Computational
Social Science

Modeling Temporal Data .I

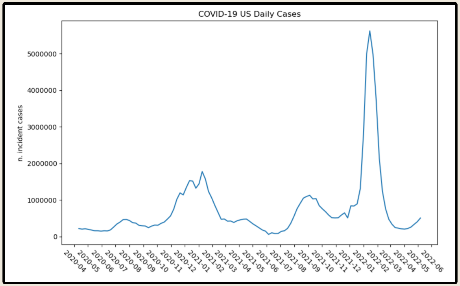
Roberto Cerina

19.03.2024

University of Amsterdam

Temporal Data

- Let y_t represent the value of the time series at time t, where $t = 1, 2, \ldots, T$, for a total of T time periods.
- The values can represent counts, measurements, or observations recorded over time.
- E.g. consider a daily count of events: $\{y_1, y_2, \ldots, y_T\} = \{2, 3, 5, 2, 1, \ldots\}$, where each $y_t \in \mathbb{N}_0$ and represents the count of events on day t.
- The goal of time-series modeling is typically to forecast future values of the series...
- We might further interested in making inference on the impact of past values on present ones.



Random Walk

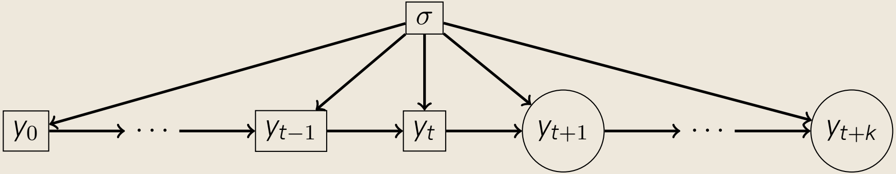
- Our best point-estimate guess at today's value is yesterday's value;
- We allow for unsystematic deviations from yesterday's value.

$$
\begin{aligned}
y_t &\sim N(\mu_t, \sigma^2) \\
\mu_t &= y_{t-1}
\end{aligned}
\qquad\qquad
\begin{aligned}
y_t &= y_{t-1} + \epsilon_t \\
y_t - y_{t-1} &= \epsilon_t \\
\epsilon_t &\sim N(0, \sigma^2)
\end{aligned}
$$

- Uncertainty depends on the average (squared) daily deviations;
- σ is the only parameter left to estimate in this model
- As we have done up to now, we continue assuming the σ^2 can be measured from the sample without any uncertainty (no posterior distribution):

$$\hat{\sigma} = \sqrt{\frac{1}{T}\sum_t (y_t - y_{t-1})^2}$$

Random Walk

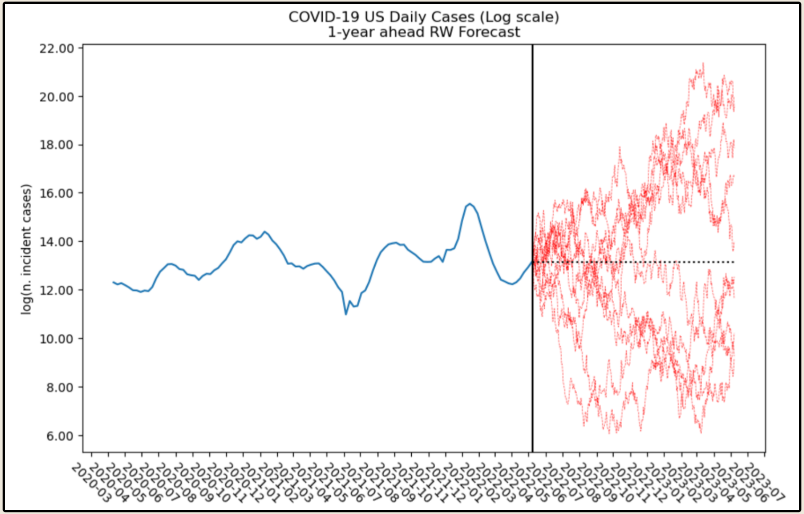


Random Walk

- Forecasting $\rightarrow$ predictions for 'steps ahead'...

