

4 Trends, Cycles and Seasonality

4.1 Deterministic trends

A deterministic trend model takes the form

$$y_t = \mu(t) + \xi_t$$

where ξ_t is a stationary process (EG some ARMA) and $\mu(t)$ is a deterministic function of time. This function can be modelled as a polynomial in time, where $\mu(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \dots + \beta_h t^h$. The linear trend model $\mu(t) = \beta_0 + \beta_1 t$ is obtained by setting $h = 1$.

The drawback with deterministic time trend models is that a time series rarely keeps on the same path indefinitely. Structural time series models achieve this aim directly by allowing the trend parameters to evolve over time as stochastic processes.

4.2 ARIMA models

ARMA(p, q) models may be fitted to nonstationary series by differencing the observations. When forecasting is carried out, the differencing must be reversed; Box and Jenkins (1976) called this operation integration. If differencing d times gives a stationary and invertible model, then the original series is said to be *integrated of order d* , or $y_t \sim I(d)$.

Definition 4.2.1

An auto-regressive integrated moving average (ARIMA) model of order (p, d, q) is defined as:

$$\phi(L)\Delta^d y_t = \theta(L)\varepsilon_t \quad (4.1)$$

and denoted as $ARIMA(p, d, q)$.

Explanation. *When a series has been overdifferenced, it will be stationary, but strictly non-invertible due to the presence of a unit root in the MA polynomial*
Suppose we have a random walk model:

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

where $\varepsilon_t \sim WN(0, \sigma^2)$. Clearly the first difference is stationary:

$$\Delta y_t = \varepsilon_t$$

however suppose we over-difference:

$$\Delta^2 y_t = \Delta y_t - \Delta y_{t-1} = \varepsilon_t - \varepsilon_{t-1}.$$

We have introduced a unit root in the MA polynomial: $\Theta(L) = 1 - L$, and the series is no longer invertible. \square

4.2.1 Prediction

Prediction from stationary models tend towards the mean as the lead time, ℓ , increases. When ℓ is large, the dynamic structure of the model is irrelevant. This is no longer the case with integrated

models, where the forecast function contains a deterministic component. The term $\phi(L)\Delta^d y_t$ in (4.1) is expanded to give an AR polynomial of order $p + d$:

$$\phi(L)\Delta^d = \varphi(L) = 1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_{p+d} L^{p+d}$$

and predictions are made from the recursive equation

$$\hat{y}_{T+\ell|T} = \varphi_1 \hat{y}_{T+\ell-1|T} + \dots + \varphi_{p+d} \hat{y}_{T+\ell-p-d|T} + \hat{\varepsilon}_{T+\ell|T} + \dots + \theta_q \hat{\varepsilon}_{T+\ell-q|T} \quad \ell = 1, 2, \dots$$

The eventual forecast function (EG for stationary processes forecasts converge to the mean) is a polynomial of order $d - 1$. When the model includes a constant it is a polynomial of order d .

Example (Prediction in ARIMA(0,1,1)).

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

$$\begin{cases} \hat{y}_{T+1|T} = y_T + \theta \varepsilon_T \\ \hat{y}_{T+\ell|T} = \hat{y}_{T+\ell-1|T} \end{cases}$$

Thus for all lead times, the forecasts made at time T follow a horizontal line. The disturbance term can be obtained from the recursive equation:

$$\varepsilon_t = y_t - y_{t-1} - \theta \varepsilon_{t-1}$$

with $\varepsilon_1 = 0$. Substitute repeatedly for lagged values of ε_t :

$$\varepsilon_2 = y_2 - y_1 - \theta \varepsilon_1 = y_2 - y_1$$

$$\varepsilon_3 = y_3 - y_2 - \theta(y_2 - y_1) = y_3 - (1 + \theta)y_2 + \theta y_1$$

$$\varepsilon_4 = y_4 - y_3 - \theta(y_3 - (1 + \theta)y_2 + \theta y_1) = y_4 - (1 + \theta)y_3 + \theta(1 + \theta)y_2 - \theta^2 y_1$$

\vdots

$$\varepsilon_T = y_T - (1 + \theta) \sum_{j=1}^{T-1} (-\theta)^{j-1} y_{T-j}$$

So the expression for the one-step ahead predictor can be written as:

$$\begin{aligned} \hat{y}_{T+1|T} &= y_T + \theta \varepsilon_T \\ &= y_T + \theta \left(y_T - (1 + \theta) \sum_{j=1}^{T-1} (-\theta)^{j-1} y_{T-j} \right) \\ &= (1 + \theta) \sum_{j=0}^{T-1} (-\theta)^j y_{T-j} \end{aligned}$$

This predictor is an EWMA, with weights that decrease geometrically with time.

