

Tabla de contenidos

Lista de figuras

Lista de tablas

inputchapters/theoretical $\textit{framework}$ /intro.tex

Capítulo 1

Preprocesamiento de los datos

1.1. Tipos de tendencia

Seguindo a ?, una serie temporal y_t se puede descomponer en la suma de un componente de tendencia, un componente estacionario y un ruido. El componente estacional se puede modelar empleando la metodología de *Box-Jenkins*. A su vez, la tendencia puede ser tanto determinística como estocástica.

Si se supone que una serie temporal varía siempre de un período a otro en una cantidad fija tal que $\Delta y_t = a_0$, entonces la solución para ésta ecuación lineal en diferencias es $y_t = y_0 + a_0 t$ donde y_0 es la condición inicial para el período cero. Por lo tanto, la solución para $\Delta y_t = a_0$ es una tendencia lineal determinística y y_0 es el intercepto y a_0 la pendiente. Si se agrega a lo anterior un componente estacionario $A(L)\varepsilon_t$, se tiene

$$y_t = y_0 + a_0 t + A(L)\varepsilon_t \quad (1.1)$$

Cuando se tiene este tipo de modelos, se dice que son estacionarios alrededor de una tendencia (TS¹).

Si se supone ahora que el cambio esperado en y_t es de a_0 unidades más un término de ruido blanco tal que

$$\Delta y_t = a_0 + \varepsilon_t \quad (1.2)$$

Y si y_0 es la condición inicial, entonces la solución general para la ecuación en diferencias de primer orden representada por (??) es

¹TS por sus siglas en inglés para Trend Stationary

$$y_t = y_0 + \sum_{i=1}^t \varepsilon_i + a_0 t$$

Se dice que la secuencia y_t tiene una tendencia estocástica porque cada perturbación ε_i imparte un cambio permanente y aleatorio en la media condicional de las series.

1.1.1. Eliminación de la tendencia

Cuando una serie temporal es estacionaria, los efectos de las perturbaciones son temporales, es decir, con el tiempo sus efectos desaparecen y la serie vuelve a su nivel de media de largo plazo. Por otro lado, aquellas series temporales con tendencia estocástica no van a volver a niveles de largo plazo. Una tendencia puede contener componentes tanto estocásticos como deterministas. Éstos componentes tienen implicancias a la hora de transformar a las series temporales para que sean estacionarias. Los métodos usuales consisten en la diferenciación y remoción de la tendencia².

Cuando una serie contenga raíces unitarias, se podrá volver estacionaria con el método de diferenciación. En este sentido, la d –ésima diferencia de un proceso con d raíces unitarias es estacionario y la misma es integrada de orden d y se expresa como $I(d)$.

Cuando una serie contenga raíz unitaria, se dice que estacionaria en diferencias (DS)³. En la literatura se ha encontrado que, en general, las variables macroeconómicas tienden a ser DS más que TS.

² *Detrending en inglés*

³ DS: difference stationary

1.2. Ajuste estacional

Para realizar el ajuste estacional de las series temporales, se va a emplear un procedimiento de descomposición de las mismas basado en Loess⁴ y denomina STL⁵.

El STL es un procedimiento de filtrado que permite estimar los componentes de tendencia, estacionalidad y de residuo de una serie temporal usando el suavizador de Loess. En este sentido, $Y_v = T_v + S_v + R_v$, como Y_v el conjunto de datos, T_v la tendencia, S_v el componente estacional y R_v el residuo, para $v = 1, \dots, N$.

Dado que el procedimiento tiene como base el suavizador de Loess, se empieza por tomar unos puntos que pueden ser x_i y y_i , medidas de una variable independiente y otra dependiente, respectivamente, para $i = 1, \dots, n$. La curva de regresión Loess $\hat{g}(x)$ es un suavizador de y dado x , para cualquier valor de x , aún habiendo datos faltantes⁶. Considerando lo anterior, se calcula $\hat{g}(x)$ eligiendo un entero positivo q , que por ahora se supone $q \leq n$ y q representa los valores de x_i cercanos a x . A cada x_i se le da un peso de vecindario que se basa en su distancia con respecto a x . Para ésto también se define $\lambda_q(x)$ como la distancia del x_i más lejano en la q -ésima posición respecto a x .