$$
\begin{aligned}
\text{1-step ahead: } & \hat{y}_{t+1} \sim N(y_t, \hat{\sigma}^2) \\
\text{2-steps ahead: } & \hat{y}_{t+2} \sim N(\hat{y}_{t+1}, \hat{\sigma}^2) \\
& \vdots \\
\text{k-steps ahead: } & \hat{y}_{t+k} \sim N(\hat{y}_{t+(k-1)}, \hat{\sigma}^2)
\end{aligned}
$$

Posterior Predictive Distribution



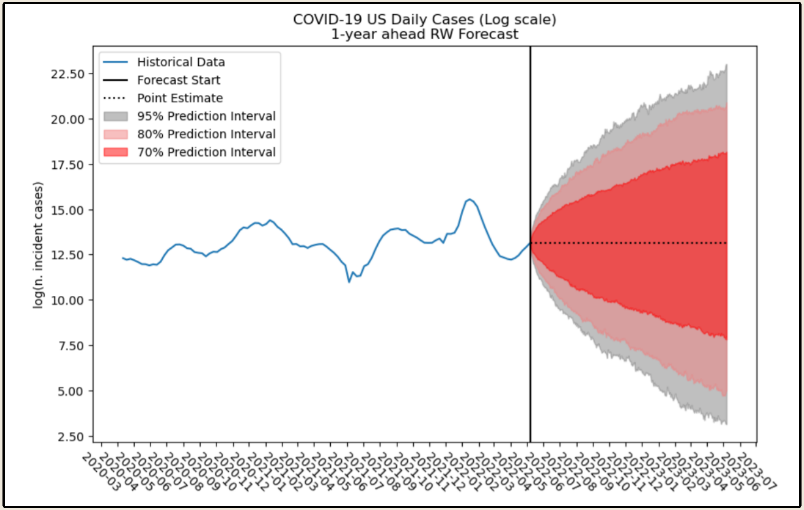
Posterior Predictive Distribution

❶ define a number of steps-ahead k;

❷ simulate S values for y_{t+1} according to its distribution:
$\tilde{y}^s_{t+1} \sim N(y_t, \hat{\sigma})$;
from each simulated y^s_{t+1}, simulate a value for for y_{t+2}:
$\tilde{y}^s_{t+2} \sim N(\tilde{y}^s_{t+1}, \hat{\sigma})$;
$\vdots$
from each simulated $y^s_{t+(k-1)}$, simulate a value for for y_{t+k}:
$\tilde{y}^s_{t+k} \sim N(\tilde{y}^s_{t+(k-1)}, \hat{\sigma})$;

You can then use Monte Carlo methods to make inference about predictions at any given future time point $t+k$.

Posterior Predictive Distribution



Stationarity

- Random Walks are **non-stationary**
  - ✏ *A process is **stationary** if its statistical properties (i.e. its expected value, variance, etc.) do not change over time.*
  - We can show non-stationarity of RWs by re-writing the model as a sum of error terms...
  - This is akin to considering the 'marginal' distribution of y_t, as opposed to the conditional distribution of $y_t \mid y_{t-1}$ which we have seen so far...
  - This is the theoretical, expected distribution of y_t, in a world where we do not know $y_t - 1$;
  - Think of this as the predictive distribution for k-steps ahead.

  - Assume $y_0 = 0$, such that:

$$
\begin{aligned}
y_1 &= \epsilon_1 \\
y_2 &= y_1 + \epsilon_2 = \epsilon_1 + \epsilon_2 \\
&\vdots \\
y_k &= \textstyle\sum_{t=1}^{k} \epsilon_t
\end{aligned}
$$

  - It follows that: $\text{Var}(y_k) = \sum_{t=1}^{k} \text{Var}(\epsilon_t) = k\sigma^2$
  - ✏ Hence the variance of the RW process gets larger over time...
- The implications of non-stationarity will become clear when we look at more complex models.