Example (Prediction in ARIMA(0,1,1) + constant). When the model above includes a constant:

$$\Delta y_t = \theta_0 + \varepsilon_t + \theta \varepsilon_{t-1}$$

the forecasts are

$$\hat{y}_{T+1|T} = y_T + \theta_0 + \theta \varepsilon_T$$

and

$$\hat{y}_{T+\ell|T} = y_T + \ell\theta_0 + \theta\varepsilon_T$$

The eventual forecast function is therefore a linear time trend.

Example (ARIMA(1,1,0)).

$$\varphi(L) = (1 - \phi L)(1 - L) = 1 - (1 + \phi)L + \phi L^2$$

and forecasts are constructed from the difference equation:

$$\hat{y}_{T+\ell|T} = (1 + \phi)\hat{y}_{T+\ell-1|T} - \phi\hat{y}_{T+\ell-2|T}$$

The corresponding forecast function:

$$\hat{y}_{T+\ell|T} = y_T + (y_T - y_{T-1})\frac{\phi(1 - \phi^\ell)}{1 - \phi} \quad \ell = 1, 2, \dots$$

As $\ell \rightarrow \infty$, $\hat{y}_{T+\ell|T} \rightarrow y_T + (y_T - y_{T-1})\frac{\phi}{1 - \phi}$, a horizontal line. In contrast to the *ARIMA*(0, 1, 1) model where the eventual forecast function is also horizontal, but depends on all past values, the eventual forecast function here depends only on the last two observations.

Note:-

Model Selection: First plot the series in levels and differences, judge which show trending movements. Checking the correlogram is also useful, for stationary processes the autocorrelations tend towards zero as the lag increases, whilst non-stationary processes do not damp so quickly. Keep differencing the model until a correlogram having the characteristics of a stationary model is obtained.

4.3 Structural Time Series Models

4.3.1 Local level model

Definition 4.3.1: Local level model

The local level model consists of a random walk plus noise:

$$\begin{cases} y_t = \mu_t + \varepsilon_t & \varepsilon_t \sim NID(0, \sigma_\varepsilon^2) \\ \mu_t = \mu_{t-1} + \eta_t & \eta_t \sim NID(0, \sigma_\eta^2) \end{cases}$$

where $\mathbb{E}(\varepsilon_t \eta_t) = 0$.

When σ_η^2 is zero the level is constant and the process is just noise around a mean. The signal noise ratio $q = \sigma_\eta^2 / \sigma_\varepsilon^2$ plays a key role in determining how observations should be weighted for prediction and signal extraction. The higher q , the more past observations are discounted in forecasting. By taking first differences of y_t :

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

we can see that the reduced form is an *ARIMA*(0,1,1). Setting the autocovariances of the first

differences equal to those of an MA(1) model $[y_t = \xi_t + \theta_{t-1}]$ yields:

$$\begin{aligned}
\gamma_1 &= -\sigma_\varepsilon^2 := \theta\sigma_\xi^2 \\
\Rightarrow \sigma_\xi^2 &= -\frac{\sigma_\varepsilon^2}{\theta} \\
\gamma_0 &= \sigma_\eta^2 + 2\sigma_\varepsilon^2 := (1 + \theta^2)\sigma_\xi^2 \\
\Rightarrow \sigma_\eta^2 + \sigma_\varepsilon^2 &= (1 + \theta^2) \left(-\frac{\sigma_\varepsilon^2}{\theta} \right) \\
\Rightarrow q + 2 &= -\frac{1 + \theta^2}{\theta} & \Rightarrow \theta &= \frac{-q - 2 \pm \sqrt{q^2 + 4q}}{2}
\end{aligned}$$

4.3.2 Local linear trend model

A linear time trend of the form $\mu(t) = \beta_0 + \beta_1 t$ can be constructed recursively from $\mu_t = \mu_{t-1} + \beta_1$ with $\beta_0 = \mu_0$. It can then be made stochastic by adding a disturbance to the level μ_t (random walk + drift), or the slope itself can be made stochastic - resulting in the local linear trend model:

