Macroeconometrics

Lecture 3

Filtering Part 2

Convenor: Stefano Grassi

Approaches to economic time series analysis Cox (1981)

Observation driven models Dynamic stochastic difference equations (the behaviour is explained in terms of the past values of the process and of the innovations) e.g. ARIMA approach and vector autoregressions

Parameter driven models The series is generated by a combination of unobserved components. Structural time series models.

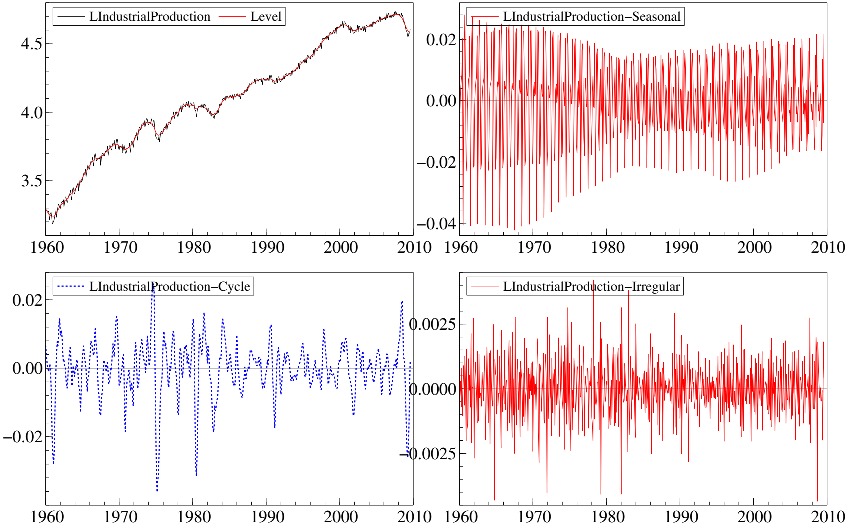

Figure: Decomposition of the USA Industrial Production

- State space models are a class of dynamic models such that the series under investigation are related to quantities called states, which are characterised by simple temporal dependence structure.
- The states have sometimes substantial interpretation. Key estimation problems in macroeconomics concern latent variables, such as the output gap, potential output, the non-accelerating-inflation rate of unemployment, or NAIRU, core inflation, and so forth. These will be typically our states.

State space models and methods

- A linear state space model consists of a *measurement equation* and a *transition equation*.
- The *measurement equation* relates the $N \times 1$ vector time series y_t to an $m \times 1$ vector of unobservable components called *states*, α_t:

$$y_t = Z_t \alpha_t + \varepsilon_t, \qquad t = 1, 2, \ldots, n, \tag{1}$$

  where Z_t is an $N \times m$ matrix, and $\varepsilon_t \sim \mathrm{N}(0, H_t)$.
- The *transition equation* is a dynamic linear model for the states α_t, taking the form of a first order vector autoregression:

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t, \tag{2}$$

  where T_t is $m \times m$, R_t is $m \times g$ and $\eta_t \sim \mathrm{N}(0, Q_t)$.

- State space methods are tools for inference in SSM: estimate any unknown parameters, estimate the states and assess the uncertainty of the estimates, perform diagnostic checking and assess goodness of fit, forecast future states and observations, and so forth.
- State space methods are algorithms with a recursive structure.
- The *Kalman filter* is the basic one: computes the one-step-ahead prediction errors, provides the evaluation of the the likelihood function via the prediction error decomposition and inferences on the unobserved states (prediction and filtering).
- *Smoothing filters*: provide improved inferences about the state components that use all the available information. The smoothing errors, a by product of smoothing filters, are essential in cross validation and diagnostic checking.

Example 1: SSR for the MA(1) model

$$y_t = \epsilon_t + \theta_1 \epsilon_{t-1}, \qquad \epsilon_t \sim \text{NID}(0, \sigma^2)$$

Let $\alpha_t = \begin{bmatrix} \epsilon_t \\ \epsilon_{t-1} \end{bmatrix}$, $\epsilon_t \sim \text{NID}(0, \sigma^2)$, $Z = [1, \theta]$, $\eta_t = \epsilon_{t+1}$,

$$T = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad R = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\alpha_1 \sim \text{N}\left(0, \sigma^2 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\right)$$

Example 2: SSR for the AR(2) model

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t, \qquad \epsilon_t \sim \text{NID}(0, \sigma^2)$$

Let $\alpha_t = [y_t, y_{t-1}]'$, $\epsilon_t \sim \text{NID}(0, \sigma^2)$, $Z = [1, 0]$, $\eta_t = \epsilon_{t+1}$

$$T = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix}, \quad R = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\alpha_1 \sim \text{N}\left(0, \sigma^2 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\right)$$

Trend models

- The trend component is concerned with the long run behaviour of the series; Beveridge and Nelson (1981) define it as "the value the series would take if it were on its long run path". Operationally, for an I(1) process it is defined as y_t plus all forecastable future changes, beyond the mean rate of drift. Also Harvey (1989, 6.1.1) uses a prediction argument, defining the trend as the "part of the series which when extrapolated gives the clearest indication of the future long-term movements in the series".
- The specification of a time series model for the trend component varies according to the features displayed by the series under investigation and any prior knowledge.

Local level model

$$\begin{aligned} y_t &= \mu_t + \epsilon_t, & \epsilon_t &\sim \text{NID}(0, \sigma^2_\epsilon) \\ \mu_t &= \mu_{t-1} + \eta_t, & \eta_t &\sim \text{NID}(0, \sigma^2_\eta) \end{aligned} \tag{3}$$

ϵ_t and η_t are independent.
$\sigma^2_\eta = 0 \Rightarrow$ constant level: $\mu_t = \mu$,
$\sigma^2_\epsilon = 0 \Rightarrow y_t \sim$ RW: $y_t = \mu_t$
Stationary representation: $\Delta y_t = \eta_t + \Delta\epsilon_t$.

$$\text{E}(\Delta y_t) = 0$$

$$\begin{aligned} \gamma(0) &= \text{E}(\Delta y_t^2) &&= \sigma^2_\eta + 2\sigma^2_\epsilon \\ \gamma(1) &= \text{E}(\Delta y_t \Delta y_{t-1}) &&= -\sigma^2_\epsilon \\ \gamma(\tau) &= \text{E}(\Delta y_t \Delta y_{t-\tau}) &&= 0, \quad \tau > 1 \end{aligned}$$

- The ACF exhibits a cut-off at lag one, with

$$\rho(1) = -\frac{\sigma_\epsilon^2}{\sigma_\eta^2 + 2\sigma_\epsilon^2} = -\frac{1}{q+2}$$

taking values in $[0, -1/2]$. where $q = \sigma_\eta^2/\sigma_\epsilon^2$ is the signal-noise ratio.