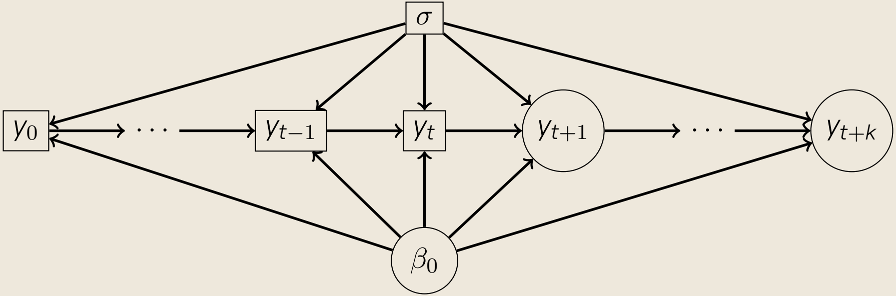
Random Walk with Drift

- Our best point-estimate guess at today's value is yesterday's value + a constant (drift);

$$
\begin{aligned}
y_t \sim& N(\mu_t, \sigma^2) \\
\mu_t =& \beta_0 + y_{t-1}
\end{aligned}
\qquad\qquad
\begin{aligned}
y_t =& \beta_0 + y_{t-1} + \epsilon_t \\
y_t - y_{t-1} =& \beta_0 + \epsilon_t \\
\epsilon_t \sim& N(0, \sigma^2)
\end{aligned}
$$

- Assume σ^2 can be measured as with the RW;
- Estimate β_0 via Maximum Likelihood (this is a simple 'homogeneous expectations' regression model on the change in y);
- Empirical Posterior distribution of β_0 can be obtained by drawing samples from:
  $\beta_0 \sim N(\hat{\beta}_0, \hat{\sigma}_{\beta_0})$

Random Walk with Drift

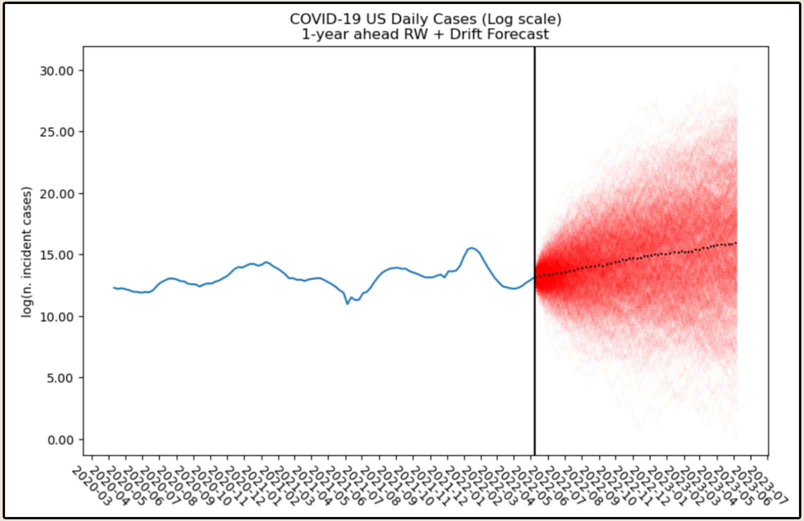


Stationarity

- This process is also **non-stationary**;
- The variance of the marginal distribution of y_t is $k\sigma^2$, same as standard RW;
- But the expected value is also time-dependent...
  - Assume $y_0 = 0$, such that:

    $$\begin{aligned} y_1 &= \beta_0 + \epsilon_1 \\ y_2 &= \beta_0 + y_1 + \epsilon_2 = 2\beta_0 + \epsilon_1 + \epsilon_2 \\ &\vdots \\ y_k &= k\beta_0 + \textstyle\sum_{t=1}^{k} \epsilon_t \end{aligned}$$
  - It follows that: $\mathrm{E}(y_k) = k\beta_0$
  - Hence the expected value of the RW with drift process gets larger (in absolute terms) over time...