Definition 4.3.2: Local linear trend model

$$\begin{aligned}
y_t &= \mu_t + \varepsilon_t \quad \varepsilon_t \sim NID(0, \sigma_\varepsilon^2) \\
\mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t \quad \eta_t \sim NID(0, \sigma_\eta^2) \\
\beta_t &= \beta_{t-1} + \zeta_t \quad \zeta_t \sim NID(0, \sigma_\zeta^2)
\end{aligned}$$

where all disturbances are mutually independent.

Only when $\sigma_\zeta^2 = 0$ the slope is fixed and the trend reduces to random walk plus drift. Allowing $\sigma_\zeta^2 > 0$, but setting $\sigma_\eta^2 = 0$ makes μ_t an integrated random walk.

The SSF of the local linear trend model is:

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

where the state vector is $\alpha_t = \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix}$ and covariance matrix is $Q = \begin{bmatrix} \sigma_\eta^2 & 0 \\ 0 & \sigma_\zeta^2 \end{bmatrix}$.

Claim 4.3.1. The reduced form of the local linear trend model is an ARIMA(0,2,2) model.

Proof. Begin with the trend, and difference until we find a stationary expression for y_t :

$$\begin{aligned}
\beta_t &= \beta_{t-1} + \zeta_t \\
\Rightarrow \Delta\beta_t &= \zeta_t \\
\mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t \\
\Rightarrow \Delta\mu_t &= \beta_{t-1} + \eta_t \\
\Rightarrow \Delta^2\mu_t &= \Delta\beta_{t-1} + \eta_t + \eta_{t-1} = \zeta_{t-1} + \eta_t + \eta_{t-1} \\
y_t &= \mu_t + \varepsilon_t \\
\Rightarrow \Delta^2y_t &= \Delta^2\mu_t + (1 - L)^2\varepsilon_t = \zeta_{t-1} + \eta_t + \eta_{t-1} + \varepsilon_t - 2\varepsilon_{t-1} + \varepsilon_{t-2}
\end{aligned}$$

The ACF is thus:

$$\begin{aligned}\gamma(0) &= \sigma_\zeta^2 + 2\sigma_\eta^2 + 6\sigma_\varepsilon^2 \\ \gamma(1) &= -\sigma_\eta^2 - 4\sigma_\varepsilon^2 \\ \gamma(2) &= \sigma_\varepsilon^2 \\ \gamma(\tau) &= 0 \quad \tau > 2\end{aligned}$$

Which we could match to the ACF of an ARIMA(0,2,2) model and solve for the parameters. \square

Question 1

Show that the reduced form of the modified damped trend model with a mean is ARIMA(1,1,2):

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

In other words the slope is an AR(1) with mean β .

Solution:-

$$\begin{aligned}\beta_t &= (1 - \rho)\beta + \rho\beta_{t-1} + \zeta_t \\ \Rightarrow (1 - \rho L)\beta_t &= (1 - \rho)\beta + \zeta_t \\ \Rightarrow \beta_t &= \beta + \frac{\zeta_t}{1 - \rho L} \quad \text{since } \beta \text{ constant} \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t \\ \Rightarrow \Delta\mu_t &= \beta_{t-1} + \eta_t \\ &= \beta + \frac{\zeta_{t-1}}{1 - \rho L} + \eta_t \\ y_t &= \mu_t + \varepsilon_t \\ \Rightarrow \Delta y_t &= \Delta\mu_t + \varepsilon_t - \varepsilon_{t-1} \\ &= \beta + \frac{\zeta_{t-1}}{1 - \rho L} + \eta_t + \varepsilon_t - \varepsilon_{t-1} \\ \Rightarrow (1 - \rho L)\Delta y_t &= (1 - \rho)\beta + \zeta_{t-1} + (1 - \rho L)(\eta_t + \varepsilon_t - \varepsilon_{t-1})\end{aligned}$$

The left hand side is an AR(1) on the differenced series, and the RHS is a mean + an MA(2) component. Thus it is an ARIMA(1,1,2) model.

4.4 Trend-cycle models

We consider stochastic cycles with frequency λ_c and damping term ρ .

