2 Unobserved Components. State Space Models. Kalman Filter.

2.1 Unobserved Components

A time series model may sometimes be formulated in terms of components. These components are not observed directly, but are assumed to have ARMA representations. A simple model, consisting of an AR(1) embedded in white noise, is:

$$y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim \text{NID}(0, \sigma_\varepsilon^2), \quad t = 1, \dots, T$$
$$\mu_t = \phi\mu_{t-1} + \eta_t, \quad \eta_t \sim \text{NID}(0, \sigma_\eta^2), \quad |\phi| < 1$$

where η_t and ε_t are independent. When $|\phi| < 1$, the process is stationary, when $|\phi| = 1$ gives a random walk + noise. An important value $q = \frac{\sigma_\eta^2}{\sigma_\varepsilon^2}$ is the signal-to-noise ratio.

From lecture 1, we know the autocovariances of the AR(1) process:

$$\gamma_\mu(\tau) = \phi^\tau \gamma_\mu(0) = \phi^\tau \frac{\sigma_\eta^2}{1 - \phi^2}$$

and because $\gamma_\varepsilon(\tau) = 0$ for $\tau \neq 0$, we have:

$$\gamma_y(\tau) = \gamma_\mu(\tau) + \gamma_\varepsilon(\tau)$$
$$\Rightarrow \rho_y(\tau) = \frac{\gamma_y(\tau)}{\gamma_y(0)}$$
$$= \frac{\gamma_\mu(\tau)}{\gamma_\mu(0) + \gamma_\varepsilon(0)}$$
$$= \frac{\phi^\tau \sigma_\eta^2 (1 - \phi^2)^{-1}}{\sigma_\eta^2 (1 - \phi^2)^{-1} + \sigma_\varepsilon^2} \quad \tau = 1, 2, \dots$$

We can substitute ϕy_{t-1} to derive the reduced form:

$$y_t = \mu_t + \varepsilon_t$$
$$= \phi\mu_{t-1} + \eta_t + \varepsilon_t$$
$$= \phi(y_{t-1} - \varepsilon_{t-1}) + \eta_t + \varepsilon_t$$
$$y_t - \phi y_{t-1} = \eta_t + \varepsilon_t - \phi\varepsilon_{t-1}$$

This is equivalent to an ARMA(1,1) process:

$$y_t - \phi y_{t-1} = \nu_t + \theta\nu_{t-1}.$$

The MA parameter can be founded by equating the autocovariances of the right hand side of the two processes:

Reduced Form

$$\gamma_y(0) = Var(\eta_t) + Var(\varepsilon_t) + \phi^2 Var(\varepsilon_t)$$
$$= \sigma_\eta^2 + \sigma_\varepsilon^2(1 + \phi^2)$$
$$\gamma_y(1) = Cov(-\phi\varepsilon_{t-1}, \varepsilon_{t-1})$$
$$= -\phi\sigma_\varepsilon^2$$

ARMA(1,1)

$$\gamma_y(0) = Var(\nu_t) + \theta^2 Var(\nu_t)$$
$$= \sigma_\nu^2(1 + \theta^2)$$
$$\gamma_y(1) = Cov(\nu_t + \theta\nu_{t-1}, \nu_{t-1} + \theta\nu_{t-2})$$
$$= \theta\sigma_\nu^2$$

Equating the expression for $\gamma_y(1)$ gives:

$$\sigma_\nu^2 = -\frac{\phi \sigma_\varepsilon^2}{\theta}$$

Define the signal to noise ratio $q = \frac{\sigma_\eta^2}{\sigma_\varepsilon^2}$, then equate the expressions for $\gamma_y(0)$:

$$\sigma_\eta^2 + \sigma_\varepsilon^2(1 + \phi^2) = \sigma_\nu^2(1 + \theta^2)$$

$$\sigma_\eta^2 + \sigma_\varepsilon^2(1 + \phi^2) = -\frac{\phi \sigma_\varepsilon^2}{\theta}(1 + \theta^2)$$

$$\frac{\sigma_\eta^2}{\sigma_\varepsilon^2} + 1 + \phi^2 = -\frac{\phi}{\theta}(1 + \theta^2)$$

$$\Rightarrow 0 = \phi\theta^2 + (q + 1 + \phi^2)\theta + \phi$$

$$\Rightarrow \theta = \frac{-(q + 1 + \phi^2) \pm \sqrt{(q + 1 + \phi^2)^2 - 4\phi^2}}{2\phi}$$

Drop the negative root, since this implies a $|\theta| > 1$. Thus

$$\theta = -\frac{q + 1 + \phi^2}{2\phi} + \frac{\sqrt{(q + 1 + \phi^2)^2 - 4\phi^2}}{2\phi}$$

Thus $\phi \leq \theta \leq 0$, but when $\theta = -\phi$, which corresponds to $q = 0$, the ARMA(1,1) has a common factor and reduces to white noise. When $\phi = 1$

$$\theta = -\frac{q + 2}{2} + \frac{\sqrt{q^2 + 4q}}{2}$$

Alternatively we can write express q as a function of ϕ and θ:

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

2.2 Filtering and Prediction

Prediction requires filtering the observations so as to give the best estimate of μ_t at the end of the sample. This estimate is then extrapolated. The filtered estimate at the end of the sample is known as the **nowcast**.

