

Modelos para series de tiempo no estacionarias

Curso de Series Cronológicas

Silvia Rodríguez Collazo

Licenciatura de Estadística

Facultad de Ciencias Económicas y de Administración

Un modelo univariado que puede ser escrito como:

$$Y_t = \mu + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots = \mu + \psi(L) \varepsilon_t \text{ donde } \sum_{j=0}^{\infty} |\psi_j| < \infty, \text{ las raíces}$$

de $\psi(L)$ se encuentran fuera del círculo unidad y $\{\varepsilon_t\}$ es una secuencia ruido blanco con media cero y varianza σ_ε^2 . Interesa destacar dos características de este modelo:

- La esperanza incondicional de la variable es constante e independiente del período t . $E(Y_t) = \mu$
- La predicción a s pasos $\hat{Y}_{t+s} = \hat{E}(Y_{t+s}/y_t, y_{t-1}, \dots)$ converge a la media incondicional, es decir $\lim_{s \rightarrow +\infty} \hat{Y}_{t+s/t} = \mu$

Sin embargo, hay series que no tienen estas características, por ejemplo por la existencia de tendencia.

Qué es una tendencia?

Tomemos la respuesta que elaboran Stock y Watson (1988), como primer acercamiento consideremos una **tendencia determinística lineal**, en la que la serie crece a una cantidad fija por ejemplo en cada trimestre.

Pero si modelamos la tendencia como un incremento e cierta cantidad fija(1%) **en promedio** en cada trimestre, pero en un trimestre dado, el cambio en la tendencia se puede desviar de esa media en una cantidad aleatoria impredecible. Porque es una cantidad aleatoria impredecible es que se denomina a esta tendencia, **tendencia estocástica**. Un ejemplo de esto es el Random Walk (RW) con drift, que veremos más adelante.

Desde una perspectiva de predicción, la característica fundamental del un RW, en contraste con una serie estacionaria, es que el crecimiento es aleatorio, predecir su nivel implica un nivel de incertidumbre que crece a medida que el horizonte de predicción de aleja del momento t .

Con esta formulación sobre tendencia estocástica, un cambio aleatorio en la tendencia en un trimestre da una nueva base para el crecimiento que ocurrirá en el período siguiente.

Podemos definir no estacionariedad por la negativa, un proceso es no estacionario si no cumple las condiciones de estacionariedad.

Vamos a caracterizar tres tipos de no estacionariedad acorde al incumplimiento de las propiedades que verifican los procesos estacionarios

- Evolutividad de la media. $E(X_t) \neq \mu$
- Evolutividad de la varianza $V(X_t) \neq \sigma^2$
- Cambios en las autocorrelaciones a lo largo del tiempo.

En lo que sigue vamos a hacer un análisis de la no estacionariedad de los procesos considerando:

- Tipos específicos de no estacionariedad
- Estudiando las propiedades estadísticas de distintos tipos de representaciones (modelos)

Existen dos formas usuales de describir la tendencia:

a) Incluyendo una **tendencia determinística**, es decir $Y_t = \alpha + \delta t + \psi(L)\varepsilon_t$

En este caso la media el proceso, μ , se reemplaza por una función lineal del tiempo, $\mu = \alpha + \delta t$

A este tipo de procesos se los denomina estacionarios en tendencia (en inglés **Trend stationary, TS**), pues si se sustrae la tendencia $\mu = \alpha + \delta t$ el proceso es estacionario.

b) Un proceso con **raíz unitaria**

$$(1 - L)Y_t = \delta + \psi(L)\varepsilon_t$$

donde la representación estacionaria $(1 - L)Y_t = \Delta Y_t$ describe las variaciones de la serie.

Siendo $\psi(1) \neq 0$ y δ la media de $(1 - L)Y_t$

Si $\psi(L) = 1$, el proceso que se obtiene es:

$$Y_t = Y_{t-1} + \delta + \varepsilon_t$$

y se denomina **paseo aleatorio con deriva δ** .