Sea W una función de peso tricubo tal que

$$W(u) = \begin{cases} (1 - u^3)^3 & \text{para } 0 \leq u < 1 \\ 0 & \text{para } u \geq 1 \end{cases}$$

Y el peso del vecindario para cada x_i es

$$v_i(x) = W\left(\frac{|x_i - x|}{\lambda_q(x)}\right)$$

Por lo que $u = \frac{|x_i - x|}{\lambda_q(x)}$. En este sentido, los x_i más cercanos a x tienen mayor peso. Los pesos van a decrecer a medida que x_i se aleje de x y serán iguales a cero en el q -ésimo punto más alejado.

Luego, se ajusta a los datos un polinomio de grado d con peso $v_i(x)$ en (x_i, y_i) ⁷. De esta manera, el valor del polinomio ajustado localmente en x es

⁴ Acrónimo en inglés que refiere a Locally Weighted Scatterplot Smoothing que en español sería Suavizado de Dispersión Ponderado Localmente

⁵ Acrónimo en inglés para Seasonal-Trend Decomposition procedure base on Loess que en español sería Procedimiento de Descomposición Estacional-Tendencial basado en Loess.

⁶ Ésto le confiere ciertas ventajas frente a otros métodos.

⁷ Cuando $d = 1$, es un polinomio localmente lineal y cuando $d = 2$ es cuadrático.

$\hat{g}(x)$.

Si ahora $q > n$ y $\lambda_n(x)$ es la distancia del x_i más alejado a x , se define

$$\lambda_q(x) = \lambda_n(x) \frac{q}{n}$$

Y se sigue como en el caso anterior, con la definición de los pesos de los vecindarios usando los valores de $\lambda_n(x)$.

Tanto los valores q como d deben ser elegidos y a medida que q aumenta, $\hat{g}(x)$ es más suave. Por otra parte, es razonable emplear $d = 1$, en el polinomio, cuando los patrones subyacentes de los datos sean una curva suave mientras que cuando se tenga una curvatura pronunciada, con picos y valles, $d = 2$ es una mejor opción. Adicionalmente, se supone que cada observación (x_i, y_i) tiene un peso ρ_i que expresa la fiabilidad de la observación respecto a otras, y bajo determinados supuestos se puede expresar de distante forma.

Considerando lo anterior, se desarrolla el procedimiento STL que consiste en dos procesos de tipo recursivo: un bucle interno anidado dentro de un bucle externo. En cada pasada del bucle interno, se actualizan tanto el componente de tendencia como el estacional. Cada corrida completa del bucle interno consiste en $n_{(i)}$ pasadas. Por su parte, cada pasada del bucle externo consiste en la pasada del bucle interno seguido del cálculo de los pesos de robustez. Estos pesos se usan en la próxima pasada del bucle interno para reducir la influencia de comportamientos pasajeros y erróneos en los componentes mencionados. Además, la pasada inicial del bucle externo se llevará a cabo con todos los pesos de robustez iguales a uno y luego se llevan a cabo $n_{(0)}$ pasadas del bucle externo. A su vez, se supone que la cantidad de observaciones del componente estacional, en cada período o ciclo, es $n_{(p)}$. Si la serie es mensual con periodicidad anual entonces $n_{(p)} = 12$. Se definen a las $n_{(p)}$ subseries como subseries-ciclo.