Posterior Predictive Distribution



Posterior Predictive Distribution

1. define a number of steps-ahead k;
2. simulate S values for β_0 according to its distribution:
   $\beta_0^s \sim N(\hat{\beta}_0, \hat{\sigma}_{\beta_0})$
3. simulate S values for y_{t+1} according to its distribution:
   $\tilde{y}_{t+1}^s \sim N(\beta_0^s + y_t, \hat{\sigma})$;
   from each simulated y_{t+1}^s, simulate a value for for y_{t+2}:
   $\tilde{y}_{t+2}^s \sim N(\beta_0^s + \tilde{y}_{t+1}^s, \hat{\sigma})$;
   $\vdots$
   from each simulated $y_{t+(k-1)}^s$, simulate a value for for y_{t+k}:
   $\tilde{y}_{t+k}^s \sim N(\beta_0^s + \tilde{y}_{t+(k-1)}^s, \hat{\sigma})$;

You can then use Monte Carlo methods to make inference about predictions at any given future time point $t + k$.

Stationarity

- ✏ For RWs, we can live with non-stationarity:
  - (quasi) *maximum ignorance* model....
  - RW incorporates exclusively the **average magnitude of step-wise change**;
  - RW is ignorant about the **direction of step-wise change**;
  - RW is non-stationary and **degenerates** as $t \to \infty$;
- ✏ We would use RWs in highly uncertain events with a short k-steps-ahead forecast window;
- ✏ The goal is typically to incorporate uncertainty about an event in decision making, not to predict exact values...

Moving-Average Model

- $MA(q)$, where q indicates the number of lags to be used in the model ;
  - Example: $MA(q)$
  - θ_1 captures the lingering effect of ‘shocks’ in past time-periods.

$$
\begin{aligned}
y_t &\sim N(\mu_t, \sigma^2) \\
\mu_t &= \beta_0 + \theta_1 \epsilon_{t-1}
\end{aligned}
\qquad\qquad
\begin{aligned}
y_t &= \beta_0 + \theta_1 \epsilon_{t-1} + \epsilon_t \\
\epsilon_t &\sim N(0, \sigma^2)
\end{aligned}
$$

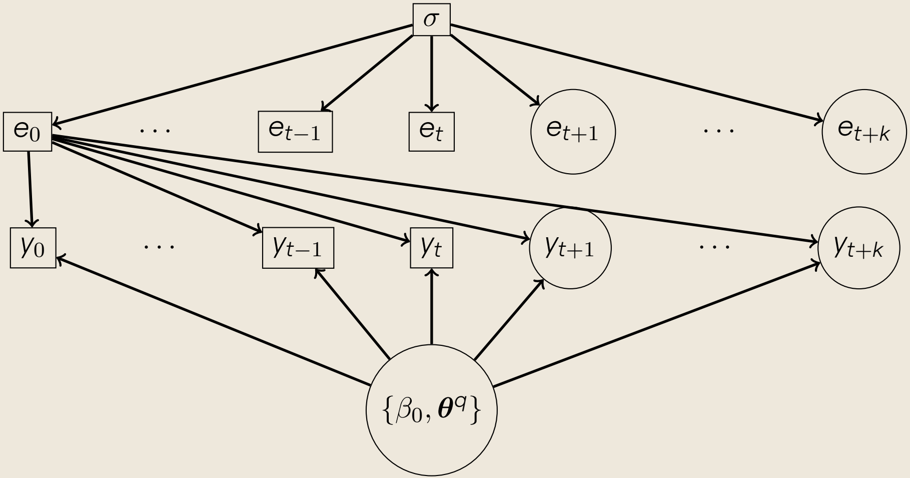
Estimation

- Estimating MA models can be somewhat tricky...
- past residuals are necessary to estimate θ_1...
- but change every-time point and are therefore dependent on θ_1...