A estos procesos se les denomina procesos estacionarios en diferencias o en inglés, **Difference stationary, DS**.

La razón por la que se supone que $\psi(1) \neq 0$ es la siguiente:

$\psi(z) = 1 + \psi_1 z^1 + \psi_2 z^2 + \psi_3 z^3 + \dots$. Si sustituimos por $z = 1$ se obtiene $\psi(1) = 1 + \psi_1 + \psi_2 + \psi_3 + \dots$ que es distinto de cero.

Por otro lado, si Y_t es estacionaria admite la representación $Y_t = \mu + \chi(L)\varepsilon_t$. Y al diferenciarla resulta:

$$(1-L)Y_t = (1-L)\chi(L)\varepsilon_t = \psi(L)\varepsilon_t$$

con $\psi(L) = (1-L)\chi(L)$ que en $L = 1$ es igual a cero. $\psi(1) = 0$

Cuando se estipula que $\psi(1) \neq 0$ se deja fuera la posibilidad de que la serie original sea estacionaria.

Observación: Si Y_t fuera $I(1)$ (integrada de orden uno), ΔY_t es $I(0)$ (integrado de orden cero) pues $\Delta Y_t = \chi(L)\varepsilon_t$ y $\chi(1) = 1 + \chi_1 + \chi_2 + \dots$ que es diferente de cero.

Consideremos el proceso TS

$$Y_t = \alpha + \delta t + \mu_t,$$

donde μ_t , es un proceso ARMA con media cero estacionario e invertible.

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p) \mu_t = (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) \varepsilon_t$$

Si se factoriza el polinomio autorregresivo se obtiene

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p) = (1 - \lambda_1 L)(1 - \lambda_2 L) \dots (1 - \lambda_p L)$$

Si todos los autovalores $\lambda_1, \lambda_2, \dots, \lambda_p$ están dentro del círculo unitario, el proceso μ_t es estacionario:

$$\mu_t = \frac{1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q}{(1 - \lambda_1 L)(1 - \lambda_2 L) \dots (1 - \lambda_p L)} \varepsilon_t \cong \psi(L) \varepsilon_t \text{ con } \sum_{j=0}^{\infty} |\psi_j| < \infty$$

y las raíces de $\psi(z)$ fuera del círculo unidad.

Cuando $|\lambda_i| < 1 \forall_i$ el proceso Y_t es un proceso TS.