Con respecto al *bucle interno*, consiste en un suavizador estacional que actualiza primero a ese componente, seguido por la suavización y actualización de la tendencia. Se supone que $S_v^{(k)}$ y $T_v^{(k)}$, para $v = 1, \dots, N$, son los componentes estacional y de tendencia al final de la k -ésima pasada. Las actualizaciones en el $(k+1)$ -ésimo paso se calculan siguiendo una serie de pasos. El primer paso es el *Detrending*⁸, o sea, se remueve la tendencia $Y_v - T_v^{(k)}$. Cuando haya valores faltantes también los habrá en la misma posición para éste cálculo. El conjunto de valores suavizados para todos las subseries-ciclo es una serie estacional temporaria $C_v^{(k+1)}$ con $N + 2n_{(p)}$ valores que varían entre $v = -n_{(p)} + 1$ y $v = N + n_{(p)}$. En el segundo paso se emplea un suavizador de las subseries-ciclo. Cada subserie-ciclo de la serie a la que se removió la

⁸Remoción de la tendencia

tendencia es suavizada por Loess con $q = n_{(s)}$ y $d = 1$. En el tercer paso se hace el filtrado de paso-bajo de las subseries-ciclo suavizadas. El filtrado se aplica sobre $C_v^{(k+1)}$ y consiste en una media móvil de largo $n_{(p)}$ seguida por otra del mismo largo, luego por una de largo 3, y luego por un suavizador loess con $d = 1$ y $q = n_{(l)}$. El resultado es $L_v^{(k+1)}$ y se define para las posiciones $v = 1, \dots, N$. El cuarto paso consiste en la remoción de la tendencia de las subseries-ciclo. Se sustrae el componente estacional en el $(k + 1)$ -ésimo bucle $S_v^{(k+1)} = C_v^{(k+1)} - L_v^{(k+1)}$ para $v = 1, \dots, N$. El quinto paso consiste en el ajuste estacional, o sea, se calcula como $Y_v - S_v^{(k+1)}$. En el sexto paso se suaviza la tendencia. La serie con ajuste estacional es suavizada con loess para $q = n_{(t)}$ y $d = 1$. El componente de tendencia para el $(k + 1)$ -ésimo bucle es $T_v^{(k+1)}$ con $v = 1, \dots, N$.

Con respecto al *bucle externo*, una vez que se haya realizado una corrida inicial del bucle interno, se obtienen unas estimaciones para T_v y S_v . Con esto es posible calcular el componente residual $R_v = Y_v - T_v - S_v$. Se define un peso para cada punto en el que se observe Y_v . Estos pesos de robustez reflejan cuán extremo es R_v . Un dato atípico que devuelva un $|R_v|$ grande tendrá un peso bajo o igual a cero. Sea $h = 6 \cdot \text{mediana}(|R_v|)$, entonces el peso de robustez para un punto en el tiempo v será $w_v = B(|R_v|/h)$, donde B es la función de pesos bi-cuadrada tal que

$$B(u) = \begin{cases} (1 - u^2)^2 & \text{para } |u| < 1 \\ 0 & \text{para } |u| \geq 1 \end{cases}$$

Ahora, se vuelve a repetir el bucle interno, pero en los pasos 2 a 6, los pesos del vecindario para un valor en el momento v es multiplicado por w_v . Las iteraciones de robustez para el bucle exterior se llevan a cabo $n_{(0)}$ veces.

Con respecto a la elección de los parámetros, en total STL cuenta con 6: $n_{(p)}$ es la cantidad de observaciones en cada ciclo del componente estacional; $n_{(i)}$, la cantidad de pasadas en el bucle interno; $n_{(0)}$, la cantidad de iteraciones de robustez en el bucle exterior; $n_{(l)}$, el parámetro de suavizamiento en el filtro de paso-bajo; $n_{(t)}$, el parámetro de suavizamiento para el componente de tendencia; $n_{(s)}$, el parámetro de suavizamiento para el componente estacional⁹.

Una vez estimado el componente estacional mediante STL, se va a realizar el ajuste estacional de cada una de las series en primeras diferencias. Se elige este proceso en vez de realizar una diferenciación estacional para no perder observaciones.

⁹Los autores plantean que la selección de los primeros 5 es directa mientras que el último parámetro se elige según el caso de estudio.