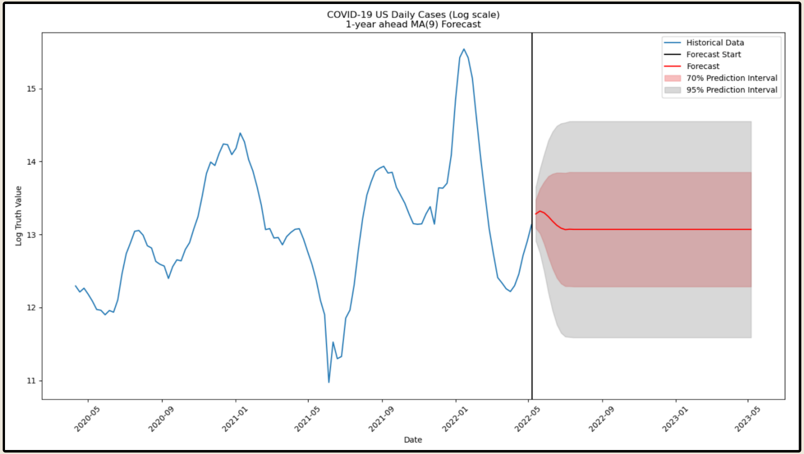
$$
\begin{aligned}
y_0 &= \beta_0 + \epsilon_0 & \qquad e_0 &= y_0 - \hat{y}_0 \\
y_1 &= \beta_0 + \theta_1 \epsilon_0 + \epsilon_1 & \qquad e_1 &= y_1 - \hat{y}_1 \\
&\vdots
\end{aligned}
$$

- To calculate the errors at each time-point I need to already have an estimate of θ_1...
- An analytical solution is not possible because of this ‘circular’ property of the model...
- We use numerical methods ! (Optimisation)

Moving-Average Model



Posterior Predictive Distribution



Posterior Predictive Distribution

❶ Extract residuals from the fitted model – here we can avoid taking the full distribution for simplicity, and work with the point-estimate of the residuals:
$\hat{e}_t = y_t - \hat{y}_t, \ \ \forall t \in \{0, ..., T\}$

❷ define a number of steps-ahead k;

❸ simulate S values for β_0 and $\boldsymbol{\theta}$ according their posterior distribution:
$\{\beta_0, \boldsymbol{\theta}\}^s \sim N(\{\hat{\beta}_0, \hat{\boldsymbol{\theta}}\}, \hat{\boldsymbol{\Sigma}})$

❹ simulate S values for y_{t+1} according to its distribution:
$\tilde{y}^s_{t+1} \sim N(\beta^s_0 + \theta^s_1 \hat{e}_t + ... + \theta^s_q \hat{e}_{t-q+1)}, \hat{\sigma});$

❺ simulate S values for e_{t+1} from its distribution:
$\tilde{e}^s_{t+1} \sim N(0, \hat{\sigma})$

$\vdots$

from each simulated $e^s_{t+(k-1)}$, simulate a value for for y_{t+k}:
$\tilde{y}^s_{t+k} \sim N(\beta^s_0 + \theta^s_1 \tilde{e}_{t+(k-1)} + ... + \theta^s_q \hat{e}^s_{t-q+k}, \hat{\sigma});$

Stationary $MA(q)$: Expected Value

- $y_t = \beta_0 + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q}$
- $E[y_t] = E[\beta_0 + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q}]$

$$E[X_t] = \beta_0$$

Stationary $MA(q)$: Variance

- $y_t = \beta_0 + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q}$
- $Var(y_t) = Var(\beta_0 + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q})$

✏ Assume β_0 is a known constant so it does not contribute to Variance in y

- $Var(y_t) = Var(\epsilon_t) + \theta_1 Var(\epsilon_{t-1}) + \theta_2 Var(\epsilon_{t-2}) + \ldots + \theta_q Var(\epsilon_{t-q})$

✏ since $Var(ce_t) = c^2 Var(e_t)$

$$Var(y_t) = \sigma^2(1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2)$$

Auto-Regressive Model

- $AR(p)$, where p indicates the number of lags to be used in the model ;
  - Example: $AR(1)$
  - β_1 captures systematic **autocorrelation**;
  - noise ϵ_t captures unsystematic / random influences on the current value y_t.