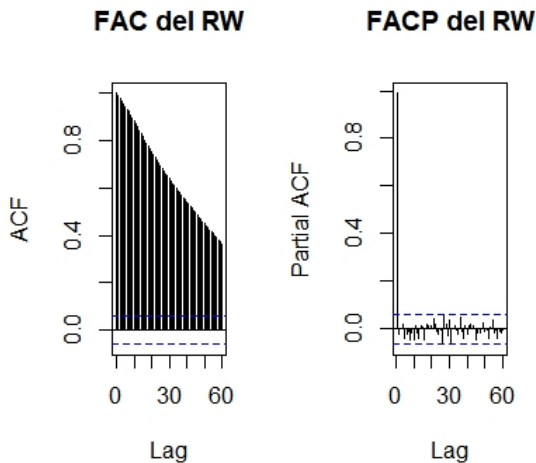


Figura: ACF y PACF de un proceso camino aleatorio, RW

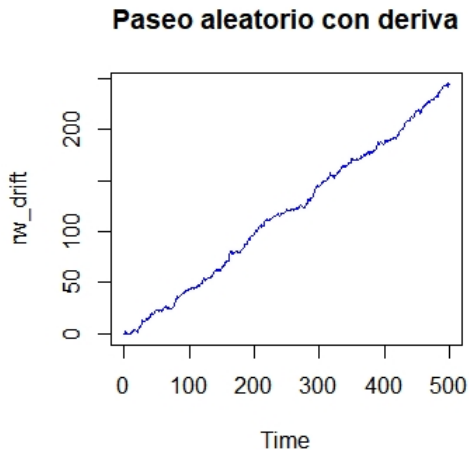


Figura: Gráfico de un proceso camino aleatorio con deriva

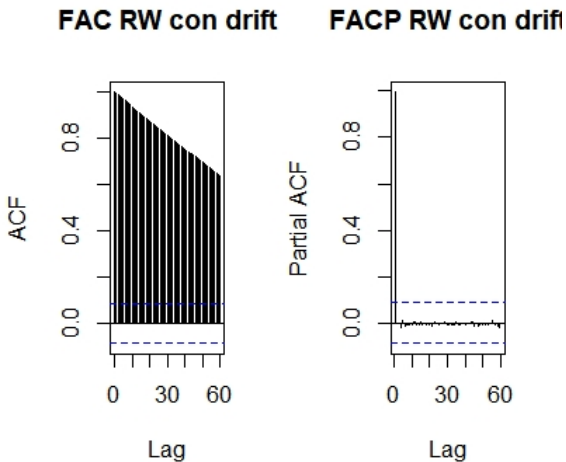


Figura: ACF y PACF de un proceso camino aleatorio con deriva

Proceso TS, estacionario en torno a una tendencia determinística

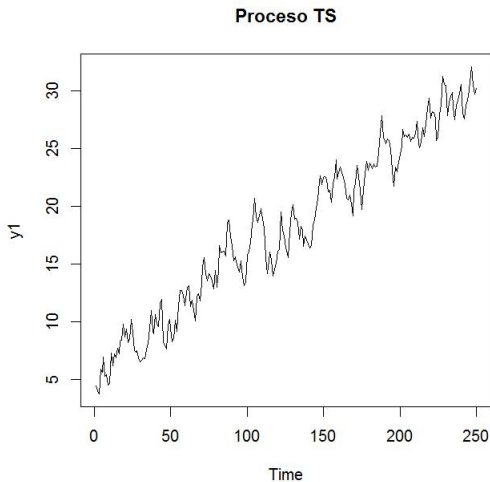


Figura: Gráfico de un proceso TS

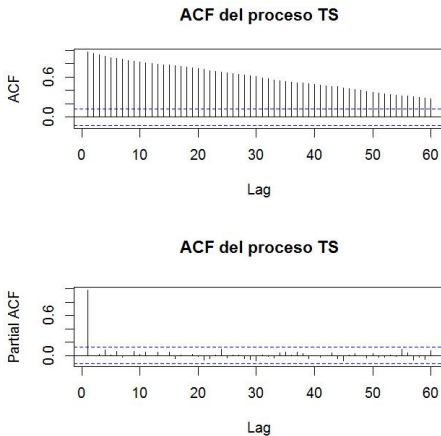


Figura: ACF y PACF de un proceso TS

Para predecir un proceso estacionario con tendencia determinística se adiciona el componente determinístico $\alpha + \delta t$ a la predicción del componente estacionario.

$$Y_t = \alpha + \delta t + \psi(L)\varepsilon_t$$

$$\hat{Y}_{t+s/t} = \alpha + \delta(t+s) + \psi_S \varepsilon_t + \psi_{S+1} \varepsilon_{t-1} + \psi_{S+2} \varepsilon_{t-2} + \dots$$

Donde $\alpha + \delta(t+s)$ son la parte no estacionaria, con ello la esperanza de Y_t es diferente en cada momento del tiempo. A medida que el horizonte de predicción s crece, la sumabilidad absoluta de $\{\psi_j\}$ implica que la predicción converge en media cuadrática a la tendencia:

$$E \left[\hat{Y}_{t+s/t} - \alpha - \delta(t+s) \right]^2 \xrightarrow{s \rightarrow \infty} 0$$

Para predecir un proceso con una raíz unitaria debemos recordar que ΔY_t es un proceso estacionario, por lo que:

$$\Delta \hat{Y}_{t+s} = \hat{E}[(Y_{t+s} - Y_{t+s-1})/Y_t, Y_{t-1} \dots] = \delta + \underbrace{\psi_0 \varepsilon_{t+s} + \psi_1 \varepsilon_{t+s-1} + \dots}_{=0}$$

$$+ \psi_S \varepsilon_t + \psi_{S+1} \varepsilon_{t-1} + \psi_{S+2} \varepsilon_{t-2} + \dots$$