Definition 4.4.1: Stochastic cycle

The stochastic cycle model is given by:

$$\begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} = \rho \begin{bmatrix} \cos \lambda_c & \sin \lambda_c \\ -\sin \lambda_c & \cos \lambda_c \end{bmatrix} \begin{bmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix} + \begin{bmatrix} \kappa_t \\ \kappa_t^* \end{bmatrix} \quad (4.2)$$

where κ_t is a white noise process with the same variance as κ_t^* and $0 \leq \rho \leq 1$.

Observe that the model collapses to an AR(1) model when the cycle has frequency zero or π (giving positive and negative AR coefficients respectively¹).

Definition 4.4.2: Trend-cycle models

The trend plus cycle plus noise model is:

$$\begin{aligned} y_t &= \mu_t + \psi_t + \varepsilon_t \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t \\ \beta_t &= \beta_{t-1} + \zeta_t \\ \psi_t &= \rho \cos \lambda_c \psi_{t-1} + \rho \sin \lambda_c \psi_{t-1}^* + \kappa_t \end{aligned}$$

Claim 4.4.1. The reduced form of the trend-cycle model is an ARIMA(2,2,4) model.

Proof. Recall that the reduced form of the local linear trend model is an ARIMA(0,2,2), thus the second difference of y_t has the form:

$$\Delta^2 y_t = \zeta_{t-1} + \eta_t + \eta_{t-1} + \varepsilon_t - 2\varepsilon_{t-1} + \varepsilon_{t-2} + \Delta^2 \psi_t$$

The stochastic seasonal is defined via the recursions in (4.2):

$$\begin{aligned} \psi_t &= \rho \cos \lambda_c \psi_{t-1} + \rho \sin \lambda_c \psi_{t-1}^* + \kappa_t \\ \Rightarrow (1 - \rho \cos \lambda_c L) \psi_t &= \rho \sin \lambda_c L \psi_t^* + \kappa_t \\ \psi_t^* &= -\rho \sin \lambda_c \psi_{t-1} + \rho \cos \lambda_c \psi_{t-1}^* + \kappa_t^* \\ \Rightarrow (1 - \rho \cos \lambda_c L) \psi_t^* &= -\rho \sin \lambda_c L \psi_t + \kappa_t^* \end{aligned}$$

We can thus express the evolution of the cycle ψ_t as:

$$\begin{aligned} (1 - \rho \cos \lambda_c L) \psi_t &= \rho \sin \lambda_c L \left(\frac{-\rho \sin \lambda_c L \psi_t + \kappa_t^*}{1 - \rho \cos \lambda_c L} \right) + \kappa_t \\ (1 - \rho \cos \lambda_c L)^2 \psi_t &= \rho \sin \lambda_c L (-\rho \sin \lambda_c L \psi_t + \kappa_t^*) + (1 - \rho \cos \lambda_c L) \kappa_t \\ (1 - 2\rho \cos \lambda_c L + (\rho L)^2 (\cos^2 \lambda_c + \sin^2 \lambda_c)) \psi_t &= \kappa_t + \rho L (\sin \lambda_c \kappa_t^* - \cos \lambda_c \kappa_t^*) \\ \psi_t &= \frac{\kappa_t + \rho L (\sin \lambda_c \kappa_t^* - \cos \lambda_c \kappa_t^*)}{1 - 2\rho \cos \lambda_c L + \rho^2 L^2} \end{aligned}$$

This has the structure of an ARMA(2,1) model. We can take second differences and substitute

¹See lecture on Spectral Analysis

into the expression for $\Delta^2 y_t$:

$$\begin{aligned}\Delta^2 y_t &= \zeta_{t-1} + \eta_t + \eta_{t-1} + \varepsilon_t - 2\varepsilon_{t-1} + \varepsilon_{t-2} + \Delta^2 \psi_t \\ &= \zeta_{t-1} + \eta_t + \eta_{t-1} + \varepsilon_t - 2\varepsilon_{t-1} + \varepsilon_{t-2} + \frac{(1-L)^2(\kappa_{t-1} + \rho L(\sin \lambda_c \kappa_{t-1}^* - \cos \lambda_c \kappa_{t-1}^*))}{1 - 2\rho \cos \lambda_c L + \rho^2 L^2}\end{aligned}$$