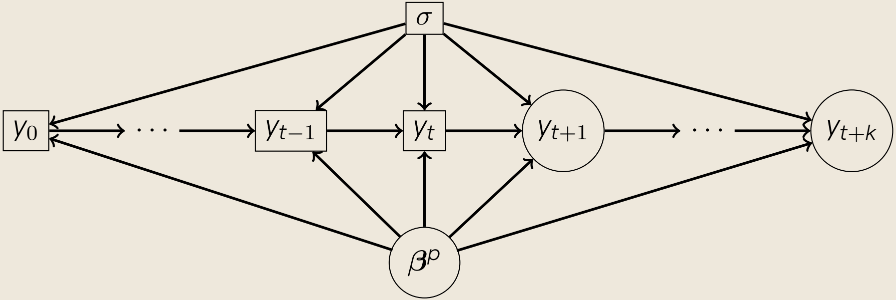
$$y_t \sim N(\mu_t, \sigma^2) \qquad y_t = \beta_0 + \beta_1 y_{t-1} + \epsilon_t$$

$$\mu_t = \beta_0 + \beta_1 y_{t-1} \qquad \epsilon_t \sim N(0, \sigma^2)$$

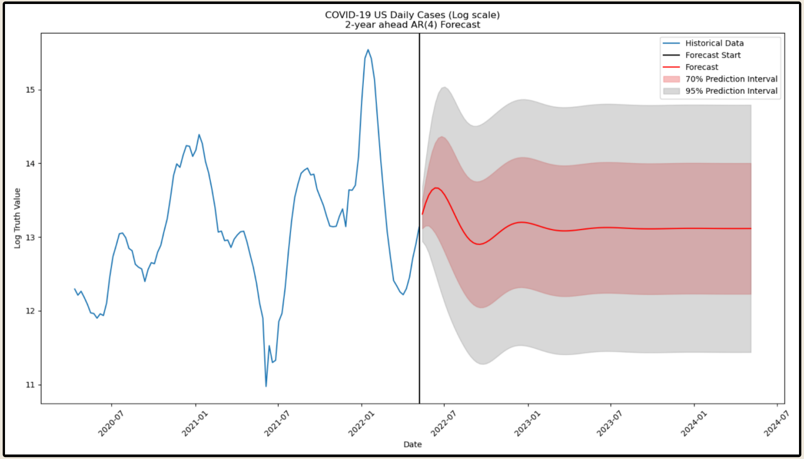
- $\beta_1 = 1$ recovers a Random Walk 'with drift'.
- The model parameters can be estimated via Maximum Likelihood (this is a classic 'heterogeneous expectations' regression model):
  - The posterior distribution of $\beta_0, ..., \beta_p$ is :
    $\boldsymbol{\beta} \sim MN(\hat{\boldsymbol{\beta}}_{MLE}, (\boldsymbol{X}^T\boldsymbol{X})^{-1}\sigma^{2\star})$
    $\boldsymbol{\Sigma} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\sigma^{2\star}$[1]

[1] $\boldsymbol{X}$ in this equation is the matrix of 'lags', so we are treating past values of the outcome as we would any independent variable.

Auto-Regressive Model



Posterior Predictive Distribution



Posterior Predictive Distribution

1. define a number of steps-ahead k;
2. simulate S values for $\beta_1, ..., \beta_p$ according to the posterior distribution:
   $\boldsymbol{\beta}^s \sim N(\hat{\boldsymbol{\beta}}, \boldsymbol{\Sigma})$
3. simulate S values for y_{t+1} according to its distribution:
   $\tilde{y}^s_{t+1} \sim N(\beta^s_0 + \beta^s_1 y_t + ... + \beta^s_p y_{t-p+1)}, \hat{\sigma})$;
   from each simulated y^s_{t+1}, simulate a value for for y_{t+2}:
   $\tilde{y}^s_{t+2} \sim N(\beta^s_0 + \beta^s_1 \tilde{y}_{t+1} + ... + \beta^s_p y_{t-p+2}, \hat{\sigma})$;
   $\vdots$
   from each simulated $y^s_{t+(k-1)}$, simulate a value for for y_{t+k}:
   $\tilde{y}^s_{t+k} \sim N(\beta^s_0 + \beta^s_1 \tilde{y}_{t+(k-1)} + ... + \beta^s_p y^s_{t-p+k}, \hat{\sigma})$;

You can then use Monte Carlo methods to make inference about predictions at any given future time point $t + k$.