Podemos escribir a Y_{t+s} como el nivel de la variable más la suma de los cambios entre t y $t+s$

$$Y_{t+s} = (Y_{t+s} - Y_{t+s-1}) + (Y_{t+s-1} - Y_{t+s-2}) + \dots + (Y_{t+1} - Y_t) + Y_t$$

$$Y_{t+s} = \Delta Y_{t+s} + \Delta Y_{t+s-1} + \dots + \Delta Y_{t+1} + Y_t$$

$$\text{Así, } \hat{Y}_{t+s/t} = \Delta \hat{Y}_{t+s/t} + \Delta \hat{Y}_{t+s-1/t} + \dots + \Delta \hat{Y}_{t+1/t} + Y_t$$

$$\hat{Y}_{t+s/t} = (\delta + \psi_S \varepsilon_t + \psi_{S+1} \varepsilon_{t-1} + \psi_{S+2} \varepsilon_{t-2} + \dots)$$

$$+ (\delta + \psi_{S-1} \varepsilon_t + \psi_S \varepsilon_{t-1} + \psi_{S+1} \varepsilon_{t-2} + \dots) + \dots$$

$$\dots + (\delta + \psi_1 \varepsilon_t + \psi_2 \varepsilon_{t-1} + \psi_3 \varepsilon_{t-2} + \dots) + Y_t$$

Reordenando:

$$\hat{Y}_{t+s/t} = s\delta + (\psi_s + \psi_{s-1} + \psi_{s-2} + \dots)\varepsilon_t + (\psi_{s+1} + \psi_s + \psi_{s-1} + \dots)\varepsilon_{t-1} \\ + (\psi_{s+2} + \psi_{s+1} + \psi_s + \dots)\varepsilon_{t-2} + \dots + Y_t$$

Algunos aspectos interesantes pueden visualizarse a través de casos especiales:

Paseo aleatorio con deriva

$$Y_t = Y_{t-1} + \delta + \varepsilon_t \text{ ó } \\ \Delta Y_t = \delta + \varepsilon_t$$

En este ejemplo $\psi_1 = \psi_2 = \dots = 0$, y por lo tanto la predicción a s pasos sería:

$$\hat{Y}_{t+s/t} = s\delta + Y_t$$

Esto significa que dado Y_t se espera que crezca a una tasa constante δ por período.

Sea un **ARIMA (0,1,1)** con $\psi_1 = \theta, \psi_2 = \psi_3 = \dots = 0$, es decir que:

$$Y_t = \delta + Y_{t-1} + \theta \varepsilon_{t-1} + \varepsilon_t$$

$$\hat{Y}_{t+s/t} = s\delta + (\psi_s + \psi_{s-1} + \dots + \psi_1)\varepsilon_t + (\psi_{s+1} + \psi_s + \dots + \psi_2)\varepsilon_{t-1} + (\psi_{s+2} + \psi_{s+1} + \dots + \psi_3)\varepsilon_{t-2} + \dots + Y_t$$

Como $\psi_1 = \theta, \psi_2 = \psi_3 = \dots = 0$,

la predicción a s pasos resulta

$$\hat{Y}_{t+s/t} = s\delta + Y_t + \theta \varepsilon_t$$

En este ejemplo, el nivel y la innovación pasada y corriente de la serie, Y_t y ε_t respectivamente, definen la base a partir de la cual se espera que la serie crezca a una tasa constante δ por período.

Observaciones:

El parámetro δ en el proceso con raíz unitaria juega un papel similar al δ en los procesos estacionarios con tendencia determinística.

La diferencia fundamental entre las 2 especificaciones (TS y DS) es el punto de corte u ordenada en el origen. Para el proceso TS, la predicción converge a una recta cuyo punto de corte es el mismo que para Y_t . En cambio, la ordenada en el origen para la predicción de un proceso con raíz unitaria cambia con cada nueva observación.