We can rearrange to show:

$$\begin{aligned}(1 - 2\rho \cos \lambda_c L + \rho^2 L^2)\Delta^2 y_t &= (1 - 2\rho \cos \lambda_c L + \rho^2 L^2)(\zeta_{t-1} + \eta_t + \eta_{t-1} + \varepsilon_t - 2\varepsilon_{t-1} + \varepsilon_{t-2}) \\ &\quad + (1-L)^2(\kappa_{t-1} + \rho L(\sin \lambda_c \kappa_{t-1}^* - \cos \lambda_c \kappa_{t-1}^*))\end{aligned}$$

The left hand side is an AR(2) on the second difference of the series, and the right hand side is an MA(4) (the deepest lag is $\rho^2 \varepsilon_{t-4}$). Thus the reduced form of the trend-cycle model is an ARIMA(2,2,4) model. \square

The state space form of the model is:

$$\begin{aligned}y_t &= \begin{bmatrix} 1 & 0 & 1 & 0 \end{bmatrix} \alpha_t + \varepsilon_t \\ \alpha_t = \begin{bmatrix} \mu_t \\ \beta_t \\ \psi_t \\ \psi_t^* \end{bmatrix} &= \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \rho \cos \lambda_c & \rho \sin \lambda_c \\ 0 & 0 & -\rho \sin \lambda_c & \rho \cos \lambda_c \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \\ \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix} + \begin{bmatrix} \eta_t \\ \zeta_t \\ \kappa_t \\ \kappa_t^* \end{bmatrix}\end{aligned}$$

Note that this matrix is block diagonal, this is because the trend and cycle are uncorrelated.

Definition 4.4.3: Cyclical trend model

The cyclical trend model incorporates the cycle into the slope by moving it into the level equation:

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

We can write this in state space form:

$$\begin{aligned}y_t &= \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \alpha_t + \varepsilon_t \\ \alpha_t = \begin{bmatrix} \mu_t \\ \beta_t \\ \psi_t \\ \psi_t^* \end{bmatrix} &= \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \rho \cos \lambda_c & \rho \sin \lambda_c \\ 0 & 0 & -\rho \sin \lambda_c & \rho \cos \lambda_c \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \\ \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix} + \begin{bmatrix} \eta_t \\ \zeta_t \\ \kappa_t \\ \kappa_t^* \end{bmatrix}\end{aligned}$$

and see that the trend and cycle are now correlated!

Claim 4.4.2. The cyclical trend model, where the cycle appears in the slope by moving it into the level equation:

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

has reduced form ARIMA(2,2,4).

Proof. Fuck off □

Question 2

Show that the damped trend model:

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

Is a special case of the cyclical trend model.

Solution:-

Let us suppose our cycles have frequency zero, then $\lambda_c = 0$. Thus $\sin \lambda_c = 0$ and the second part of the recursion is redundant. Using $\cos 0 = 1$:

$$\psi_t = \rho\psi_{t-1} + \kappa_t$$

Consider the special case of the cyclical trend model where β_t does not appear in the level equation:

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

However this is exactly the same as the damped trend model, just with a change of notation! We can see this clearly by writing the structural forms side by side:

Damped Trend Model

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

Cyclical Trend Model

$$\begin{aligned}y_t &= \mu_t + \varepsilon_t \\ \mu_t &= \mu_{t-1} + \psi_{t-1} + \eta_t \\ \psi_t &= \rho\psi_{t-1} + \kappa_t\end{aligned}$$

These are the same, just interchanging the roles of β_t and ψ_t .

4.5 Unit roots and Stationarity tests

Two classes of tests are available to help distinguish between models that are integrated of different orders. In stationarity tests, nonstationarity appears under the alternative hypothesis. In unit root tests, the situation is reversed, with nonstationarity corresponding to the null hypothesis. This duality between the two sets of tests can also be seen in terms of AR and MA models. For example, the local level model is equivalent to an MA(1) in first differences, and the test of a zero variance driving the random walk corresponds to a test that the MA parameter is strictly noninvertible.

