Modelos univariados para procesos no estacionarios (en media y varianza) - Raíces unitarias

Fernando Borraz

Extracción de señales - dECON - 2017

Introducción

- Desde un punto de vista práctico el supuesto de estacionariedad puede ser muy fuerte
- Varias de las variables económicas mas utilizadas en la práctica (PBI, consumo, precios, dinero, etc) presentan tendencia en sus niveles y en ocasiones en la varianza
- Removiendo tendencias simplemente tomando primeras diferencia o realizando una regresión contra una tendencia lineal
- Ahora lo que veremos será determinar de manera exacta cuando usar cada método para remover tendencias y que las series sean estacionarias y así poder aplicar el teorema de Wold

Introducción

- Razones para desagregar una serie de tiempo económica en sus componentes inobservables (Espasa y Cancelo 1993)
- "El análisis de la coyuntura de un fenómeno económico normalmente no se realiza directamente a través de la evolución de sus datos originales, ya que éstos contienen muchas oscilaciones de escaso interés económico que pueden llevar a conclusiones equivocadas. Por el contrario, dicho análisis se efectúa a través de señales extraídas de los datos, que sirven de estimación de lo que un analista puede considerar como un aspecto esencial del fenómeno en cuestión."

Introducción

- Los datos originales de una serie económica presentan grandes oscilaciones, muchas de las cuales no son de interés desde el punto de vista del análisis económico. Así, los aspectos esenciales de un fenómeno económico no son directamente observables, con lo cual es necesario disponer de procedimientos que permita cuantificarlos a partir de la información disponible (los datos originales)

- Al analista corresponde definir cuál o cuáles componentes o señales son de interés

Componentes inobservables de una serie económica

- i) Tendencia (l_t): evolución subyacente de la serie o de largo plazo. Es la evolución firme que hay detrás de la trayectoria

- ii) Estacionalidad (s_t): oscilaciones sistemáticas a lo largo del año. Tiene periodicidad de un múltiplo del año. Son oscilaciones regulares y sistemáticas dentro del año

- iii) Ciclo (c_t): desviaciones sistemáticas que no son estacionales. Son menos sistemáticas y pueden durar entre 2 y 5 años

- iv) Irregular(ε_t): schock aleatorio. Son ruidos de corto plazo. Son no sistemáticas (ruido blanco)

Componentes inobservables de una serie económica

- Por lo que podemos descomponer una serie observada Y en cuatro componentes no observables $y_t = l_t + c_t + s_t + \varepsilon_t$

- O si la serie es multiplicativa:
$y_t = l_t c_t s_t \varepsilon_t \Rightarrow Ln(y_t) = Ln\left(l_t\right) + Ln\left(c_t\right) + Ln\left(s_t\right) + Ln\left(\varepsilon_t\right)$

- Veremos procedimientos para extraer señales, es decir, para cuantificar los elementos no observables l, c, s, ε

- En muestras pequeñas es difícil separar tendencias de ciclos

- Nuestro objetivo puede ser estudiar ser la tendencia o algún otro de los componentes no observados. Por ejemplo c es lo que se estudian los modelos RBC (real bussiness cycle)

Componentes inobservables de una serie económica

- Desde el punto de vista metodológico, existe un incentivo por utilizar procedimientos que tengan propiedades conocidas y deseables desde el punto de vista teórico para analizar las fluctuaciones macroeconómicas

- Este aspecto implica contar con procedimientos de descomposición que especifiquen las características del componente que se desea estudiar y que sea posible su estimación a partir del análisis de los datos originales

Componentes inobservables de una serie económica

- Extracción de señales es una expresión tomada del campo de la ingeniería: los sistemas de audio tienen un parámetro, el denominado "coeficiente de señal ruido", que define la calidad de los sonidos emitidos

- En este campo, la extracción de señales se realiza a partir de un sistema que filtra la "señal de interés" (la música) eliminando el ruido (distorsiones que afectan al sistema de audio)

- El punto de partida de los procedimientos de extracción de señales es, entonces, que en la práctica la información que proporciona la observación y medición de un fenómeno (económico) se encuentra contaminada. Por ello, es necesario "filtrar" los datos de modo de recuperar la "señal" que ellos contienen

- Observación: La señal "pura" nunca se observa, por lo que siempre se está expuesto a cometer errores

Componentes inobservables de una serie económica

- El tema metodológico planteado no es en absoluto trivial y se traduce básicamente en un problema de filtrado de los datos originales

- Un filtro es una combinación lineal de las observaciones (originales) de una variable para distintos momentos del tiempo, que se realiza con la finalidad de obtener una estimación de una señal "deseada" por el analista