Example (Random Walk + Noise). The conditional distribution of $y_{t+\ell}$ is obtained by writing:

$$y_{T+\ell} = \mu_T + \sum_{j=1}^{\ell} \eta T + j + \varepsilon_{T+\ell}$$

Thus the MMSE predictor is $\tilde{y}_{T+\ell|T} = \mathbb{E}(\mu_{T+\ell}|Y_t) = \mathbb{E}(\mu_T|Y_t)$, and so the forecast function is a horizontal straight line.
To find the MSE write:

$$y_{T+\ell} = \mu_T + \sum_{j=1}^{\ell} \eta T + j + \varepsilon_{T+\ell}$$

$$= \mathbb{E}(\mu_T|Y_t) + (\mu_T - \mathbb{E}(\mu_T|Y_t)) + \sum_{j=1}^{\ell} \eta T + j + \varepsilon_{T+\ell}$$

The conditional variance of $y_{T+\ell}$ is:

$$\sigma_{T+\ell|T}^2 = MSE(\tilde{y}_{T+\ell|T}) = Var(\mu_T|Y_t) + \ell\sigma_\eta^2 + \sigma_\varepsilon^2$$

> The MSE increases linearly with the forecast horizon, with $Var(\mu_T|Y_t)$ being the price paid for not knowing μ_T. Given σ_η^2 and σ_ε^2, a 95% prediction interval for $y_{T+\ell}$ is:
>
> $$\tilde{y}_{T+\ell|T} \pm 1.96\sigma_{T+\ell|T}$$

How do we filter for AR(1) + noise? Filtering is a two step process, firstly we obtain the prediction equations, then we update the predictions using new observations.

Prediction

Suppose we know the mean and variance of μ_{t-1} conditional on Y_{t-1}: $\mu_{t-1}|Y_{t-1} \sim N(\mu_{t-1|t-1}, p_{t-1|t-1})$. We can calculate the conditional mean and variance of μ_t as follows:

$$\begin{aligned}
\mu_{t|t-1} &= \mathbb{E}(\mu_t|Y_{t-1}) \\
&= \mathbb{E}(\phi\mu_{t-1} + \eta_t|Y_{t-1}) \\
&= \phi\mathbb{E}(\mu_{t-1}|Y_{t-1}) \\
&= \phi\mu_{t-1|t-1}
\end{aligned}$$

We can write $\mu_t = \phi\mu_{t-1|t-1} + \phi(\mu_{t-1} - \mu_{t-1|t-1}) + \eta_t$. Thus:

$$\mu_t - \mu_{t|t-1} = \mu_t - \phi\mu_{t-1|t-1} = \phi(\mu_{t-1} - \mu_{t-1|t-1}) + \eta_t$$

and we can derive the conditional variance as:

$$\begin{aligned}
p_{t|t-1} &= Var(\mu_t|Y_{t-1}) \\
&= \mathbb{E}[(\mu_t - \mu_{t|t-1})^2|Y_{t-1}] \\
&= \mathbb{E}[(\phi(\mu_{t-1} - \mu_{t-1|t-1}) + \eta_t)^2|Y_{t-1}] \\
&= \phi^2\mathbb{E}[(\mu_{t-1} - \mu_{t-1|t-1})^2|Y_{t-1}] + \sigma_\eta^2, \quad \text{since } \eta_t \perp Y_{t-1} \\
&= \phi^2 p_{t-1|t-1} + \sigma_\eta^2 \\
&\Rightarrow \mu_t|Y_{t-1} \sim N(\phi\mu_{t-1|t-1}, \phi^2 p_{t-1|t-1} + \sigma_\eta^2)
\end{aligned}$$

The conditional distribution of $y_t = \mu_t + \varepsilon_t$ is normal with mean and variance:

$$\begin{aligned}
\mathbb{E}(y_t|Y_{t-1}) &= \mathbb{E}(\mu_t|Y_{t-1}) + \mathbb{E}(\varepsilon_t|Y_{t-1}) \\
&= \phi\mu_{t-1|t-1} \\
Var(y_t|Y_{t-1}) &= \mathbb{E}[(y_t - \mu_{t|t-1})^2|Y_{t-1}] \\
&= \mathbb{E}[(\mu_t - \mu_{t|t-1})^2|Y_{t-1}] + \mathbb{E}[\varepsilon_t^2|Y_{t-1}] \\
&= \phi^2 p_{t-1|t-1} + \sigma_\eta^2 + \sigma_\varepsilon^2 \\
&\Rightarrow y_t|Y_{t-1} \sim N(\phi\mu_{t-1|t-1}, \phi^2 p_{t-1|t-1} + \sigma_\eta^2 + \sigma_\varepsilon^2)
\end{aligned}$$

Updating

Having obtained the prediction equations, the next step is to update the estimates of the mean and variance of μ_t to take account of the information in the new observation. I.e., we want to find $\mu_t|Y_t \sim N(\mu_{t|t}, p_{t|t})$.