4.5.1 Unit root tests: Dickey-Fuller

The basic unit root test is of the null hypothesis that a series is a random walk against the alternative that it's a stationary AR(1). Thus in

$$y_t = \phi y_{t-1} + \varepsilon_t$$

with y_0 fixed, the null hypothesis is $\phi = 1$ and the alternative is $\phi < 1$. The Dickey-Fuller test is a one sided test based on the 't-statistic' obtained from the OLS estimate of ϕ . This is not asymptotically normal, but critical values are tabulated by Dickey and Fuller (1979). There are different critical values when the model includes a constant and/or a time trend:

$$y_t = \alpha + \beta t + \phi y_{t-1} + \varepsilon_t \quad \varepsilon_t \sim NID(0, \sigma^2)$$

Note:-

If the true model is a random walk plus drift and a constant, but no time trend ($y_t = \alpha + y_{t-1} + \varepsilon_t$) then the t statistic is asymptotically normal. See Hamilton (1994, p 495-7) if you're feeling brave.

We can subtract y_{t-1} from both sides to obtain the dickey fuller test:

Definition 4.5.1: Dickey-Fuller test

$$\Delta y_t = \rho y_{t-1} + \varepsilon_t \quad \rho = \phi - 1$$

$H_0 : \rho = 0$ and $H_1 : \rho < 0$ using Dickey-Fuller tables.

More generally we can consider the augmented Dickey-Fuller test, based on an AR(p):

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t$$

Definition 4.5.2: Augmented Dickey-Fuller test

$$\Delta y_t = \rho y_{t-1} + \sum_{i=1}^{p-1} \gamma_i \Delta y_{t-i} + \varepsilon_t$$

where

$$\gamma_i = - \sum_{j=i+1}^p \phi_j \quad \text{and} \quad \rho = \sum_{i=1}^p \phi_i - 1$$

$H_0 : \rho = 0$ and $H_1 : \rho < 0$ using Dickey-Fuller tables.

The distribution of $\hat{\rho}$ is unaffected by the lagged differences since it converges at a faster rate. Thus we use the standard DF tables. That $\rho = 0$ implies a unit root follows because when it holds the polynomial associated with the AR(p) factorises as $\phi(L) = (1 - L)\phi^*(L)$. Setting $L=1$ then gives $\phi(1) = 0$ and the equation for ρ given above.

4.5.2 Stationarity tests

Consider the random walk plus noise model (or local level model):

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

When $\sigma_\eta^2 = 0$ the random walk becomes a constant level. Thus a nonstationary model becomes stationary and standard asymptotic theory no longer applies to a test of $H_0 : \sigma_\eta^2 = 0$ against $H_1 : \sigma_\eta^2 > 0$.

Definition 4.5.3: Nyblom-Makelainen test

$$NM = \frac{1}{T^2 s^2} \sum_{i=1}^T \left[\sum_{t=1}^i e_t \right]^2 > c$$

where $e_t = y_t - \bar{y}$, $s^2 = T^{-1} \sum_{t=1}^T e_t^2$ and c is a critical value.

Asymptotically NM has a Cramer-von Mises distribution under the null when the irregular's are iid. This model can be extended such that ε_t follows any indeterministic stationary process, giving us the KPSS test.

Definition 4.5.4: KPSS test

Replace s^2 in the NM statistic with an estimator of the long run variance of ε_t :

$$\sigma_L^2(m) = \hat{\gamma}(0) + 2 \sum_{\tau=1}^m w(\tau, m) \hat{\gamma}(\tau)$$

where $w(\tau, m)$ is a weighting function, such as $w(\tau, m) = 1 - |\tau|/(m + 1)$, and $\hat{\gamma}(\tau)$ is the sample autocovariance at lag τ . The lag length m is specified by the researcher.

4.6 Seasonality

4.6.1 Deterministic seasonality

A seasonal component, γ_t may be added to a model consisting of a trend and irregular to give the *basic structural model* (BSM):

$$y_t = \mu_t + \gamma_t + \varepsilon_t$$

where the components evolve as before. The signal noise ratio associated with the seasonal component is $q_\omega = \sigma_\omega^2 / \sigma_\varepsilon^2$.

a fixed seasonal pattern may be modelled as:

$$\gamma_t = \sum_{j=1}^s \gamma_j z_{jt}$$

where s is the number of seasons and the dummy variables z_{jt} take the value 1 in the j th season and 0 otherwise. In order to not confound trend with seasonality, we constrain $\sum_{j=1}^s \gamma_j = 0$.