Entre los procesos TS y DS hay diferencias importantes en sus implicaciones en la varianza del error de predicción.

Para un **proceso estacionario con tendencia determinística, TS**, el error de predicción s pasos adelante es:

$$Y_{t+s} - Y_{t+s/t} = [\alpha + \delta(t+s) + \varepsilon_{t+s} + \psi_1 \varepsilon_{t+s-1} + \psi_2 \varepsilon_{t+s-2} + \dots + \psi_{s-1} \varepsilon_{t+1} + \psi_s \varepsilon_t + \psi_{s+1} \varepsilon_{t-1} + \dots] \\ [\alpha + \delta(t+s) + \psi_s \varepsilon_t + \psi_{s+1} \varepsilon_{t-1} + \dots]$$

$$Y_{t+s} - \hat{Y}_{t+s/t} = \varepsilon_{t+s} + \psi_1 \varepsilon_{t+s-1} + \psi_2 \varepsilon_{t+s-2} + \dots + \psi_{s-1} \varepsilon_{t+1}$$

Por lo tanto, el error cuadrático medio (ECM) es:

$$E \left[Y_{t+s} - \hat{Y}_{t+s/t} \right]^2 = (1 + \psi_1^2 + \psi_2^2 + \dots + \psi_{s-1}^2) \sigma_\varepsilon^2$$

$$\lim_{s \rightarrow \infty} E \left[Y_{t+s} - \hat{Y}_{t+s/t} \right]^2 = (1 + \psi_1^2 + \psi_2^2 + \dots + \psi_{s-1}^2) \sigma_\varepsilon^2$$

La incertidumbre que se agrega a medida que la predicción es a más pasos es pequeña, y en el límite el error cuadrático medio del predictor es igual a la varianza incondicional del componente $\psi(L)$ estacionario.

En cambio, para el **proceso con una raíz unitaria** el error de predicción a s períodos adelante es:

$$Y_{t+s} - \hat{Y}_{t+s/t} = [\Delta Y_{t+s} + \Delta Y_{t+s-1} + \dots + \Delta Y_{t+1} + Y_t] - [\Delta \hat{Y}_{t+s/t} + \Delta \hat{Y}_{t+s-1/t} + \dots + \Delta \hat{Y}_{t+1/t} + Y_t]$$

$$Y_{t+s} - \hat{Y}_{t+s/t} = (\varepsilon_{t+s} + \psi_1 \varepsilon_{t+s-1} + \psi_2 \varepsilon_{t+s-2} + \dots + \psi_{s-1} \varepsilon_{t+1}) + (\varepsilon_{t+s-1} + \psi_1 \varepsilon_{t+s-2} + \dots + \psi_{s-2} \varepsilon_{t+1}) + \dots (\varepsilon_{t+1})$$

$$Y_{t+s} - \hat{Y}_{t+s/t} = \varepsilon_{t+s} + (1 + \psi_1) \varepsilon_{t+s-1} + (1 + \psi_1 + \psi_2) \varepsilon_{t+s-2} + \dots (1 + \psi_1 + \psi_2 + \dots + \psi_{s-1}) \varepsilon_{t+1}$$

La esperanza del error de predicción (ECM) es:

$$E \left[Y_{t+s} - \hat{Y}_{t+s/t} \right]^2 = \sigma_\varepsilon^2 \left[(1 + \psi_1)^2 + (1 + \psi_1 + \psi_2)^2 + (1 + \psi_1 + \psi_2 + \dots + \psi_{s-1})^2 \right]$$

A diferencia del proceso TS la esperanza del cuadrado del error de predicción no converge a un valor fijo, sino que se incrementa a medida que el horizonte de predicción se aleja.