The current information is contained in the conditional distribution of $(\mu_t, y_t)'$ which is bivariate normal:

$$\begin{aligned}
\mathbb{E}\mu_t|Y_{t-1} &= \mu_{t|t-1}, \quad \mathbb{E}y_t|Y_{t-1} = \mu_{t|t-1} \\
Var(\mu_t|Y_{t-1}) &= p_{t|t-1}, \quad Var(y_t|Y_{t-1}) = p_{t|t-1} + \sigma_\varepsilon^2 \\
Cov(\mu_t, y_t|Y_{t-1}) &= Cov(\mu_t, \mu_t + \varepsilon_t|Y_{t-1}) = p_{t|t-1}
\end{aligned}$$

$$\Rightarrow \begin{bmatrix} \mu_t \\ y_t \end{bmatrix}\bigg|Y_{t-1} \sim N\left(\begin{bmatrix} \mu_{t|t-1} \\ \mu_{t|t-1} \end{bmatrix}, \begin{bmatrix} p_{t|t-1} & p_{t|t-1} \\ p_{t|t-1} & p_{t|t-1} + \sigma_\varepsilon^2 \end{bmatrix}\right)$$

This updated distribution is the distribution of μ_t is the distribution of μ_t conditional on Y_{t-1}, and y_t. To calculate $\mu_{t|t}$ and $p_{t|t}$, we invoke the following lemma:

> **Lemma 2.2.1** (Marginal and conditional normal distributions).
>
> $$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$
>
> Consider two distributions:
>
> $$x_1 \sim N(\mu_1, \Sigma_{11}) \quad x_2 \sim N(\mu_2, \Sigma_{22})$$
>
> The conditional distribution is given by:
>
> $$x_1 | x_2 \sim N(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

Applying this lemma we can find the conditional distribution of μ_t given y_t:

$$\mu_{t|t} = \mu_{t|t-1} + \frac{Cov(\mu_t, y_t|Y_{t-1})}{Var(y_t|Y_{t-1})}(y_t - \mu_{t|t-1})$$

$$= \mu_{t|t-1} + \frac{p_{t|t-1}}{p_{t|t-1} + \sigma_\varepsilon^2}(y_t - \mu_{t|t-1})$$

$$p_{t|t} = p_{t|t-1} - \frac{Cov(\mu_t, y_t|Y_{t-1})^2}{Var(y_t|Y_{t-1})}$$

$$= p_{t|t-1} - \frac{p_{t|t-1}^2}{p_{t|t-1} + \sigma_\varepsilon^2}$$

The updating and prediction equations can thus be written together as a single set of recursions:

$$\mu_{t+1|t} = \phi\mu_{t|t} = \phi\mu_{t|t-1} + \phi\frac{p_{t|t-1}}{p_{t|t-1} + \sigma_\varepsilon^2}(y_t - \mu_{t|t-1})$$

$$p_{t+1|t} = \phi^2 p_{t|t} + \sigma_\eta^2 = \phi^2 p_{t|t-1} - \phi^2\frac{p_{t|t-1}^2}{p_{t|t-1} + \sigma_\varepsilon^2} + \sigma_\eta^2$$

These equations can be repeatedly applied as new observations become available. This is a special case of the Kalman filter (**???**). Indeed the first equation can be written as:

$$\mu_{t+1|t} = \phi\mu_{t|t-1} + k_t(y_t - \mu_{t|t-1}) \quad \text{where} \quad k_t = \phi\frac{p_{t|t-1}}{p_{t|t-1} + \sigma_\varepsilon^2}$$

is known as the **Kalman gain**.

How do we start the filter?

When $|\phi| < 1$, we can use the unconditional distribution of μ_0: $\mu_0 \sim N(0, \sigma_\eta^2/(1 - \phi^2))$ and so $\mu_{0|0} = 0$ and $p_{0|0} = \sigma_\eta^2/(1 - \phi^2)$.

When $|\phi| = 1$ we start with $\mu_1|y_1 \sim N(y_1, \sigma_\varepsilon^2)$. This is equivalent (**???**) to starting with $\mu_0 \sim N(0, \kappa)$ where $\kappa \to \infty$ (Diffuse prior). In both cases the filter converges to a steady state with $p_t = p > 0$.

> **Explanation.** Indeed we can actually set an arbitrary value for μ in $\mu_0 \sim N(\mu, \kappa)$, to see this we first set up the predictive equation for p_1:
>
> $$p_{t+1|t} = p_{t|t} + \sigma_\eta^2 \Rightarrow p_{1|0} = p_{0|0} + \sigma_\eta^2 = \kappa + \sigma_\eta^2$$