- Spectral density of Δy_t

$$f(\lambda) = \frac{1}{2\pi}[\sigma_\eta^2 + 2(1 - \cos\lambda)\sigma_\epsilon^2]$$

Reduced form representation: IMA(1,1)

$$\Delta y_t = (1+\theta L)\xi_t, \quad \xi_t \sim \text{NID}(0,\sigma^2)$$

$$\begin{array}{rcccl} \gamma(0) & = & \text{E}(\Delta y_t^2) & = & (1+\theta^2)\sigma^2 \\ \gamma(1) & = & \text{E}(\Delta y_t \Delta y_{t-1}) & = & \theta\sigma^2 \\ \gamma(\tau) & = & \text{E}(\Delta y_t \Delta y_{t-\tau}) & = & 0, \quad \tau > 1 \end{array}$$

from the identity:

$$\rho(1) = \frac{\theta}{1+\theta^2} = -\frac{1}{q+2}$$

can determine

$$\theta = [(q^2+4q)^{1/2} - 2 - q]/2$$

hence θ constrained to lie in the range [-1,0]. Also,

$$\sigma_\eta^2 = (1+\theta)^2\sigma^2$$

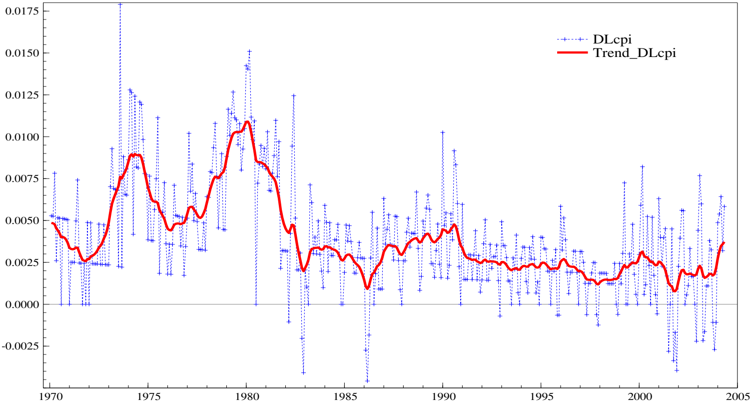

Figure: U.S. monthly inflation and estimates of local level.

The Kalman filter

- Let us define $Y_t = \{y_1, y_2, \dots, y_t\}$ (information up to time t)

$$\tilde{\alpha}_{t+1|t} = \mathrm{E}(\alpha_{t+1}|Y_t), \quad \mathrm{Var}(\alpha_{t+1}|Y_t) = P_{t+1|t}$$

- The **Kalman filter** (KF) is the following recursive algorithm: for $t = 1, \dots, n$,

$$\begin{aligned} \nu_t &= y_t - Z_t\tilde{\alpha}_{t|t-1}, & F_t &= Z_t P_{t|t-1} Z_t' + H_t \\ & & K_t &= T_t P_{t|t-1} Z_t' F^{-1}, \\ & & L_t &= T_t - K_t Z_t \\ \tilde{\alpha}_{t+1|t} &= T_t\tilde{\alpha}_{t|t-1} + K_t\nu_t, & P_{t+1|t} &= T_t P_{t|t-1} T_t' + R_t Q_t R_t' \end{aligned}$$

$\nu_t = y_t - \mathrm{E}(y_t|Y_{t-1})$ are the innovations or one-step-ahead prediction errors, with variance matrix F_t. Moreover assume that α_t given Y_{t-1} is $\mathrm{N}(\tilde{\alpha}_{t|t-1}, P_{t|t-1})$

Smoothing

- Smoothing refers to optimal estimation of the unobserved components based also on future observations.
- *Fixed-interval smoothing* is concerned with computing the full set of smoothed estimates $\tilde{\alpha}_{t|n} = \mathrm{E}(\alpha_t|Y_n)$, along with $\mathrm{Cov}(\alpha_t|Y_n) = P_{t|n}$ for the entire sample, $t = 1, \ldots, n$.
- *Fixed-point smoothing* computes smoothed estimates with reference to a fixed point in time, $\tilde{\alpha}_{t|t+j}$, where t is fixed and $j = 0, 1, \ldots n - t$.
- *Fixed-lag smoothing* computes $\tilde{\alpha}_{t-j|t}$, where $j > 0$ is fixed and $t = j + 1, \ldots, n$.

- Fixed interval smoothing algorithm (de Jong, JASA 1988). Backwards recursive algorithm starting at $t = n$, with initial values $r_n = 0$ and $N_n = 0$:

$$
\begin{aligned}
r_{t-1} &= L_t' r_t + Z_t' F_t^{-1} \nu_t, & N_{t-1} &= Z_t' F_t^{-1} Z_t + L_t' N_t L_t, \\
\tilde{\alpha}_{t|n} &= \tilde{\alpha}_{t|t-1} + P_{t|t-1} r_{t-1}, & P_{t|n} &= P_{t|t-1} - P_{t|t-1} N_{t-1} P_{t|t-1},
\end{aligned}
$$

  where $L_t = T_t - K_t Z_t$.
- A preliminary forward KF pass is required to store the quantities $\tilde{\alpha}_{t|t-1}$, $P_{t|t-1}$, ν_t, F_t and K_t.

Diagnostic checking

The assumptions behind the local level model are that the disturbances ϵ_t and η_t are normally distributed and serieally indipendent with constant variance. Under those assumptions the standardized one-step forecasts errors are:

$$e_t = \frac{\nu_t}{\sqrt{F_t}} \tag{4}$$

should be normally distributed with unitary variance. Departing from this underlines that the model is not well specified.