1.3. Pruebas de raíces unitarias y de estacionariedad

1.3.1. Pruebas de Dicky-Fuller Aumentado

En la literatura existen varios métodos para testear la presencia de raíces unitarias en series temporales. Sin embargo, en el presente estudio se va a considerar la prueba de Dicky-Fuller Aumentado (ADF¹⁰) para poner a prueba la hipótesis nula de existencia de raíces unitarias en los componentes del IPC, es decir, para cada una de las series se pone a prueba

$$H_0) \gamma = 0$$

$$H_1) \gamma \neq 0$$

En este sentido, se consideran tres ecuaciones de regresión distintas para poner a prueba H_0 ,

$$\Delta y_t = \gamma y_{t-1} + \sum_{i=2}^p \beta_i \Delta y_{t-i+1} + \varepsilon_t \quad (1.3)$$

$$\Delta y_t = a_0 + \gamma y_{t-1} + \sum_{i=2}^p \beta_i \Delta y_{t-i+1} + \varepsilon_t \quad (1.4)$$

$$\Delta y_t = a_0 + \gamma y_{t-1} + a_2 t + \sum_{i=2}^p \beta_i \Delta y_{t-i+1} + \varepsilon_t \quad (1.5)$$

Se van a emplear tres estadísticos τ , τ_μ , τ_τ para probar la hipótesis nula $H_0) \gamma = 0$ en cada caso y además, se tienen otros tres F -estadísticos ϕ_1, ϕ_2, ϕ_3 para hacer pruebas conjuntas sobre los coeficientes.

Los estadísticos ϕ_1, ϕ_2, ϕ_3 se construyen como pruebas F :

$$\phi_i = \frac{[SSR_{restringido} - SSR_{no restringido}] / r}{SSR_{no restringido} / (t - k)}$$

donde SSR es la suma de los cuadrados de los residuos en los modelos restringidos y no restringidos, r es la cantidad de restricciones, T es la cantidad de observaciones, k es la cantidad de parámetros estimados en el modelo irres-

¹⁰ADF por sus siglas en inglés.

tricto, $i = 1, 2, 3$. A su vez, $T - k$ van a ser los grados de libertad del modelos sin restricciones. Los valores de los coeficientes estimados se van a comparar con los valores críticos de tablas reportados por Dickey y Fuller(1981). La hipótesis nula indica que el proceso de generación de los datos es la del modelo restringido contra la hipótesis alternativa de que los datos son generados por el modelo sin restringir. Cuando los valores de ϕ_1, ϕ_2, ϕ_3 sean mayores a los valores críticos reportados por Dickey y Fuller(1981) se rechaza la hipótesis nula, cuando sean menores a los valores críticos entonces no se rechaza la hipótesis nula.

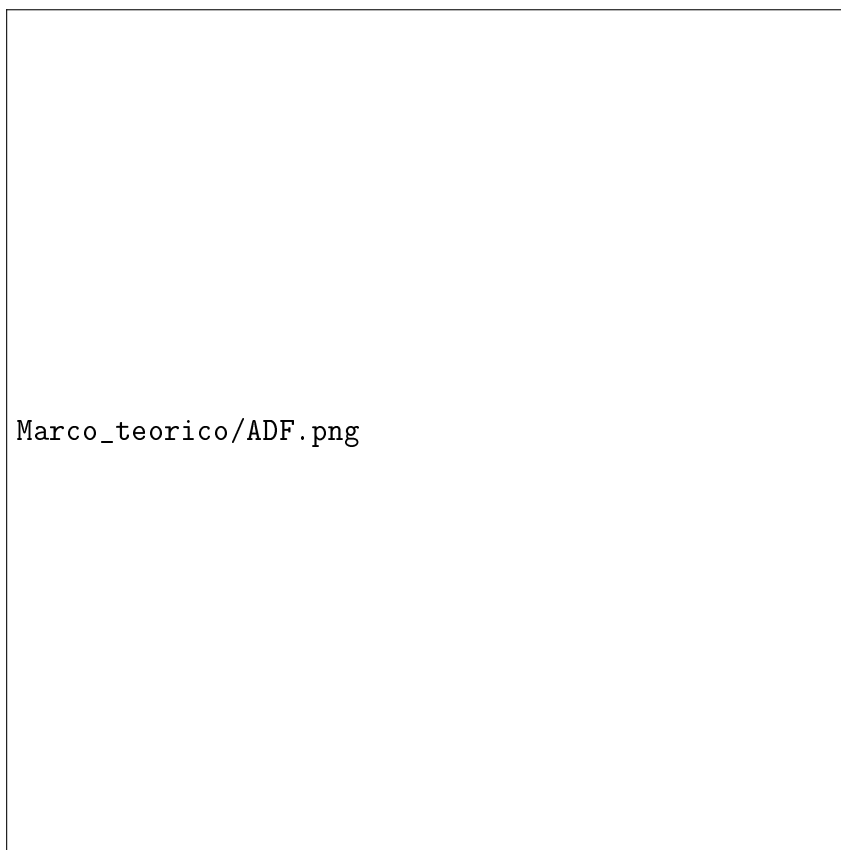
En el siguiente cuadro ?? se consideran los tres modelos y cada una de las hipótesis a testear con sus respectivos estadísticos.