Now we update given y_1:

$$p_{1|1} = p_{1|0} - \frac{p_{1|0}^2}{p_{1|0} + \sigma_\varepsilon^2}$$

$$= \kappa + \sigma_\eta^2 - \frac{(\kappa + \sigma_\eta^2)^2}{\kappa + \sigma_\eta^2 + \sigma_\varepsilon^2}$$

$$= (\kappa + \sigma_\eta^2)\left(1 - \frac{\kappa + \sigma_\eta^2}{\kappa + \sigma_\eta^2 + \sigma_\varepsilon^2}\right)$$

$$= (\kappa + \sigma_\eta^2)\left(\frac{\sigma_\varepsilon^2}{\kappa + \sigma_\eta^2 + \sigma_\varepsilon^2}\right)$$

$$= \frac{\kappa + \sigma_\eta^2}{\kappa + \sigma_\eta^2 + \sigma_\varepsilon^2}\sigma_\varepsilon^2$$

$$\Rightarrow \lim_{\kappa \to \infty} p_{1|1} = \sigma_\varepsilon^2$$

Now we can update μ:

$$\mu_{1|1} = \mu1|0 + \frac{p_{1|0}}{p_{1|0} + \sigma_\varepsilon^2}(y_1 - \mu_{1|0})$$

$$= mu + \frac{\kappa + \sigma_\eta^2}{\kappa + \sigma_\eta^2 + \sigma_\varepsilon^2}(y_1 - \mu)$$

$$\lim_{\kappa \to \infty} \mu_{1|1} = \mu + y_1 - \mu = y_1$$

Since the conditional distribution is also normal, we get the distribution of $\mu_1|y_1 \sim N(y_1, \sigma_\varepsilon^2)$ as required. $\qquad\square$

For a random walk the filter is an exponentially weighted moving average:

Example $(\phi = 1)$. First recall the updating and prediction equation of the variance:

$$p_{t+1|t} = \phi^2 p_{t|t-1} - \phi^2 \frac{p_{t|t-1}^2}{p_{t|t-1} + \sigma_\varepsilon^2} + \sigma_\eta^2$$

Indeed setting $\phi = 1$ gives:

$$p_{t+1|t} = p_{t|t-1} - \frac{p_{t|t-1}^2}{p_{t|t-1} + \sigma_\varepsilon^2} + \sigma_\eta^2$$

$$\sigma_\varepsilon^{-2} p_{t+1|t} = \sigma_\varepsilon^{-2} p_{t|t-1} - \frac{\sigma_\varepsilon^{-2} p_{t|t-1}^2}{p_{t|t-1} + \sigma_\varepsilon^2} + \sigma_\varepsilon^{-2}\sigma_\eta^2$$

$$= \sigma_\varepsilon^{-2} p_{t|t-1} - \frac{(\sigma_\varepsilon^{-2})^2 p_{t|t-1}^2}{\sigma_\varepsilon^{-2} p_{t|t-1} + 1} + q \quad \text{since } q = \sigma_\varepsilon^{-2}\sigma_\eta^2$$

Re-define $p_{t|t-1} = \sigma_\varepsilon^{-2} p_{t|t-1}$ and we obtain the Ricatti equation:

$$p_{t+1|t} = p_{t|t-1} - \frac{p_{t|t-1}^2}{p_{t|t-1} + 1} + q$$

Setting $p_{t+1|t} = p_{t|t-1} = p$ as in a steady state gives the algebraic Riccati equation:

$$p = p - \frac{p^2}{p+1} + q = \frac{p}{p+1} + q$$

For a general ϕ we can write:

$$p = \phi^2 \left(p - \frac{p^2}{p+1} \right) + q$$

Solving the quadratic Ricatti equation:

$$\frac{p^2}{p+1} = q$$
$$\Rightarrow 0 = p^2 - qp - q$$
$$\Rightarrow p = \frac{q \pm \sqrt{q^2 + 4q}}{2}$$
$$\Rightarrow p = \frac{q + \sqrt{q^2 + 4q}}{2} \quad \text{since } p > 0$$

Recall the updating equation for the state (with $\phi = 1$), we can show that it takes the form of an EWNA:

$$\mu_{t+1|t} = \mu_{t|t-1} + \frac{p_{t|t-1}}{p_{t|t-1} + \sigma_\varepsilon^2} (y_t - \mu_{t|t-1})$$
$$:= \mu_{t|t-1} + \kappa(y_t - \mu_{t|t-1})$$
$$= (1 - \kappa)\mu_{t|t-1} + \kappa y_t$$

with

$$\kappa = \frac{p_{t|t-1}}{p_{t|t-1} + \sigma_\varepsilon^2}$$
$$= \frac{p}{p+1} \quad \text{with } p_{t|t-1} = \sigma_\varepsilon^{-2} p_{t|t-1} \text{ and } p_{t|t-1} = p$$

Recall

$$p = \frac{q + \sqrt{q^2 + 4q}}{2}, \quad \theta = -\frac{q+2}{2} + \frac{\sqrt{q^2 + 4q}}{2}$$

$$\Rightarrow \theta + 1 + q = p$$
$$= \frac{p}{p+1} + q$$
$$\Rightarrow \theta + 1 = \frac{p}{p+1}$$

Thus

$$\kappa = \frac{p}{p+1}$$
$$= \theta + 1$$
$$= -\frac{q}{2} + \frac{\sqrt{q^2 + 4q}}{2}$$