Maximum Likelihood Estimation

- Let $\theta \in \Theta$ denote a vector containing the so-called hyperparameters, i.e. the vector of structural parameters other than the scale factor σ^2.
- The state space model depends on θ via the system matrices $Z_t = Z_t(\theta), G_t = G_t(\theta), T_t = T_t(\theta), H_t = H_t(\theta)$.

Maximum Likelihood Estimation for Local Level model

Consider the local level model:

$$
\begin{aligned}
y_t &= \alpha_t + \epsilon_t \qquad \epsilon_t \sim \mathrm{N}(0, \sigma_\epsilon^2) \\
\alpha_t &= \alpha_{t-1} + \eta_t \quad \eta_t \sim \mathrm{N}(0, \sigma_\eta^2)
\end{aligned} \tag{5}
$$

The parameter to be estimated are σ_ϵ^2 and σ_η^2. Those parameter are restricted in the region $[0, +\infty)$. It is much better to maximize the function in the domain $(-\infty, +\infty)$.

Reparametrization

The vector of parameters, θ, has two unrestricted elements, which are related to the model's hyperparameters by:

$$\sigma_\eta^2 = \exp(2\theta_1), \qquad \sigma_\epsilon^2 = \exp(2\theta_2),$$

or in the inverse way:

$$\theta_1 = \frac{1}{2}\log(\sigma_\eta^2) \qquad \theta_2 = \frac{1}{2}\log(\sigma_\epsilon^2)$$

Likelihood using the Kalman filter

- Let $\mathrm{L}(Y_n|\theta)$ denote the log-likelihood function, that is the log of the joint density of the sample time series $\{y_1, \ldots, y_n\}$ as a function of the parameters θ.
- The log-likelihood can be evaluated by the prediction error decomposition:

$$\mathrm{L}(Y_n|\theta) = \log f(y_1, \ldots, y_n|\theta) = \sum_{t=1}^{n} \log f(y_t|Y_{t-1};\theta).$$

- The predictive density $f(y_t|Y_{t-1};\theta)$ is evaluated with the support of the **Kalman Filter**.

Proof

- In this case we assume **NORMALITY**.
- Recall that $F_t = \mathrm{Var}(y_t | Y_{t-1})$ and $v_t = y_t - Z_t \tilde{\alpha}_{t|t-1}$.
- Then we can substitute $\mathrm{N}(Z_t \tilde{\alpha}_{t|t-1}, F_t)$ for $f(y_t | Y_{t-1})$ and we get, apart from constant :

$$\log \mathrm{L}(Y_n|\theta) = -\frac{1}{2}\left(\sum_{t=1}^{n} \log|F_t| + \sum_{t=1}^{n} \nu_t' F_t^{-1} \nu_t\right).$$

- The likelihood function can be maximized numerically by a quasi-Newton optimization routine.
- Analytical expressions for the score vector, with respect to the parameters in G_t and H_t, and for the information matrix are available.

Example

- This is a series of readings of the annual flow of the Nile river at Aswan for 1871 to 1970. This series is originally considered by Cobb (1978) and analysed more recently by Balke (1993).
- A second data set used in this lesson is the airline data, consisting of the number of UK airline passengers (in thousands, from January 1949 to December 1960), see Box and Jenkins (1976).

Estimation of a local level model

- Suppose that we want to estimate the series Nile river
- Suppose that we believe that the appropriate model is the local level model

$$\begin{aligned} y_t &= \mu_t + \epsilon_t, & \epsilon_t &\sim \text{NID}(0, \sigma_\epsilon^2) \\ \mu_t &= \mu_{t-1} + \eta_t, & \eta_t &\sim \text{NID}(0, \sigma_\eta^2) \end{aligned} \tag{6}$$

ϵ_t and η_t are independent. $\sigma_\eta^2 = 0 \Rightarrow$ constant level: $\mu_t = \mu$, $\sigma_\epsilon^2 = 0 \Rightarrow y_t \sim$ RW: $y_t = \mu_t$

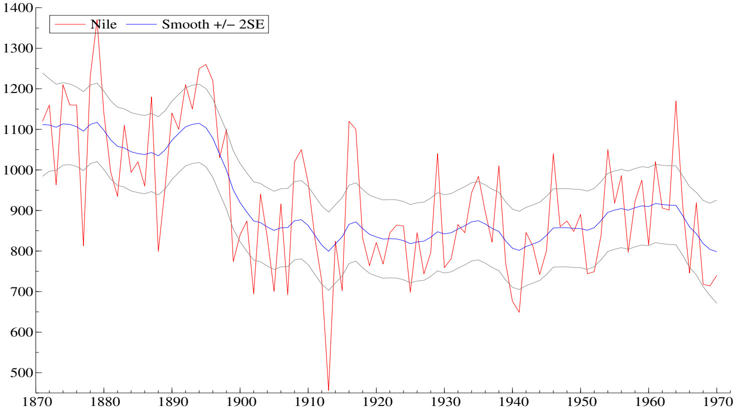

Figure: LLM for the Nile river data.

Seasonal adjustment

- Seasonal adjustment is a relatively easy task when time series are modelled as an unobserved components time series model in which a seasonal component is included.
- The estimated seasonal component is subtracted from the original time series in order to get the seasonally adjusted series. In the same way the original time series is detrended by subtracting the estimated trend component.

Cycles

- One definition of *business cycle* broadly refers to the recurrent, though not exactly periodic, deviations around the long term path of the series.
- A model for the cyclical component should be capable of reproducing commonly acknowledged essential features, such as the presence of strong autocorrelation, determining the recurrence and alternation of phases, and the dampening of the fluctuations, or zero long run persistence.