Sin embargo, asintóticamente se aproxima a una función lineal de s con pendiente: $(1 + \psi_1 + \psi_2 + \dots)^2 \sigma_\varepsilon^2$

En resumen, en el proceso TS la esperanza del cuadrado del error de predicción recorre una banda finita a medida que el horizonte de predicción se aleja, mientras que en los procesos con una raíz unitaria crece linealmente con el horizonte de predicción.

OBS: Esta característica de los procesos con una raíz unitaria implica que el desvío estándar del error de predicción crece con la raíz cuadrada de s .

El intervalo de confianza al 95% para Y_{t+s} se expande más lentamente que el nivel de la serie: $\left[\hat{Y}_{t+s/t} = s\delta + Y_t \right]$ y el desvío crece con la \sqrt{s} .

- Esto significa que los datos con una RU y con una deriva positiva muestran una tendencia creciente si se los observa por un período suficientemente largo. La tendencia introducida por la deriva, asintóticamente domina la variabilidad creciente debido al componente de raíz unitaria.
- Un paseo aleatorio sin deriva no muestra una tendencia a retornar al valor inicial o a la media incondicional, un paseo aleatorio con deriva no muestra un comportamiento a retornar a una tendencia determinística, pero la serie está dominada asintóticamente por el término $\delta > 0$

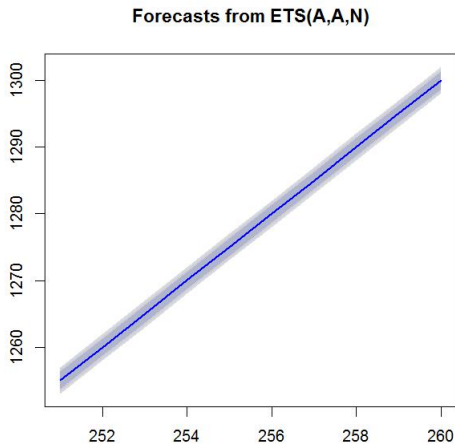


Figura: Intervalo de confianza de la predicción de un proceso TS

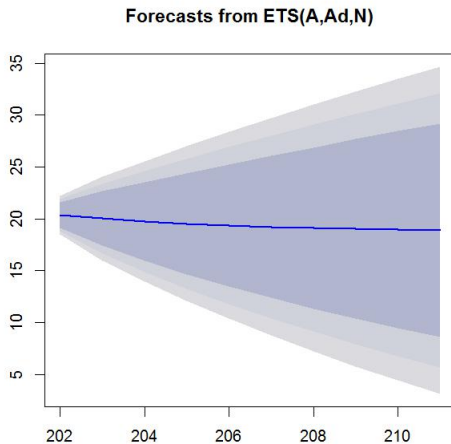


Figura: Intervalo de confianza de la predicción de un proceso DS

Otra diferencia entre los procesos TS y DS se ve en el efecto que las innovaciones tienen sobre la serie. Consideremos las consecuencias sobre Y_{t+s} si ε_t se incrementa en una unidad y el resto de las innovaciones no son afectadas.

En un proceso TS

$$\frac{\partial Y_{t+s}}{\partial \varepsilon_t} = \psi_s \text{ y } \lim_{s \rightarrow \infty} \frac{\partial Y_{t+s}}{\partial \varepsilon_t} = 0$$

Recordar que :

$$\hat{Y}_{t+s} = \alpha + \delta(t+s) + \psi_s \varepsilon_t + \psi_{s+1} \varepsilon_{t-1} + \psi_{s+2} \varepsilon_{t-2} + \dots,$$

por lo que **el efecto de cualquier perturbación estocástica tiende a desaparecer en el tiempo.**