2.3 Innovations Form

In a signal plus noise model, the update in the Kalman Filter depends on the prediction errors or *innovations*,

$$\nu_t = y_t - \mu_{t|t-1}.$$

The innovations form of the AR(1) + noise model is:

$$y_t = \mu_{t|t-1} + \nu_t$$
$$\mu_{t+1|t} = \phi\mu_{t|t-1} + k_t\nu_t$$

This is different from the structural model since we are examining conditionals ($t|t-1$ not t subscript). Thus the dynamic equations (above) are a function of past observations. The Kalman gain k_t depends on ϕ and q. In the steady state, $k_t = \kappa$ is constant:

$$y_t = \mu_{t|t-1} + \nu_t$$
$$\mu_{t+1|t} = \phi\mu_{t|t-1} + \kappa\nu_t$$
$$\Rightarrow y_t = \phi\mu_{t-1|t-2} + \kappa\nu_{t-1} + \nu_t$$
$$= \phi(y_{t-1} - \nu_{t-1}) + \kappa\nu_{t-1} + \nu_t$$
$$= \phi y_{t-1} + (\kappa - \phi)\nu_{t-1} + \nu_t$$

I.e. we get an ARMA(1,1) with MA parameter $\theta = \kappa - \phi$.

2.4 ML Estimation

The joint density for the full set of T observations, $y_1, \ldots, y_T$ is:

$$p(\mathbf{y}; \mathbf{\Psi}) = \prod_{t=1}^{T} p(y_t|Y_{t-1}; \mathbf{\Psi})$$

where $\mathbf{\Psi}$ is the vector of parameters. In a Gaussian model:

$$y_t|Y_{t-1} \sim N(\tilde{y}_{t|t-1}, f_t)$$

where $\tilde{y}_{t|t-1} = \phi\mu_{t-1|t-1}$ and $f_t = p_{t|t-1} + \sigma_\varepsilon^2$. Thus the log-likelihood can be derived:

$$L(\phi, \sigma_\eta^2, \sigma_\varepsilon^2) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi f_t}} \exp\left(-\frac{(y_t - \tilde{y}_{t|t-1})^2}{2f_t}\right)$$
$$= \left(\frac{1}{\sqrt{2\pi}}\right)^T \left(\prod_{t=1}^{T} \frac{1}{\sqrt{f_t}}\right) \exp\left(-\sum_{t=1}^{T} \frac{(y_t - \tilde{y}_{t|t-1})^2}{2f_t}\right)$$
$$\ln L(\phi, \sigma_\eta^2, \sigma_\varepsilon^2) = -\frac{T}{2}\ln 2\pi - \frac{1}{2}\sum_{t=1}^{T}\ln f_t - \frac{1}{2}\sum_{t=1}^{T} \frac{(y_t - \tilde{y}_{t|t-1})^2}{f_t}$$

The prediction error $y_t - \tilde{y}_{t|t-1}$ is now a residual, whose interpretation is the same as $\varepsilon(\Psi)$ in an ARMA from lecture 1. When f_t is constant, as it is in the steady-state, maximising lnL is equivalent to minimising a sum of squares function, again as in the ARMA.

Diagnostics and testing with the residuals

It follows from $y_t|Y_{t-1} \sim N(\tilde{y}_{t|t-1}, f_t)$ that the prediction errors ν_t are normally distributed and serially independent with mean zero and variance f_t. When a model has been estimated, diagnostic checking can be based on the standardised residuals, $\nu_t/\sqrt{f_t}$.

However, unless q and ϕ are known, they will not be NID. There are two options here, carry out

a modified Box-Pierce test based on χ^2_{P-2} or do an LM test.

Tests on variance parameters

The null hypothesis that $\sigma^2_\varepsilon = 0$ can be tested against $\sigma^2_\varepsilon > 0$ using a likelihood ratio test, the distribution is a mixture of of χ^2's:

$$LR \sim \frac{1}{2}\chi^2_0 + \frac{1}{2}\chi^2_1$$

The χ^2_0 is a degenerate distribution with all it's mass at the origin, thus the size of the test can be set to α by using the 2α quantile of the χ^2_1 distribution. When $\phi = 1$ (random walk) this test is still valid since the model is stationary and invertible in first differences.

The null hypothesis that $\sigma^2_\eta = 0$ vs $\sigma^2_\eta > 0$ does not produce a LR statistic with the previous distribution, because when the null is true, the first differenced observations $\Delta y_t = \Delta \varepsilon_t$ are strictly non-invertible. Such non-standard behaviour is a feature of the situations when the model is stationary under the null, but non-stationary under the alternative.

> **Note:-**
>
> A test of the hypothesis that $\phi = 1$ against the alternative $\phi < 1$ encounters similar problems. Indeed the LR test of $\sigma^2_\eta = 0$ cannot be implemented in the stationary AR(1) + noise since ϕ is not identified when $\sigma^2_\eta = 0$.

2.5 Signal Extraction

Signal extraction is concerned with finding the optimal estimate or estimator of an unobserved component at some point within the sample. Whilst filtering aims to find the expected value of the signal conditional on information at time t, smoothing seeks to account for the information made available after t.

In the AR(1) + noise model, the smoothed estimate of the component of interest μ_t, is its expectation conditional on the full set of observations. When the model is Gaussian, the estimate is liner so:

$$\tilde{\mu}_{t|T} = \mathbb{E}(\mu_t|Y_T) = \sum_{j=t-T}^{t-1} w_{j,t} y_{t-j}$$

where the w's are non-random weights.