Periodic functions

- Periodic functions: $\cos(t)$ and $\sin(t)$ are periodic with period 2π: $\cos(t + 2\pi j) = \cos t$, $\sin(t + 2\pi j) = \sin t$.
- Linear transformation of the argument: $\cos(\lambda_c t - \theta)$ is periodic with period $2\pi/\lambda_c$. $\lambda_c/2\pi$ is the frequency (number of cycles in the unit interval).
- Using the addition formulae

$$\begin{aligned}\cos(\alpha \pm \beta) &= \cos\alpha\cos\beta \mp \sin\alpha\sin\beta \\ \sin(\alpha \pm \beta) &= \sin\alpha\cos\beta \mp \cos\alpha\sin\beta\end{aligned}$$

we can re-express the periodic function. We introduce an amplitude parameter, a, which is the maximum of the function:

$$a\cos(\lambda_c t - \theta) = \alpha\cos\lambda_c t + \alpha^*\sin\lambda_c t$$

where $\alpha = a\cos\theta$ and $\alpha^* = a\sin\theta$ so that $a = \sqrt{\alpha^2 + \alpha^{*2}}$.

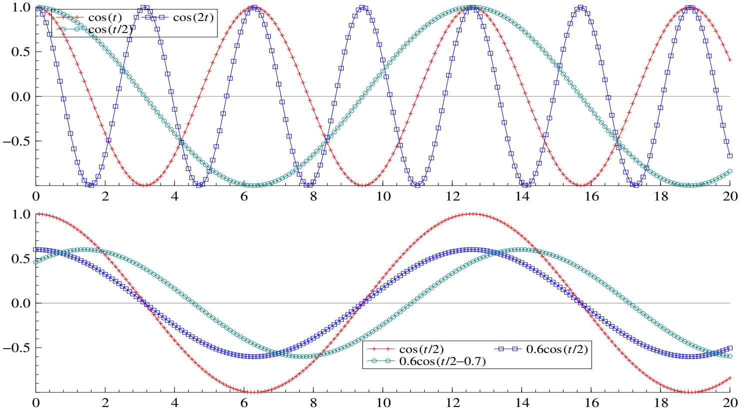

Figure: Plot of $\cos(\lambda_c t - \theta)$

- Deterministic cycle model:

$$\psi_t = \alpha \cos(\lambda_c t) + \alpha^* \sin(\lambda_c t),$$

  λ_c takes a particular value in $[0, \pi]$.
- This defines a perfectly periodic function of time, repeating itself every $\bar{p} = 2\pi/\lambda_c$ time units, where $\bar{p}$ is the period, with constant amplitude $a = (\alpha^2 + \alpha^{*2})^{1/2}$
- It can be rewritten as

$$\psi_t = a \cos(\lambda_c t - \theta).$$

- A stochastic cycle can be obtained by letting the coefficients α and α^* follow an AR(1) process with coefficient ρ, $0 \leq \rho \leq 1$, that is responsible for the dampening of the fluctuations.
- Hence:

$$\begin{aligned} \psi_t &= [\cos \lambda_c t, \quad \sin \lambda_c t] \begin{bmatrix} \alpha_t \\ \alpha_t^* \end{bmatrix} \\ \alpha_{t+1} &= \rho \alpha_t + \tilde{\kappa}_t \\ \alpha_{t+1}^* &= \rho \alpha_t^* + \tilde{\kappa}_t^* \end{aligned} \quad (7)$$

  where $\tilde{\kappa}_t$ and $\tilde{\kappa}_t^*$ are mutually independent NID disturbances with zero mean and common variance σ_κ^2.
- The AR processes α_t, α_t^* are related to the amplitude of the oscillation as ψ_t can be rewritten:

$$\psi_t = a_t \cos(\lambda_c t - \vartheta_t)$$

  where $a_t = (\alpha_t^2 + \alpha_t^{2*})^{.5}$ is the time varying amplitude and $\vartheta_t = \tan^{-1}(\alpha_t^* / \alpha_t)$ is the phase shift.

Equivalently, a deterministic cycle can be generated recursively by

$$\begin{aligned} \psi_{t+1} &= \cos\lambda_c \psi_t + \sin\lambda_c \psi_t^*, \\ \psi_{t+1}^* &= -\sin\lambda_c \psi_t + \cos\lambda_c \psi_t^*, \end{aligned}$$

with starting values $\psi_0 = \alpha$ and $\psi_0^* = \alpha^*$, a stochastic cycle is constructed multiplying the right hand side of these two equations by ρ (*damping factor*), and adding stochastic disturbances in the form of NID sequences, giving:

$$\begin{bmatrix} \psi_{t+1} \\ \psi_{t+1}^* \end{bmatrix} = \rho \begin{bmatrix} \cos\lambda_c & \sin\lambda_c \\ -\sin\lambda_c & \cos\lambda_c \end{bmatrix} \begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} + \begin{bmatrix} \kappa_t \\ \kappa_t^* \end{bmatrix}, \tag{8}$$

where again $\kappa_t \sim \text{NID}(0, \sigma_\kappa^2)$ and $\kappa_t^* \sim \text{NID}(0, \sigma_\kappa^2)$, are mutually independent.

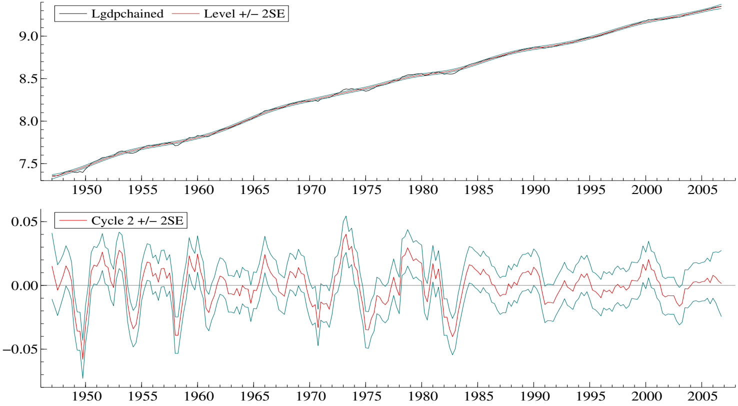

Figure: US GDP (logs). Smoothed estimates of the level and cycle.