Tabla 1.1: Modelos del test ADF

	Modelo	H_0)	Estadístico de prueba
c	$\Delta y_t = a_0 + \gamma y_{t-1} + a_2 t + \sum_{i=2}^p \beta_i \Delta y_{t-i+1} + \varepsilon_t$	$\gamma = 0$ $\gamma = a_2 = 0$ $\gamma = a_2 = a_0 = 0$	τ_τ ϕ_3 ϕ_2
b	$\Delta y_t = a_0 + \gamma y_{t-1} + \sum_{i=2}^p \beta_i \Delta y_{t-i+1} + \varepsilon_t$	$\gamma = 0$ $\gamma = a_0 = 0$	τ_μ ϕ_1
a	$\Delta y_t = \gamma y_{t-1} + \sum_{i=2}^p \beta_i \Delta y_{t-i+1} + \varepsilon_t$	$\gamma = 0$	τ

Siguiendo el manual complementario de Enders(año), se sugiere seguir una regla que ayuda a elegir el conjunto apropiado de regresores de la ecuación de estimación: si el proceso generativo de los datos contiene algún regresor determinístico y la ecuación de estimación también los contiene, se pueden realizar inferencias sobre todos los coeficientes con una prueba t o F . Esto porque una prueba que involucre una restricción entre parámetros con distintas tasas de convergencia es dominada asintóticamente por los parámetros con menores tasas de convergencia. En este sentido, las distribuciones Dickey– Fuller no estándar son necesarias sólo cuando se incluyen regresores determinísticos. Sin embargo, si se incluyen componentes superflúos, la prueba va a perder poder. Además, cuando no se conozca el proceso de generación de los datos, se sugiere seguir un procedimiento y realizar las pruebas de Dickey-Fuller Aumentado partiendo del modelo menos restrictivo en cada caso a uno más particular.

Figura 1.1: Procedimiento de la prueba ADF



Enders (año), aclara que no es esperable que se encuentren buenos resultados aplicando el procedimiento de manera mecánica y recomienda apoyarse en herramientas visuales para tomar conclusiones sobre la posible existencia de regresores determinísticos. De forma adicional, como los coeficientes a_0 y a_2 juegan distintos roles bajo las hipótesis nula y alternativa se podrían generar confusiones a la hora de interpretar los parámetros. Si bien las pruebas de ADF son útiles para detectar la presencia de raíces unitarias, los mismos tienen sus limitaciones. En cada paso, cada prueba está condicionada a que las pruebas anteriores sean correctas.

Cuando se empieza por el primer paso, es decir, con el modelo (c) con constante y con tendencia, se hace más difícil rechazar H_0), por lo tanto, cuando se rechace la hipótesis nula en un modelo (c) se tiende a rechazar también la hipótesis nula cuando no se incluyan los términos deterministas.