2.5.1 Wiener-Kolmogorov (WK) filter

Given an UC model, the WK formulae give expressions for the weights w_j. The assumption of a doubly infinite series is made, meaning there are an infinite number of observations on either side of the point of interest. The weights are constant, and so:

$$\mu_{t|\infty} = \sum_{j=-\infty}^{\infty} w_j y_{t-j} = w(L)y_t$$

where $w(L)$ is a symmetric polynomial in the lag operator. The WK formula for finding the weights in $y_t = \mu_t + \xi_t$ where μ_t and ξ_t are mutually uncorrelated ARMA models is:

$$w(L) = \frac{g_\mu(L)}{g_y(L)} = \frac{g_\mu(L)}{g_\mu(L) + g_\xi(L)}$$

where $g_\mu(L)$ and $g_\xi(L)$ are the ACGF's of μ_t and ξ_t respectively. The ACGF of y is $g_\mu(L) + g_xi(L)$.

The ACGF is defined as a polynomial in the lag operator, g(L), such that:

$$g(L) = \sum_{\tau=-\infty}^{\infty} \gamma(\tau) L^{\tau}$$

The coefficient of L^{τ} gives the autocovariance at lag τ.

Recall the ACGF of an ARMA model $y_t = \phi^{-1}(L)\theta(L)\nu_t$ where $\nu_t \sim WN(0, \sigma_\nu^2)$ is:

$$g(L) = \frac{\theta(L)\theta(L^{-1})}{\phi(L)\phi(L^{-1})}\sigma_\nu^2$$

Example. MA(1): $y_t = \nu_t + \theta\nu_{t-1}$

$$
\begin{aligned}
g(L) &= \sum_{\tau=-\infty}^{\infty} \gamma(\tau) L^{\tau} \\
&= \gamma(0) + \gamma(1)L + \gamma(-1)L^{-1} \\
&= (1 + \theta^2)\sigma_\nu^2 + \theta\sigma_\nu^2 L + \theta\sigma_\nu^2 L^{-1}
\end{aligned}
$$

In the AR(1) + noise model, the reduced form is ARMA(1,1) with MA parameter

$$\theta = -\frac{q + 1 + \phi}{2\phi} + \frac{\sqrt{(q + 1 + \phi^2)^2 - 4\phi^2}}{2\phi}$$

Provided q¿0, the WK formula gives:

$$\mu_{t|\infty} = \frac{\frac{\sigma_\eta^2}{(1-\phi L)^2}}{\sigma_\varepsilon^2 + \frac{\sigma_\eta^2}{(1-\phi L)^2}} y_t = \frac{\sigma_\eta^2}{\sigma_\nu^2(1 + \theta L)^2} y_t$$

The second equality holds since the reduced form is equated to an ARMA(1,1): $\eta_t + \varepsilon_t - \phi\varepsilon_{t-1} = \nu_t + \theta\nu_{t-1}$. Further, from equating the first autocovariances of the reduced form and the ARMA(1,1) as at the start of the lecture, we get:

$$\sigma_\nu^2 = -\frac{\phi\sigma_\varepsilon^2}{\theta} \Rightarrow \mu_{t|\infty} = -\frac{q\theta}{\phi}\frac{1}{(1 + \theta L)^2}y_t$$

On recognising that the second term is the ACGF of an AR(1) model with parameter $-\theta$, it becomes apparent that the weights decline symmetrically and exponentially.

$$w_j = -\frac{q\theta/\phi}{1 - \theta^2}(-\theta)^{|j|}$$

By setting $L = 1$[1] it can be seen that the weights sum to:

$$\sum_{j=-\infty}^{\infty} w_j = -\frac{q\theta/\phi}{1-\theta^2} \sum_{j=-\infty}^{\infty} (-\theta)^{|j|}$$

$$= -\frac{q\theta/\phi}{1-\theta^2} \left(2\sum_{j=0}^{\infty} \theta^j - 1\right)$$

$$= -\frac{q\theta/\phi}{1-\theta^2} \left(\frac{2}{1-\theta} - 1\right)$$

$$= -\frac{q\theta/\phi}{1-\theta^2} \frac{1-\theta}{1+\theta}$$

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

Recall from 50 fucking slides ago:

$$q = -\frac{(\phi+\theta)(1+\phi\theta)}{\theta}$$

Thus

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

$$= \frac{(\phi+\theta)(1+\phi\theta)}{\phi(1+\theta)^2}$$

In the case of a random walk ($\phi = 1$) this simplifies to unity. The lower q, the closer θ is to one, and the more spread out the weights. The MSE is given by

$$MSE(\mu_{t|\infty}) = \sigma_\varepsilon^2 \frac{1+\theta}{1-\theta}$$

The weights for smoothing near the end of a semi-infinite sample can be found by modifying the signal extraction formulae:

$$w_j = \frac{\phi+\theta}{\phi(1-\theta^2)}[(1+\phi\theta)(-\theta)^{|-j|} + (\phi+\theta)(\theta)^{j+2(T-t)+1}], \quad j = t - T, \ldots, 0, \ldots, \infty$$

