - homoscedastic
  - normally distributed

- Make use of standardised prediction errors

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

- Example: Inflationary gap (i.e. the difference between core deflator and actual inflationary pressure) with a local level trend model with seasonal and intervention components, and an explanatory variable (output gap)

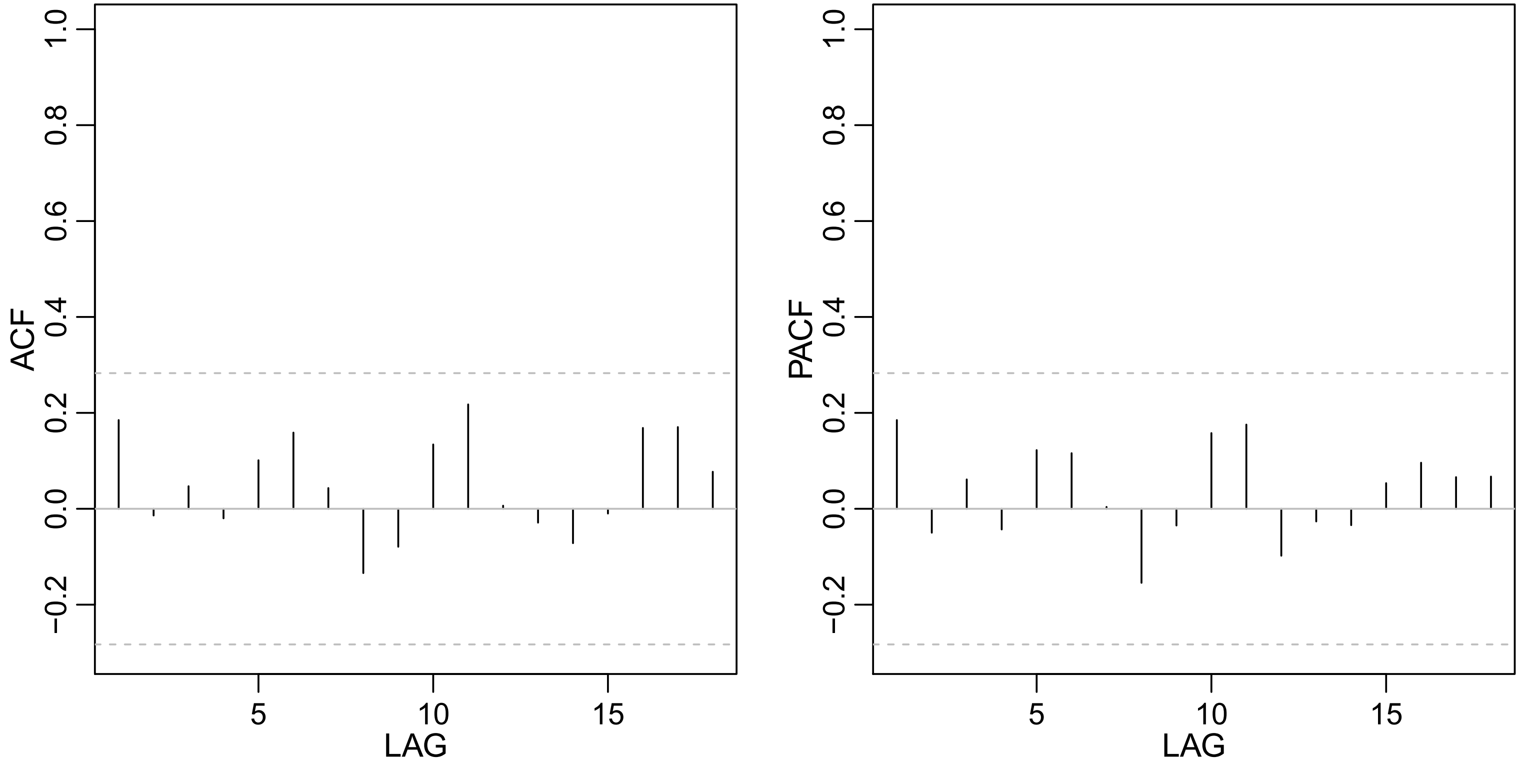

Figure : Diagnostic Tests - Example

Testing for Independence (Box-Ljung)

- The residual autocorrelation from lag k

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

- 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 $l = 1, \dots, k$
- Compare to a χ^2 distribution with $(k - w + 1)$ degrees of freedom (where w is the number of hyperparameters or disturbance variances)
- When calculated value $<$ critical value, the null of independence is not rejected - residuals are not serially correlated

Homoscedasticity of the residuals

- Compare the variance of the residuals in the first third with the variance of the residuals in the last third of the series

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

- where d is the number of diffuse initial state values
- h is the nearest integer to $(T - d)/3$
- Compare to an F-distribution with (h, h) degrees of freedom
- When $1 < H(h) < F(h, h; 0.025)$, the null of equal variances is not rejected - no departure from homoskedasticity in residuals

Normality of the residuals

- Consider 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

$$S = \frac{\frac{1}{T}\sum_{t=1}^{T}(e_t - \bar{e})^3}{\sqrt{\left(\frac{1}{T}\sum_{t=1}^{T}(e_t - \bar{e})^2\right)^3}} \quad K = \frac{\frac{1}{T}\sum_{t=1}^{T}(e_t - \bar{e})^4}{\left(\frac{1}{T}\sum_{t=1}^{T}(e_t - \bar{e})^2\right)^2}$$

- When $N < \chi^2_{(2;0.05)}$ the null hypothesis of normality is not rejected - residuals are not normally distributed

Forecasting

- To compute forecasts continue with the Kalman filter
- Assume that the last observation in the in-sample estimation period is n

$$\alpha_n = \alpha_{t-1} + K_{n-1}(y_{n-1} - F'_{n-1}\alpha_{n-1})$$

- 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)