Trigonometric seasonal component

- The so called trigonometric seasonal component is given by

$$\gamma_t = \sum_{j=1}^{[s/2]} \gamma_{j,t}^{+} \tag{9}$$

where

$$\begin{bmatrix} \gamma_{j,t+1}^{+} \\ \gamma_{j,t+1}^{*} \end{bmatrix} = \begin{bmatrix} \cos\lambda_j & \sin\lambda_j \\ -\sin\lambda_j & \cos\lambda_j \end{bmatrix} \begin{bmatrix} \gamma_{j,t}^{+} \\ \gamma_{j,t}^{*} \end{bmatrix} + \begin{bmatrix} \tilde{\kappa}_{j,t} \\ \tilde{\kappa}_{j,t}^{*} \end{bmatrix}, \tag{10}$$

where $\lambda_j = 2\pi j/s$ as the $j - th$ seasonal frequency and

$$\begin{bmatrix} \gamma_{j,t+1}^{+} \\ \gamma_{j,t+1}^{*} \end{bmatrix} \sim NID\left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma_{\kappa}^{2} I_2 \right\}, \quad j = 1, \ldots, [s/2]. \tag{11}$$

where $\tilde{\kappa}_t$ and $\tilde{\kappa}_t^{*}$ are mutually independent NID disturbances.

- Note that for s even $[s/2]$, while for s odd, $[s/2] = (s-1)/2$.

State Space Form

- The model can be cast in state space form as follows

$$y_t = \mu_t + \gamma_t^+ + \varepsilon_t \tag{12}$$

where μ_t is the trend and the γ_t^+ is the seasonality component. ε_t is the usual error

The state space representation. Case of quarterly data

$$
\begin{aligned}
y_t &= Z\alpha_t + G\varepsilon_t \\
&= \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \end{bmatrix} \alpha_t + G\varepsilon_t,
\end{aligned}
$$

where the vector of states α_t is equal to

$$
\alpha_t^{'} = \begin{bmatrix} \mu_t & \gamma_{1,t} & \gamma_{1,t}^{*} & \gamma_{2,t} & \gamma_{2,t}^{*} \end{bmatrix}.
$$

The transition equation of the model has the following structure

$$\alpha_{t+1} = T\alpha_t + H\varepsilon_t,$$

where

$$T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \cos\lambda_1 & \sin\lambda_1 & 0 & 0 \\ 0 & -\sin\lambda_1 & \cos\lambda_1 & 0 & 0 \\ 0 & 0 & 0 & \cos\lambda_2 & \sin\lambda_2 \\ 0 & 0 & 0 & -\sin\lambda_2 & \cos\lambda_2 \end{bmatrix},$$

where $\lambda_1 = \frac{\pi}{2}$ and $\lambda_2 = \pi$.

The disturbance term is equal to

$$\varepsilon_t' = \left[\begin{array}{cccccc} \frac{\epsilon_t}{\sigma_\epsilon} & \frac{\eta_t}{\sigma_\eta} & \frac{\gamma_{1,t}}{\sigma_{\gamma_1}} & \frac{\gamma^*_{1,t}}{\sigma_{\gamma_1}} & \frac{\gamma_{2,t}}{\sigma_{\gamma_2}} & \frac{\gamma^*_{2,t}}{\sigma_{\gamma_2}} \end{array} \right].$$

Finally the two matrices that premultiply the disturbance term ε_t are

$$G_t = G = \left[\begin{array}{cccccc} \sigma_\epsilon & 0 & 0 & 0 & 0 & 0 \end{array} \right],$$

and

$$H = \left[\begin{array}{cccccc} 0 & \sigma_\eta & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^*_{\gamma_1} & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{\gamma_1} & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^*_{\gamma_2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{\gamma_2} \end{array} \right].$$

Trigonometric seasonal component

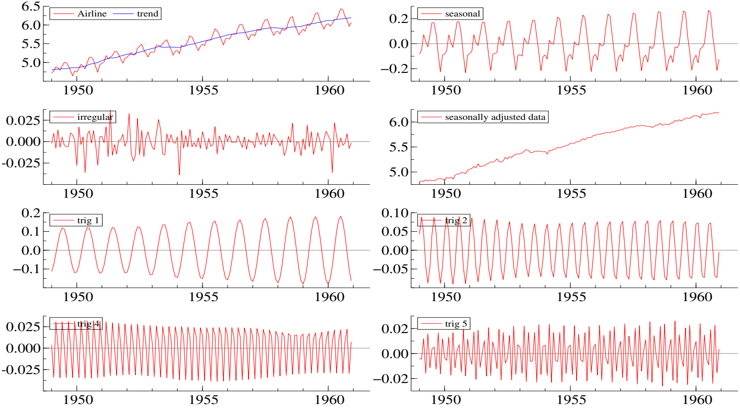

Figure: BSM for the Airline dataset.

The Kalman filter

Let us define $Y_t = \{y_1, y_2, \ldots, y_t\}$ (information up to time t)

$$\tilde{\alpha}_{t+1|t} = \mathrm{E}(\alpha_{t+1}|Y_t), \quad \mathrm{Var}(\alpha_{t+1}|Y_t) = P_{t+1|t}$$

The **Kalman filter** (KF) is the following recursive algorithm: for $t = 1, \ldots, n$,

$$\begin{array}{ll} \nu_t = y_t - Z_t\tilde{\alpha}_{t|t-1}, & F_t = Z_t P_{t|t-1} Z_t' + H_t \\ & K_t = T_t P_{t|t-1} Z_t' F^{-1}, \\ & L_t = T_t - K_t Z_t \\ \tilde{\alpha}_{t+1|t} = T_t \tilde{\alpha}_{t|t-1} + K_t \nu_t, & P_{t+1|t} = T_t P_{t|t-1} T_t' + R_t Q_t R_t' \end{array}$$

$\nu_t = y_t - \mathrm{E}(y_t|Y_{t-1})$ are the innovations or one-step-ahead prediction errors, with variance matrix F_t. Moreover assume that α_t given Y_{t-1} is $\mathrm{N}(\tilde{\alpha}_{t|t-1}, P_{t|t-1})$

Proof of the Kalman filter

Assume $\tilde{\alpha}_{t+1|t}$, $P_{t+1|t}$ given at the t-th run of the KF. The available information set is Y_t.

$$\begin{aligned} \tilde{\alpha}_{t+1|t} &= \mathrm{E}(T_t\alpha_t + R_t\eta_t|Y_t) \\ &= T_t E(\alpha_t|Y_t) \end{aligned} \tag{13}$$

$$\begin{aligned} P_{t+1|t} &= \mathrm{Var}(T_t\alpha_t + R_t\eta_t|Y_t) \\ &= T_t \mathrm{Var}(\alpha_t|Y_t) T_t^{'} + R_t Q_t R_t^{'} \end{aligned} \tag{14}$$

The one-step forecast error of y_t given Y_{t-1} are given by the following relation:

$$v_t = y_t - \mathrm{E}(y_t|Y_{t-1}) = y_t - \mathrm{E}(Z_t\alpha_t + \varepsilon_t|Y_{t-1}) = y_t - Z_t\tilde{\alpha}_{t|t-1} \tag{15}$$

this quantity is fundamental for the likelihood evaluation.