Of course seasonal patterns are not fixed in time, but the requirement that the terms sum to zero must be maintained when the seasonal pattern changes.

4.6.2 Stochastic seasonality

Let γ_{jt} denote the effect of season j at time t and define $\gamma_t = [\gamma_{1t}, \dots, \gamma_{st}]^T$. The seasonals evolve according to a multivariate random walk:

$$\gamma_t = \gamma_{t-1} + \omega_t \quad \equiv \quad \begin{bmatrix} \gamma_{1t} \\ \vdots \\ \gamma_{st} \end{bmatrix} = \begin{bmatrix} \gamma_{1,t-1} \\ \vdots \\ \gamma_{s,t-1} \end{bmatrix} + \begin{bmatrix} \omega_{1t} \\ \vdots \\ \omega_{st} \end{bmatrix}$$

where $\omega_t = [\omega_{1t}, \dots, \omega_{st}]^T$ is a zero-mean disturbance with

$$Var(\omega_t) = \sigma_\omega^2 \left(I - \frac{ii'}{s} \right) = \sigma^2 \begin{bmatrix} 1 - \frac{1}{s} & -\frac{1}{s} & \dots & -\frac{1}{s} \\ -\frac{1}{s} & 1 - \frac{1}{s} & \dots & -\frac{1}{s} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{s} & -\frac{1}{s} & \dots & 1 - \frac{1}{s} \end{bmatrix}$$

This structure is necessary to ensure that the seasonal components always sum to zero. The sum of the elements of γ_t can be written as $i'\gamma_t$. From this structure we can see that $Var(i'\omega_t) =$

$\sigma_{\omega}^2(i' - \frac{i'ii'}{s}) = \sigma_{\omega}^2(i' - i') = 0$. Combining this with an initial condition setting the sum 0 at $t=0$, we have that the expectation and variance of $i'\gamma_t$ are always zero.

Note that here although all s seasonal components are continually changing, only one affects the observation at time t , that is $\gamma_t = \gamma_{jt}$ when j is prevailing at t .

We can write this in state-space form:

$$y_t = \begin{bmatrix} 1 & 0 & z'_t \end{bmatrix} \alpha_t + \varepsilon_t$$

$$\alpha_t = \begin{bmatrix} \mu_t \\ \beta_t \\ \gamma_t \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & I_s \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \\ \gamma_{t-1} \end{bmatrix} + \begin{bmatrix} \eta_t \\ \zeta_t \\ \omega_t \end{bmatrix}$$

where the bottom row in the transition equation is just rewriting the multivariate random walk. We can write the reduced form of the dummy variable stochastic seasonal model as:

$$\gamma_t = - \sum_{j=1}^{s-1} \gamma_{t-j} + \omega_t$$

which is simply rearranging the fact that the sum of the seasonals equals a random disturbance.

Claim 4.6.1. The reduced form of the BSM with stochastic seasonality is $\Delta\Delta_s y_t \sim MA(s+1)$

Proof.

$$y_t = \mu_t + \gamma_t + \varepsilon_t$$

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

$$\beta_t = \beta_{t-1} + \zeta_t$$

$$\gamma_t = - \sum_{j=1}^{s-1} \gamma_{t-j} + \omega_t$$

We can make the stochastic trend component stationary by second differencing as usual:

$$\Delta^2 \mu_t = \Delta \beta_{t-1} + \eta_t - \eta_{t-1} = \zeta_{t-1} + \eta_t - \eta_{t-1}$$

The seasonal component can be made stationary by applying the seasonal summation operator

Definition 4.6.1: Seasonal summation operator

$$S(L) = 1 + L + L^2 + \dots + L^{s-1}$$

which is a special case of the summation operator.

Thus $S(L)\gamma_t = \omega_t$ (by rearranging the reduced form, we see the sum of seasonals is ω_t). The stationary form of the BSM is obtained by multiplying through by $\Delta^2 S(L) = \Delta\Delta_s$. This follows from

$$(1 - L)S(L) = 1 - L^s = \Delta_s \Rightarrow \Delta^2 S(L) = \Delta\Delta_s$$

The result is:

$$\Delta\Delta_s y_t = \Delta_s \eta_t + S(L)\zeta_{t-1} + \Delta^2 \omega_t + \Delta\Delta_s \varepsilon_t$$

Which is clearly an $Ma(s+1)$ from the deepest lag on epsilon. □

Note:-

If the slope were deterministic in the above, we would not need to difference out the trend and the observations are rendered stationary by the seasonal difference alone, giving $\Delta_s y_t \sim MA(s+1)$ plus a constant trend β .