A su vez, establece que el problema principal de las pruebas de Dickey-Fuller es que tanto el intercepto como la pendiente de la tendencia son, con frecuencia, estimados de manera "pobre" bajo la presencia de raíces unitarias. Incluso cuando la serie sea estacionaria los resultados de las estimaciones de éstos parámetros pueden ser débiles si la serie es persistente. En la medida que las estimaciones de a_0 y a_1 contengan errores importantes, lo mismo va a ocurrir con la estimación de γ que probablemente tendrá un error estándar grande también. El autor plantea que se puede observar lo anterior al comparar los valores críticos de Dickey-Fuller para τ , τ_μ y τ_τ con los de una tabla t —estándar. Los intervalos de confianza demasiado amplios para γ implican que se tiende a no rechazar la hipótesis nula de raíz unitaria incluso cuando el verdadero valor de γ no es cero.

Además, la prueba presenta limitaciones también frente a cambios de régimen.

1.3.2. Pruebas de Kwiatkowski, Phillips, Schmidt y Shin

Para hacer frente a las debilidades del test de Dickey-Fuller se han propuesto en la literatura varias alternativas.

Siguiendo a Kwiatkowski et al (1992), desde el punto de vista empírico, para series económicas, en general, las pruebas de raíz unitaria fallan en rechazar la hipótesis nula de raíz unitaria cuando en realidad contienen raíces unitarias. En este sentido, se va a emplear una prueba alternativa, denominada KPSS, con hipótesis nula de estacionariedad contra una alternativa de raíz unitaria para cada serie temporal.

Se toma una representación de los componentes donde cada serie es la suma de una tendencia determinística, un paseo aleatorio y un error estacionario, y se pone a prueba

H_0) la serie es estacionaria alrededor de una tendencia

que se corresponde con la hipótesis de que la varianza del paseo aleatorio (*random walk*) es igual a cero.

Se emplea un estadístico de Multiplicadores de Lagrange (ML) para testear la hipótesis nula de estacionariedad. De esta manera, siendo y_t con $t = 1, 2, \dots, T$ las series a las que se les quiere aplicar el test, se asume que se puede descomponer a la serie en la suma de un componente de tendencia determinística, un paseo aleatorio y un error estacionario se tiene que,

$$y_t = \xi t + r_t + \varepsilon_t \quad (1.6)$$

Donde r_t es un paseo aleatorio:

$$r_t = r_{t-1} + u_t, \quad (1.7)$$

donde u_t es $iid(0, \sigma_u^2)$. El valor inicial r_0 es fijo y sirve de intercepto. La hipótesis de estacionariedad es $\sigma_u^2 = 0$ y como se asume que ε_t es estacionario, bajo la hipótesis nula y_t es estacionaria alrededor de una tendencia.

En el caso particular de que en el modelo (??) se tenga $\xi = 0$, bajo la hipótesis nula y_t va a ser estacionaria alrededor de una constante (r_0).

Sean e_t con $t = 1, 2, \dots, T$, los residuos de la regresión y con un intercepto y tendencia. A su vez, sea $\hat{\sigma}_\varepsilon^2$ la estimación del error de la varianza de la regresión (suma de los residuos al cuadrado dividida T). Con lo anterior, se define el proceso de suma parcial de los residuos como

$$S_t = \sum_{i=1}^t r_i, \quad t = 1, \dots, T \quad (1.8)$$

Entonces el estadístico ML es

$$ML = \sum_{t=1}^T S_t^2 / \hat{\sigma}_\varepsilon^2 \quad (1.9)$$

En el caso de que se quiera poner a prueba la hipótesis nula de estacionariedad alrededor de una constante se define e_t como los residuos de la regresión y sobre un intercepto ($e_t = y_t - \bar{y}$).

Cabe resaltar que es una prueba de cola superior y se reportan los valores críticos. Además, para este caso se asume que los errores ε_t son $iidN(0, \sigma_\varepsilon^2)$. Sin embargo, se puede extender la prueba con supuestos más débiles sobre la distribución de los errores dado que el supuesto anterior puede ser poco realista.