The smoothing weights are obtained when $t << T$. On the other hand setting t=T gives the weights for the filtered estimator of μ_T as

$$w_j = \frac{(\phi+\theta)[(1+\phi\theta)(-\theta)^j + (\phi+\theta)(\theta)^{j+1}]}{\phi(1-\theta^2)}$$

$$= \frac{(\phi+\theta)(-\theta)^j}{\phi(1-\theta^2)}[(1+\phi\theta) - (\phi+\theta)\theta]$$

$$= \frac{\phi+\theta}{\phi} \frac{1-\theta^2}{1-\theta^2}(-\theta)^j$$

$$= \frac{\phi+\theta}{\phi}(-\theta)^j, \quad j = 0, 1, 2, \ldots$$

The weights decline exponentially when $\phi = 1$:

$$w_j = (1+\theta)(-\theta)^j, \quad j = 0, 1, 2, \ldots$$

as in the classic EWMA.

---

[1]Literally how does the lag operator even work, why can you set it to 1??? It's an OPERATOR wtf is going on

2.5.2 Auxiliary Residuals

Not covered in lectures.

2.6 State Space form

We derived the filter for an AR(1) dynamic equation in the previous section, but what if μ_t is AR(2)? It turns out this is not an issue, since an AR(2) can be expressed in first-order Markov form (a multivariate AR(1)) by writing

$$\begin{bmatrix} \mu_t \\ \mu_{t-1} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \mu_{t-2} \end{bmatrix} + \begin{bmatrix} \eta_t \\ 0 \end{bmatrix}$$

This is known as the transition equation. We can extract the first element as

$$y_t = \mu_t + \varepsilon_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \mu_t \\ \mu_{t-1} \end{bmatrix} + \varepsilon_t$$

known as the measurement equation. The role of the second element is simply to store the information in the second lag. We can generalise this further and write the state space form. The observation is related to an $m \times 1$ vector $alpha_t$, the state vector, through a measurement equation

$$y_t = z_t \alpha_t + d_t + \varepsilon_t$$

where z_t is a $1 \times m$ vector, and d_t is deterministic. The elements of α_t are generated by a first order Markov process, the transition equation:

$$\alpha_t = T_t \alpha_{t-1} + \eta_t$$

where T_t is an $m \times m$ matrix and η_t is a $m \times 1$ vector of serially uncorrelated Gaussian disturbances with $\mathbb{E}\eta_t = 0$ and $Var(\eta_t) = Q_t$ (i.e. multivariate WN). The disturbances η_t and ε_t may be contemporaneously correlated, but $\mathbb{E}\varepsilon_t' \eta_s = 0$ for all $t \neq s$.

The specification of the state space system is completed by assuming that $\mathbb{E}\alpha_0 = \alpha_{0|0}$ and $Var(\alpha_0) = P_{0|0}$ are known and $P_{0|0}$ is positive definite. Further $\mathbb{E}\eta_t \alpha_0' = 0$ and $\mathbb{E}\varepsilon_t \alpha_0' = 0$ for all t. All system matrices and vectors (z_t, d_t, T_t, Q_t) are non-stochastic.

Models can also be set up with what is called the future SSF, where the transition equation is

$$\alpha_{t+1} = T_t \alpha_t + \eta_t$$

2.6.1 ARMA models

Any ARMA(p,q) model can be written in state space form. A transition equation can be set up as follows:

$$\alpha_t = \begin{bmatrix} \phi_1 & 1 & 0 & \cdots & 0 \\ \phi_2 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \vdots & 0 & \cdots & 0 & 1 \\ \phi_m & 0 & 0 & \ldots & 1 \end{bmatrix} \alpha_{t-1} + \begin{bmatrix} 1 \\ \theta_1 \\ \vdots \\ \theta_{m-1} \end{bmatrix} \varepsilon_t$$

where m = max(p,q+1). The original ARMA can be recovered by noting that the first element of α_t is equal to y_t. The role of the measurement equation is simply to extract the first element of the state vector α_t. Thus

$$y_t = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} \alpha_t$$

There is no disturbance term.

> **Example.** The MA(1) model, $y_t = \mu_t + \xi_t + \theta\xi_{t-1}$ can be written in state space form as:
>
> $$y_t = \begin{bmatrix} 1 & 0 \end{bmatrix}\alpha_t + d_t$$
>
> where $d_t = \mu$ and
>
> $$\alpha_t = \begin{bmatrix} y_t \\ \theta\xi_t \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\alpha_{t-1} + \begin{bmatrix} 1 \\ \theta \end{bmatrix}\xi_t$$
>
> The future SSF is:
>
> $$y_t = \mu + \theta\alpha_t + \xi_t, \quad \alpha_{t+1} = \xi_t$$