4.7 Seasonal ARIMA models

Definition 4.7.1: SARIMA(p,d,q)(P,D,Q)_s

The seasonal ARIMA model is given by:

$$\phi(L)\Phi(L^s)\Delta^d\Delta_s^D y_t = \theta_0 + \theta(L)\Theta(L^s)\varepsilon_t$$

where θ_0 is a constant, while $\Phi(L^s)$ and $\Theta(L^s)$ are polynomials in the seasonal lag operator L^s :

$$\Phi(L^s) = 1 - \Phi_1 L^s - \dots - \Phi_P L^{sP} \quad \Theta(L^s) = 1 - \Theta_1 L^s - \dots - \Theta_Q L^{sQ}$$

The most common implementation of this very broad class of models is known as the *airline model*, which is a SARIMA(0,1,1)(0,1,1)_s model:

$$\Delta\Delta_s y_t = (1 + \theta L)(1 + \Theta L^s)\varepsilon_t$$

Claim 4.7.1. With a deterministic seasonal and no trend, the airline model is equivalent to the BSM when $\Theta = -1$.

Proof. The BSM is:

$$\begin{aligned} y_t &= \mu_t + \gamma_t + \varepsilon_t \\ \mu_t &= \mu_{t-1} + \eta_t \\ \gamma_t &= \sum_{j=1}^s \gamma_j z_{jt} \end{aligned}$$

We can write the seasonally differenced series as:

$$\begin{aligned} \Delta_s y_t &= \Delta_s \mu_t + \Delta_s \gamma_t + \Delta_s \varepsilon_t \\ &= \mu_t - \mu_{t-s} + \varepsilon_t - \varepsilon_{t-s} \end{aligned}$$

Since $\gamma_t - \gamma_{t-s} = 0$. We now iterate on the level equation:

$$\begin{aligned} \mu_t &= \mu_{t-1} + \eta_t \\ &= \mu_{t-2} + \eta_{t-1} + \eta_t \\ &= \dots \\ &= \mu_{t-s} + \sum_{j=1}^s \eta_{t-j} \end{aligned}$$

Thus we can write the seasonally differenced series as:

$$\Delta_s y_t = \sum_{j=1}^s \eta_{t-j} + \varepsilon_t - \varepsilon_{t-s} = S(L)\eta_t + \Delta_s \varepsilon_t$$

Now we first difference the series:

$$\begin{aligned}\Delta\Delta_s y_t &= \Delta S(L)\eta_t + \Delta\Delta_s \varepsilon_t &= \Delta_s \eta_t + \Delta\Delta_s \varepsilon_t &\text{since } \Delta S(L) = \Delta_s \\ &= (1 - L^s)(\eta_t + \varepsilon_t - \varepsilon_{t-1})\end{aligned}$$

This is an MA(s+1) process, and we can compare it to the airline model:

$$\Delta\Delta_s y_t = (1 + \Theta L^s)(\varepsilon_t + \theta \varepsilon_{t-1}) \quad \Delta\Delta_s y_t = (1 - L^s)(\eta_t + \varepsilon_t - \varepsilon_{t-1})$$

Thus the airline model is equivalent to the BSM when $\Theta = -1$, since both final terms are MA(1) processes.

The airline model thus provides a good approximation to the reduced form when the slope and seasonal are close to being deterministic. \square

4.8 Model Selection

Structural models aim to capture the most obvious parts of a time series; when there are obvious trends or seasonal patterns to capture. ARIMA models are more parsimonious, however they may miss some of the more subtle features of the data.

ARIMA selection is based on the premise that ACFs are stable over time. Even if this is the case, it is still difficult to identify complex models from the ACF alone, and we may miss important features of the series. In practice sampling error in the correlogram can make it difficult to identify even simple ARIMA models. STMs are more robust as the choice of model is not dependent on correlograms.

Model selection for ARIMAs is normally straightforward, unit root tests determine the degree of differencing and lags are included according to some information criterion or statistical significance. However, these tests often have poor size and power properties.