En cambio, para un proceso DS

$$\hat{Y}_{t+S/t} = s\delta + (\psi_S + \psi_{S-1} + \dots + \psi_1)\varepsilon_t + (\psi_{S+1} + \psi_S + \dots + \psi_2)\varepsilon_{t-1} + (\psi_{S+2} + \psi_{S+1} + \dots + \psi_3)\varepsilon_{t-2} + \dots + Y_t$$

por lo que

$$\frac{\partial Y_{t+s}}{\partial \varepsilon_t} = \frac{\partial Y_t}{\partial \varepsilon_t} + \psi_s + \psi_{s-1} + \dots + \psi_1 \quad \underbrace{\quad}_{Y_t = Y_{t-1} + \delta + \varepsilon_t} = 1 + \psi_S + \psi_{S-1} + \dots + \psi_1$$

$$\lim_{s \rightarrow \infty} \frac{\partial Y_{t+s}}{\partial \varepsilon_t} = 1 + \psi_1 + \psi_2 + \dots = \psi(1)$$

Otra diferencia entre los procesos TS y DS es el tipo de transformación a aplicar para obtener la estacionariedad.

Si el proceso es TS, por ejemplo $Y_t = \alpha + \delta t + \psi(L)\varepsilon_t$

el tratamiento apropiado consiste en sustraerle la tendencia δt a la serie para obtener estacionariedad.

En cambio, si los datos fueron generados por un proceso con raíz unitaria:

$$\Delta Y_t = \delta + \psi(L)\varepsilon_t,$$

sustraer δt de Y quita la dependencia temporal de la media, pero no de la varianza.

Por ejemplo, si los datos fueron generados por un paseo aleatorio con deriva

$Y_t = \delta + Y_{t-1} + \varepsilon_t$, al sustraerle una tendencia determinística resulta

$$Y_t - \delta t = Y_0 + (\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t) \cong Y_0 + u_t$$

Demostración:

$$Y_t = \delta + Y_{t-1} + \varepsilon_t$$

$$Y_{t-1} = \delta + Y_{t-2} + \varepsilon_{t-1}$$

$$Y_{t-2} = \delta + Y_{t-3} + \varepsilon_{t-2}$$

Por lo tanto,

$Y_t = \delta + (\delta + Y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \Rightarrow Y_t = 2\delta + (\delta + Y_{t-3} + \varepsilon_{t-2}) + \varepsilon_{t-1} + \varepsilon_t$, y así sucesivamente, hasta que se obtiene

$Y_t = \delta t + Y_0 + (\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t)$. Si se define

$$u_t = \sum_{i=1}^t \varepsilon_i \Rightarrow V(u_t) = V\left(\sum_{i=1}^t \varepsilon_i\right) \underset{\text{indep}}{=} V(\varepsilon_1) + V(\varepsilon_2) + \dots + V(\varepsilon_t) = t\sigma_\varepsilon^2 \text{ La}$$

varianza de los residuos u_t es $t\sigma_\varepsilon^2$, la varianza crece con cada observación adicional.

De esta forma, sustraerle una tendencia lineal a un proceso con una raíz unitaria **no** es suficiente para producir estacionariedad.

En este caso, **el tratamiento correcto sería diferenciar la serie, y es por esta razón que los procesos son DS (estacionarios en diferencia)**.

Por otra parte, si un proceso TS es diferenciado el resultado es:

$$\begin{aligned} Y_t &= \alpha + \delta t + \psi(L)\varepsilon_t \\ -Y_{t-1} &= \alpha + \delta(t-1) + \psi(L)\varepsilon_{t-1} \\ \Delta Y_t &= \delta + (1-L)\psi(L)\varepsilon_t \end{aligned}$$

Esta es un serie estacionaria, pero se introdujo una raíz unitaria en la parte de medias móviles, y por tanto es un proceso no invertible.

- Box ,G.; Jenkins,G.; Reinsel,G. (1994) *Time Series Analysis. Forecasting and Control* .Ed. Prentice Hall. 3º Ed.
- Hamilton, J. (1994) - *Time Series Analysis*. Princeton University Press.
- Wei W. (2006) - *Time Series Analysis.Univariate and Multivariate Methods*. Ed. Pearson Education. 2º Ed.
- Nielsen,H. (2007) «Non stationary time series and unit root testing». Lectures Notes 5
- Clements,M.; Hendry,D. (2000) «Forecasting with Difference-stationary and Trend-stationary Models».