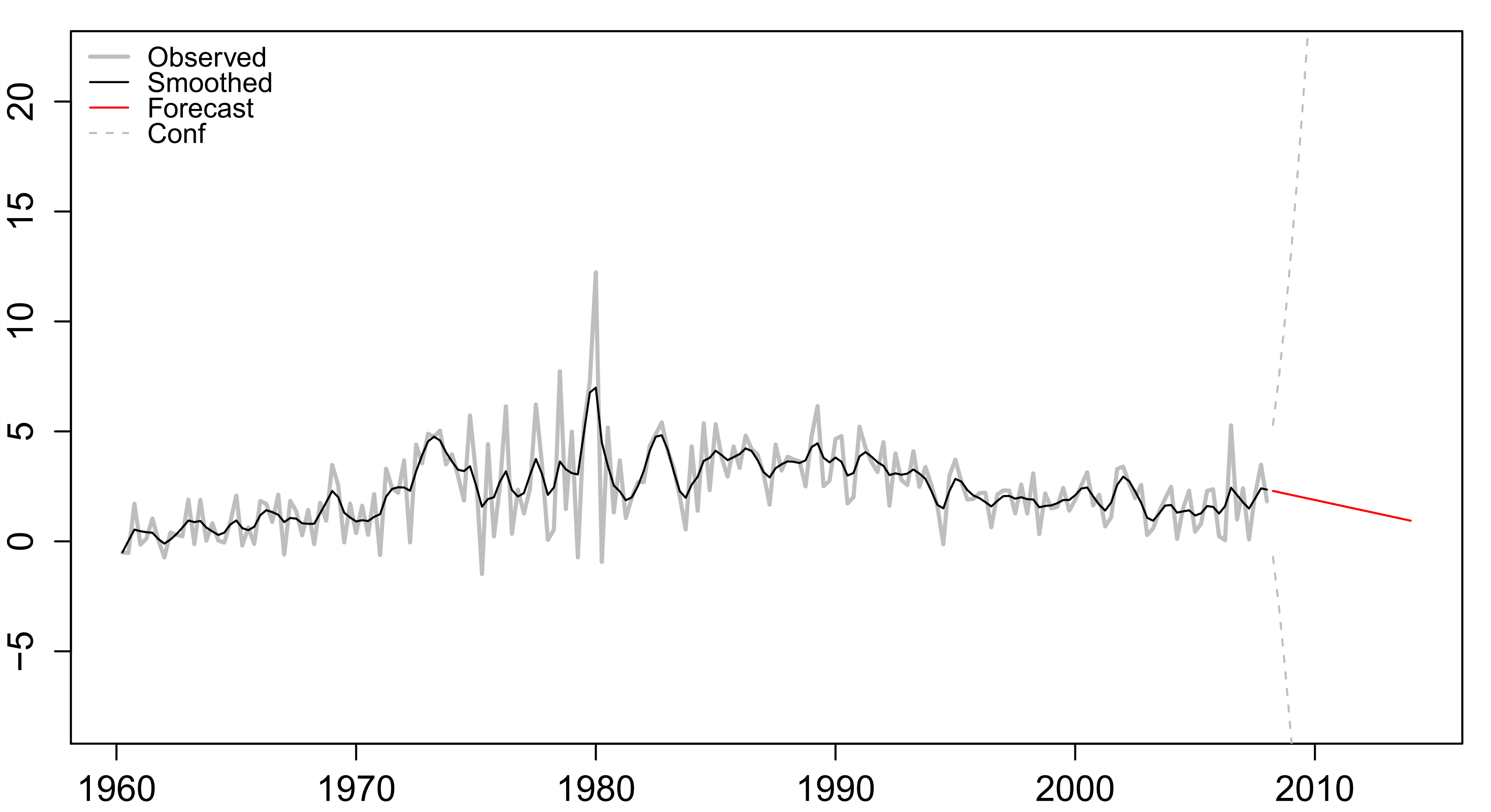

Figure : Filtered trend and forecasts - SA deflator (1960Q1-2014Q1)

Non-stationary ARIMA models

- Typical Box-Jenkins approach makes the series stationary
- Removing the trend by first differencing to create $y_t^\star$

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

- Removing the seasonal with periodicity s by differencing

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

- Or remove both the trend and the seasonal

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

- Where the variable is still not stationary take second difference

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

- After sufficient differencing the appropriate $AR(p)$, $MA(q)$ or $ARMA(p, q)$ is identified

State-Space and Box-Jenkins

- Recall that the local level model has the form

$$y_t = \mu_t + \varepsilon_t$$

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

- Where the first difference of y_t yields

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

- Since state equation $(\mu_t = \mu_{t-1} + \eta_t)$ implies that

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

- We can rewrite the first difference for Δy_t as

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

- Which is stationary and has the correlogram of an $MA(1)$

State-Space and Box-Jenkins

- State space approach explicitly models the non-stationarity of trend and the seasonal components
  - provides simultaneous decomposition of a time series into the respective dynamic components
- Box-Jenkins treats these as nuisance components which need to be removed
  - primarily concerned with short-run dynamics and forecasts
- Advantages of state space:
  - Easy to deal with missing data, time-varying regression coefficients and multivariate extensions
  - Gets around problem when we are not sure when a series is non-stationary

Conclusion

- Models with unobserved components are frequently encountered in economics and finance
- State-space models provide an encompassing approach to describe these dynamic systems
- Allow for the modelling of unobservable variables that could take a number of functional forms
- Use of the Kalman filter facilitates the identification of the unobserved components in the model
- Bayesian techniques are frequently used to estimate the parameters in these models
- These estimation techniques also lend themselves to instances where a nonlinear filter or smoother may need to be used