From the previous relation if Y_{t-1} and v_t are fixed then Y_t is fixed and vice versa. Thus

$$\mathrm{E}(\alpha_t|y_t) = \mathrm{E}(\alpha_t|Y_{t-1}, v_t) \tag{16}$$

Properties of v_t:

- $\mathrm{E}(v_t|Y_{t-1}) = \mathrm{E}(y_t - Z_t\tilde{\alpha}_t|Y_{t-1}) = 0.$
- $\mathrm{Cov}(y_j, v_t) = \mathrm{E}[y_j \mathrm{E}(v_t|Y_{t-1})'] = 0$ with $j = 1, \ldots, T-1.$

This is a very important step!!!

Multivariate normal regression

Suppose that x, y, z are three random vectors such that their joint distribution is multivariate normal. In addition, assume that the diagonal block covariance matrix Σ_{ww} is nonsingular for $w = x, y, z$ and $\Sigma_{yz} = 0$. Then,

- $\mathrm{E}(x|y,z) = E(x|y) + \Sigma_{xz}\Sigma_{zz}^{-1}(z - \mu_z)$;
- $\mathrm{Var}(x|y,z) = Var(x|y) - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{xz}$

From the previous theorem we therefore have:

$$\begin{aligned} \mathrm{E}(\alpha_t | Y_t) &= \mathrm{E}(\alpha_t | Y_{t-1}, v_t) \\ &= \mathrm{E}(\alpha_t | Y_{t-1}) + \mathrm{Cov}(\alpha_t, v_t)[\mathrm{Var}(v_t)]^{-1} v_t \\ &= \tilde{\alpha}_t + M_t F_t^{-1} v_t \end{aligned} \tag{17}$$

Where

$$\begin{aligned} M_t = \mathrm{Cov}(\alpha_t, v_t) &= \mathrm{E}[\mathrm{E}\{\alpha_t (Z_t \alpha_t + \epsilon_t - Z_t \tilde{\alpha}_t)^{'})| Y_{t-1}\}] \\ &= \mathrm{E}[\mathrm{E}\{\alpha_t (\alpha_t - \tilde{\alpha}_t)^{'} Z_t^{'} | Y_{t-1}\}] = P_t Z_t^{'} \end{aligned} \tag{18}$$

$$F_t = \mathrm{Var}(Z_t \alpha_t + \epsilon_t - Z_t \tilde{\alpha}_t) = Z_t P_t Z_t^{'} + H_t \tag{19}$$

Substituting in (15) we get the following results:

$$\begin{aligned} \tilde{\alpha}_{t+1} &= T_t \tilde{\alpha}_t + T_t M_t F_t^{-1} v_t \\ &= T_t \tilde{\alpha}_t + K_t v_t \end{aligned} \tag{20}$$

where $K_t = T_t M_t F_t^{-1} = T_t P_t Z_t^{'} F_t^{-1}$

By the second result of Multivariate normal lemma we can calculate the following quantity:

$$\begin{aligned}\text{Var}(\alpha_t|Y_t) &= \text{Var}(\alpha_t|Y_{t-1}, v_t) \\ &= \text{Var}(\alpha_t|Y_{t-1}) - \text{Cov}(\alpha_t, v_t)[\text{Var}(v_t)]^{-1}\text{Cov}(\alpha_t, v_t)' \\ &= P_t - M_t F_t^{-1} M_t' \\ &= P_t - P_t Z_t' F_t^{-1} Z_t P_t\end{aligned} \tag{21}$$

Substituting in (16) we get the following results:

$$P_{t+1} = T_t P_t L_t' + R_t Q_t R_t' \tag{22}$$

where $L_t = T_t - K_t Z_t$

Maximum Likelihood Estimation

- Let $\theta \in \Theta$ denote a vector containing the so-called hyperparameters, i.e. the vector of structural parameters other than the scale factor σ^2.
- The state space model depends on θ via the system matrices $Z_t = Z_t(\theta), G_t = G_t(\theta), T_t = T_t(\theta), H_t = H_t(\theta)$.

Consider the local level model:

$$
\begin{aligned}
y_t &= \alpha_t + \epsilon_t \qquad \epsilon_t \sim \mathsf{N}(0, \sigma_\epsilon^2) \\
\alpha_t &= \alpha_{t-1} + \eta_t \quad \eta_t \sim \mathsf{N}(0, \sigma_\eta^2)
\end{aligned} \tag{23}
$$

The parameter to be estimated are σ_ϵ^2 and σ_η^2. Those parameter are restricted in the region $[0, +\infty)$. It is much better to maximize the function in the domain $(-\infty, +\infty)$.

Reparametrization

The vector of parameters, θ, has two unrestricted elements, which are related to the model's hyperparameters by:

$$\sigma_\eta^2 = \exp(2\theta_1), \qquad \sigma_\epsilon^2 = \exp(2\theta_2),$$

or in the inverse way:

$$\theta_1 = \frac{1}{2}\log(\sigma_\eta^2) \qquad \theta_2 = \frac{1}{2}\log(\sigma_\epsilon^2)$$

Likelihood using the Kalman filter

- Let $\mathrm{L}(Y_n;\theta)$ denote the log-likelihood function, that is the log of the joint density of the sample time series $\{y_1,\ldots,y_n\}$ as a function of the parameters θ.
- The log-likelihood can be evaluated by the prediction error decomposition:

$$\mathrm{L}(Y_n;\theta) = \log f(y_1,\ldots,y_n;\theta) = \sum_{t=1}^{n} \log f(y_t|Y_{t-1};\theta).$$

- The predictive density $f(y_t|Y_{t-1};\theta)$ is evaluated with the support of the **Kalman Filter**.