2.6.2 Two AR1 components

$$y_t = \mu_{1,t} + \mu_{2,t} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$
$$\mu_i, t = \phi_i\mu_{i,t-1} + \eta_{i,t}, \quad \eta_{i,t} \sim N(0, \sigma_i^2), \quad i = 1,2$$

where $|\phi_i| < 1$, $i = 1, 2$. Using stationarity in the AR(1)'s:

$$Var(\mu_{i,t}) = \phi_i^2 Var(\mu_{i,t}) + \sigma_i^2$$
$$\Rightarrow Var(\mu_{i,t}) = (1 - \phi_i^2)^{-1}\sigma_i^2, \quad i = 1,2$$
$$\Rightarrow Var(y_t) = Var(\mu_{1,t}) + Var(\mu_{2,t}) + \sigma_1^2$$
$$= (1 - \phi_1^2)^{-1}\sigma_1^2 + (1 - \phi_2^2)^{-1}\sigma_2^2 + \sigma_\varepsilon^2$$

Iterating backwards from t:

$$y_t = \mu_{1,t} + \mu_{2,t} + \varepsilon_t$$
$$= \phi_1\mu_{1,t-1} + \phi_2\mu_{2,t-1} + \eta_{1,t} + \eta_{2,t} + \varepsilon_t$$
$$= \ldots$$
$$= \phi_1^\tau\mu_{1,t-\tau} + \phi_2^\tau\mu_{2,t-\tau} + \sum_{i=1}^{\tau} \phi_1^{i-1}\eta_{1,t-i+1} + \sum_{i=1}^{\tau} \phi_2^{i-1}\eta_{2,t-i+1} + \varepsilon_t$$

$$\Rightarrow \gamma(\tau) = Cov\left(\phi_1^\tau\mu_{1,t-\tau} + \phi_2^\tau\mu_{2,t-\tau} + \sum_{i=1}^{\tau} \phi_1^{i-1}\eta_{1,t-i+1} + \sum_{i=1}^{\tau} \phi_2^{i-1}\eta_{2,t-i+1}, \mu_{1,t-\tau} + \mu_{2,t-\tau} + \varepsilon_{t-\tau}\right)$$
$$= \phi_1^\tau Var(\mu_{1,t-\tau}) + \phi_2^\tau Var(\mu_{2,t-\tau})$$
$$= \phi_1^\tau(1 - \phi_1^2)^{-1}\sigma_1^2 + \phi_2^\tau(1 - \phi_2^2)^{-1}\sigma_2^2$$
$$\Rightarrow \rho(\tau) = \frac{\phi_1^\tau(1 - \phi_1^2)^{-1}\sigma_1^2 + \phi_2^\tau(1 - \phi_2^2)^{-1}\sigma_2^2}{(1 - \phi_1^2)^{-1}\sigma_1^2 + (1 - \phi_2^2)^{-1}\sigma_2^2 + \sigma_\varepsilon^2} \quad \tau = 1, 2, \ldots$$

The SSF is:

$$y_t = \mu_{1,t} + \mu_{2,t} + \varepsilon_t = z\alpha_t + \varepsilon_t$$

where $z = \begin{bmatrix} 1 & 1 \end{bmatrix}$ and

$$\alpha_t = \begin{bmatrix} \mu_{1,t} \\ \mu_{2,t} \end{bmatrix} = \begin{bmatrix} \phi_1 & 0 \\ 0 & \phi_2 \end{bmatrix}\alpha_{t-1} + \begin{bmatrix} \eta_{1,t} \\ \eta_{2,t} \end{bmatrix}, \quad Q = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

The initial conditions are $\alpha_{1|0} = 0$ and

$$P_{1|0} = \begin{bmatrix} \frac{\sigma_1^2}{1-\phi_1^2} & 0 \\ 0 & \frac{\sigma_2^2}{1-\phi_2^2} \end{bmatrix}$$

The KF and associated smoother deliver estimates of $\mu_{1,t}$ and $\mu_{2,t}$ and their respective conditional variances. Using the WK filter for this task is much more difficult.

Long memory and unobserved components

The characteristics of long memory can often be approximated by the sum of two stationary AR(1)'s.

2.7 Missing observations and Data Irregularities

Missing observations pose no problems when handled in the state space framework. If an observation is missing at a particular point in time, $t = \tau$, the Kalman Filter updating equations are redundant. Thus for the AR(1) + noise model, the predicted state at time τ is $\mu_{\tau|\tau-1}$ with MSE $p_{\tau|\tau-1}$. If y_τ is missing, then the update is skipped, or we set $\mu_{\tau|\tau} = \mu_{\tau|\tau-1}$ and $p_{\tau+1|\tau-1} = \phi^2 p_{\tau+1|\tau-1}$, and the prediction equations are applied to give the two-step ahead predictor $\mu_{\tau+1|\tau} = \phi \mu_{\tau|\tau-1}$ and $p_{\tau+1|\tau} = \phi^2 p_{\tau|\tau-1} + \sigma_\eta^2$. If $y_{\tau+1}$ is missing, the prediction equations are just applied again.

The likelihood function is computed in the usual way, by using the innovations from the Kalman filter, except there is no innovation corresponding to the missing observation. This generalises for any linear Gaussian state space model irrespective of the number of missing observations or where they're located. E.G. if m consecutive observations are missing, the mean and variance of an $m+1$ step ahead predictive distribution will enter the likelihood function.

Data Irregularities - Suppose a survey becomes quarterly instead of annually, we can just construct a quarterly model with missing observations for the first part of the series.