Stationary $AR(1)$: Expected Value

- Using an $AR(1)$ model as example:
  $y_t = \beta_0 + \beta_1 y_{t-1} + \epsilon_t$
- Substitute $y_{t-1} = \beta_0 + \beta_1 y_{t-2} + \epsilon_{t-1}$ into the $AR(1)$ equation:
  $Y_t = \beta_0 + \beta_1(\beta_0 + \beta_1 y_{t-2} + \epsilon_{t-1}) + \epsilon_t$
  $\vdots$
  $y_t = \beta_0 + \beta_1\beta_0 + \beta_1^2 c + \cdots + \beta_1^n y_{t-n} + \epsilon_t + \beta_1\epsilon_{t-1} + \beta_1^2\epsilon_{t-2} + \cdots + \beta_1^n\epsilon_{t-n}$
- As$t \to \infty$, assuming $|\beta_1| < 1$[2], $\beta_1^n \to 0$:
  $Y_t = \beta_0 + \beta_1\beta_0 + \beta_1^2\beta_0 + \cdots + \beta_1^n\beta_0 + \cdots + \epsilon_t + \beta_1\epsilon_{t-1} + \beta_1^2\epsilon_{t-2} + \dots$
- $\beta_0 + \beta_1\beta_0 + \beta_1^2\beta_0 + \dots$ can be expressed as an infinite geometric series, which sums to $\frac{\beta_0}{1-\beta_1}$ when $|\beta_1| < 1$, hence we are left with:
  $y_t = \frac{\beta_0}{1-\beta_1} + \textit{sum of error terms}$
- The *sum of error terms* has mean of zero by definition, therefore:

$$E(y_t) = \frac{\beta_0}{1 - \beta_1}$$

[2] More on this later, more complex for $AR(p > 1)$

Stationary $AR(1)$: Variance

- Starting with the $AR(1)$ model:
  $y_t = \beta_0 + \beta_1 y_{t-1} + \epsilon_t$
- The variance of both sides gives:
  $Var(y_t) = Var(\beta_0 + \beta_1 y_{t-1} + \epsilon_t)$
- Since β_0 is a constant and variance of a constant is 0:
  $Var(y_t) = \beta_1^2 Var(y_{t-1}) + Var(\epsilon_t)$
- via stationarity, $Var(y_t) = Var(y_{t-1})$, therefore:
  $Var(y_t) = \beta_1^2 Var(y_t) + \sigma_\epsilon^2$

$$Var(y_t) = \frac{\sigma_\epsilon^2}{1 - \beta_1^2}$$

ARMA Model

✏ You can mix the two model components to fit an ARMA(p,q) model:

$$y_t = \beta_0 + \sum_{j=1}^{p} \beta_j y_{t-j} + \sum_{h=1}^{q} \theta_h \epsilon_{t-h} + \epsilon_t$$

Stationarity

- ✏ ARMA models assume the underlying time-series is **stationary**...
- ✏ ...this is typically not the case !
- ✖ Problems with **non-stationarity**:
  - ✏ non-stationarity implies I cannot generalise out of my 'temportal window' of observations...
  - ✏ I cannot reliably make inference or forecast future values...
  - ✏ say I observe a non-stationary series $\{y_0,...,y_k\}$, and estimate an AR(1) model...
    - fitting an AR(1) model to a subsequent series $\{y_z,...,y_{k+z}\}$ (values from the same series but observed at a later stage) will yield different values of $\hat{\beta}$ and $\hat{\sigma}$;
    - The $\hat{\beta}$ and $\hat{\sigma}$ I can estimate from my sample will be **biased** (different from the true values) and **inconsistent** (increasing sample-size is not guaranteed to get them any closer to the true values).