Proof

- In this case we assume **NORMALITY**.
  Recall that $F_t = \mathrm{Var}(y_t | Y_{t-1})$ and $v_t = y_t - Z_t\tilde{\alpha}_t$.
- Then we can substitute $\mathrm{N}(Z_t\tilde{\alpha}_t, F_t)$ for $f(y_t | Y_{t-1})$ and we get, apart from constant :

$$\log \mathrm{L}(Y_n) \quad = \quad -\frac{1}{2}\left(\sum_{t=1}^{n} \log |F_t| + \sum_{t=1}^{n} \nu_t' F_t^{-1} \nu_t\right).$$

The likelihood function can be maximized numerically by a quasi-Newton optimization routine.

Testing in unobserved components models

- Interest lies in testing the null hypothesis that a component is deterministic, against the alternative that it is stochastic (and nonstationary).
- Wrt. unit roots tests, the null and the alternative are reversed.

Testing level stationarity against a RW

$$\begin{array}{rcll} y_t & = & \mu_t + \epsilon_t, & \epsilon_t \sim \text{NID}(0, \sigma^2_\epsilon) \\ \mu_t & = & \mu_{t-1} + \eta_t, & \eta_t \sim \text{NID}(0, \sigma^2_\eta) \end{array} \qquad (24)$$

ϵ_t and η_t are independent.

$$\text{H}_0 : \sigma^2_\eta = 0 \quad \text{vs} \quad \text{H}_1 : \sigma^2_\eta > 0$$

Classical likelihood based testing procedures run into difficulties, as the variance parameter lies on the boundary of the parameter space under the null; uniformly most powerful invariant test do not exist, but only MPI against a specific alternative $\text{H}_1 : \sigma^2_\eta = a$, where a is a positive number.

The local linear model and the Leser-HP filter

In the local linear trend model (LLTM) the trend μ_t is an integrated random walk:

$$\begin{array}{rcl} y_t &=& \mu_t + \psi_t, \qquad\qquad \psi_t \sim \mathrm{NID}(0, \sigma_\psi^2), \quad t = 1, 2, \ldots, n, \\ \mu_t &=& \mu_{t-1} + \beta_{t-1} + \eta_t, \quad \eta_t \sim \mathrm{NID}(0, \sigma_\eta^2), \\ \beta_t &=& \beta_{t-1} + \zeta_t, \quad \zeta_t \sim \mathrm{NID}(0, \sigma_\zeta^2). \end{array} \tag{25}$$

It is assumed that the ψ_t, η_t and ζ_t are mutually and serially uncorrelated. For $\sigma_\zeta^2 = 0$ the trend reduces to a random walk with constant drift, whereas for $\sigma_\eta^2 = 0$ the trend is an integrated random walk ($\Delta^2 \mu_t = \zeta_{t-1}$).

- The above representation encompasses a deterministic linear trend, arising when both σ_η^2 and σ_ζ^2 are zero.
- The LLTM is the model for which the Leser filter is optimal (see Leser, 1961). The latter is derived as the minimiser, with respect to $\mu_t, t = 1, \ldots, n$, of the penalised least squares criterion:

$$PLS = \sum_{t=1}^{n}(y_t - \mu_t)^2 + \lambda \sum_{t=3}^{n}(\Delta^2 \mu_t)^2.$$

- The parameter λ governs the trade-off between fidelity and it is referred to as the *smoothness* or *roughness penalty* parameter.

- The solution arising for $\lambda = 1600$ is widely popular in the analysis of quarterly macroeconomic time series as the Hodrick-Prescott filter.
- We show in the paper that the Leser-HP filter is the optimal signal extraction filter for the LLTM with $\sigma_\eta^2 = 0$ and $\lambda = \sigma_\psi^2/\sigma_\zeta^2$. A consequence of this result is that the components can be efficiently computed using the Kalman filter and smoother.
- The equivalence $\lambda = \sigma_\psi^2/\sigma_\zeta^2$ makes clear that the roughness penalty measures the variability of the cyclical (noise) component relative to that of the trend disturbance, and regulates the smoothness of the long-term component. If $\sigma_\zeta^2 \to 0$, $\lambda \to \infty$, and the trend is a straight line.
- The Leser-HP detrended or cyclical component is the smoothed estimate of the component ψ_t in LLTM representation and, although the maintained representation for the deviations from the trend is a WN component, the filter has been one of the most widely employed tools in macroeconomics to extract a measure of the business cycle.

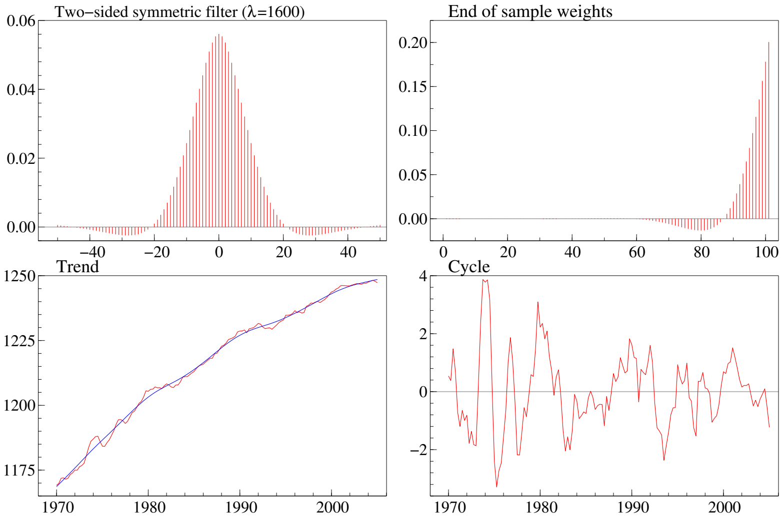

Figure: Signal extraction weights for the Leser HP filter $\lambda = 1600$ and application to Italian GDP

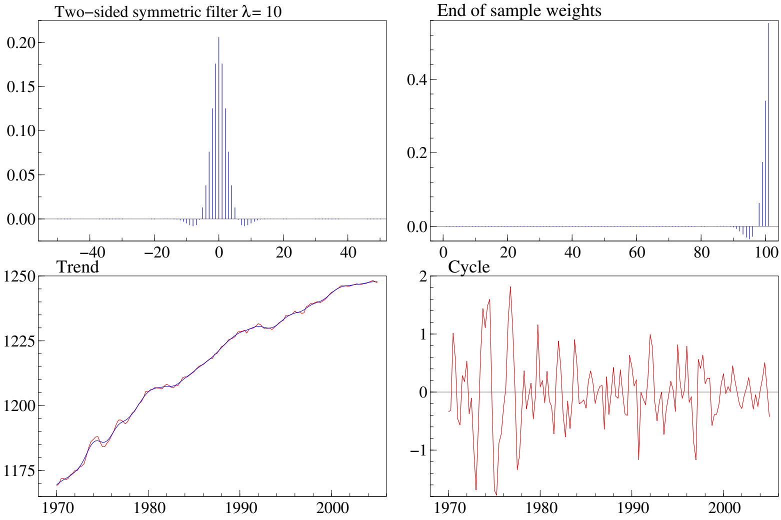

Figure: Signal extraction weights for the Leser HP filter $\lambda = 10$ and application to Italian GDP

- Problem: initialisation of the state vector when nonstationary state components are present ($\phi = 1$). This is discussed in de Jong (1988, 1991), Koopman (1997), Durbin and Koopman (2001).
- Illustration: LLM

$$
\begin{aligned}
y_t &= \mu_t + \epsilon_t \quad \epsilon_t \sim \text{NID}(0, \sigma^2_\epsilon), \\
\mu_{t+1} &= \mu_t + \eta_t, \quad \eta_t \sim \text{NID}(0, \sigma^2_\eta)
\end{aligned}
$$

Assumptions:

i Fixed initial conditions: the latent process has started at time $t = 1$ with μ_1 representing a fixed and unknown quantity.

ii Diffuse (random) initial conditions: the process has started in the remote past, $\tilde{\mu}_{1|0} = 0$, $P_{1|0} = \kappa, \kappa \to \infty$.

Preliminary

- We consider now the estimation of a bivariate model for the U.S. quarterly real GDP and the quarterly rate of inflation Δp_t, where p_t is the logarithm of the quarterly CPI for the U.S, using the data from the first quarter of 1950 to the fourth quarter of 2006.
- The KPSS test conducted on the inflation series leads to the rejection of the null of stationarity against a random walk for all the values of the lag truncation parameter up to 5; if a linear trend is considered and stationarity is tested against a random walk with drift, then the null is rejected also for much higher values of the lag truncation parameter.
- In the sequel, inflation will be taken to be integrated of order one.

Model

- The model has the following specification:
-

$$
\begin{array}{lcll}
y_t & = & \mu_t + \psi_t, & t = 1, \dots, n, \\
\mu_t & = & \mu_{t-1} + \beta_t + \eta_t, & \eta_t \sim \text{NID}(0, \sigma^2_\eta) \\
\psi_t & = & \phi_1 \psi_{t-1} + \phi_2 \psi_{t-2} + \kappa_t, & \kappa_t \sim \text{NID}(0, \sigma^2_\kappa) \\
 & & & \\
\Delta p_t & = & \tau_t + \varepsilon_{pt} & \varepsilon_{pt} \sim \text{NID}(0, \sigma^2_{p\varepsilon}) \\
\tau_t & = & \tau_{t-1} + \theta_\psi(L)\psi_t + \eta_{\tau t} & \eta_{\tau t} \sim \text{NID}(0, \sigma^2_{\tau\eta});
\end{array}
\tag{26}
$$

where η_t, κ_t, ε_{pt} are mutually independent.

- The output equation is the usual decomposition into orthogonal components.
- The inflation equation is a decomposition into a core component, τ_t, and a transitory one.
- The changes in the core component are driven by the output gap and by the idiosyncratic disturbances $\eta_{\tau t}$.
- The lag polynomial $\theta_\psi(L) = \theta_{\psi 0} + \theta_{\psi 1} L$ can be rewritten as $\theta_\psi(1) - \theta_{\psi 1}\Delta$, which enables to isolate the level effect of the gap from the change effect, which we expect to be positive, that is we expect $\theta_{\psi 1} < 0$. If $\theta_\psi(1) = 0$, the inflation equation can be rewritten $\Delta p_t = \tau_t^* - \theta_{\psi 1}\psi_t + \varepsilon_t$, with $\Delta \tau_t^* = \eta_{\tau t}$, so that output and inflation would share a common cycle.

Estimation

- The bivariate model and its GM extension(see below) under the normality assumption are estimated by maximum likelihood in the time domain. The likelihood is evaluated by the Kalman filter.

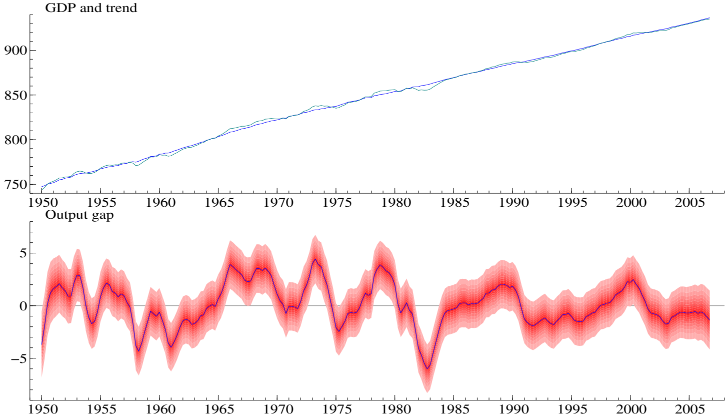

Figure: Bivariate model GDP and Inflation