- Su aplicación siempre distorsiona el proceso original al que se aplica. Muchas veces dichas distorsiones no son precisamente “deseadas”, pudiendo conducir a interpretaciones equivocadas sobre el comportamiento de la variable analizada y a estimaciones sesgadas de sus componentes inobservables. Por esta razón, resulta importante contar con herramientas que permitan analizar las propiedades estadísticas de dichas transformaciones

Tendencias

- Dado que muchas veces nos encontramos con muestras pequeñas se nos hace difícil separar tendencias de ciclos
- Lo que para una muestra pequeña puede ser una tendencia, tal vez en una muestra más grande sea un ciclo
- Por lo anterior suponemos que l_t represente a la tendencia y al ciclo
- Luego si nos interesa separar el ciclo aplicaremos un filtro (puede ser HP)
- Por lo que la serie nos queda $y_t = l_t + s_t + \varepsilon_t$

Tendencias

- Hay dos maneras de describir tendencias: determinísticas y estocásticas
- Modelos de tendencias determinísticas: tendencia se modela como un polinomio función del tiempo
- Modelos de tendencias estocáticas: una alternativa más flexible a la anterior es modelar la serie como estacionaria en primera diferencia. La serie no es estacionaria por presentar tendencia pero su primera diferencia si es estacionaria

Modelos de tendencia determinística (Trend stationary, TS)

- Se asume la siguiente descomposición: $y_t = l_t + s_t + \varepsilon_t$

- l representa una función determinística de tendencia; s es una función determinística de estacionalidad y ε es un proceso puramente aleatorio de Y_t (ε_t puede ser un modelo ARMA)

- Por ejemplo, las variables que presentan tendencia determinística se pueden modelar como:

$$y_t = l_t + s_t + \frac{(1+\theta_1 L+\theta_2 L^2 \ldots.+\theta\phi q L^q)}{(1-\phi_1 L-\phi_2 L^2 \ldots.-\phi_p L^P)}\varepsilon_t = \alpha + \beta t + \psi(L)\varepsilon_t$$

con α, β parámetros y $\psi(L)\varepsilon_t$ es la representación en medias moviles $(\sum_{j=0}^{\infty} \mid \psi_j \mid < \infty)$

Modelos de tendencia determinística (trend stationary, TS)

- $E(y_t) = \alpha + \beta t$ no constante por lo que y_t no estacionario
- ¿Por qué la tendencia es lineal si en general las series económicas presentan tendencia exponencial? $y_t = e^{\beta t}$
- $\frac{dy}{dt} = \beta y_t$ y el crecimiento es una fracción constante
- Ademas $Ln\left(e^{\beta t}\right) = \beta t$
- Por esto es conveniente tomar logaritmo a las variables

Modelos de tendencia determinística

- Dado que $y_t - l_t - s_t$ es un proceso estacionario suele denominarselo como estacionario en tendencia, trend stationary (TS)

- Para remover la tendencia debemos obtener $y_t^* = y_t - \hat{\alpha} - \hat{\beta} t$

- Una vez que eliminamos la tendencia y la estacionalidad a los datos, la serie resultante es estacionaria y se puede analizar con las herramienta que vimos de procesos estacionarios

Modelos de tendencia determinística

- Supongamos por el momento que la serie no presenta estacionalidad (más adelante veremos explicitamente el tema de la estacionalidad)

- Ej: Sea $y_t = l_t + \varepsilon_t \qquad$ con $\varepsilon_t \sim N\left(0, \sigma^2\right)$ y l_t es una función determinística del tiempo

- Sea $l_t = \alpha + \beta t$ como ejemplo de una tendencia lineal determinística$\Rightarrow$

- $y_{t+h} = \alpha + \beta(t+h) + \varepsilon_{t+h} \qquad h \geq 0 \Rightarrow$

- $E_t(y_{t+h|t}) = \alpha + \beta(t+h) \qquad h \geq 1$

- Pero este modelo no es adecuado para realizar predicciones

Modelos de tendencia determinística

- $E_t(y_{t+h|t}) = E_{t+1}(y_{t+h|t}) = ... = \alpha + \beta(t+h) \qquad h > 1$
- Por lo que la nueva información del período t+1 no cambia la predicción de y_{t+h} pues solo es una función determinística del tiempo

Modelos con tendencias estocástica (difference stationary, DS)

- Una alternativa más flexible es modelar la serie como estacionaria en primera diferencia o como un proceso con raíz unitaria:

- $y_t = \alpha + y_{t-1} + \psi(L)\varepsilon_t$

- De nuevo es útil usar variables en logaritmo de modo que los cambios en las variables representen tasas de crecimiento

- $Ln(y_t) - Ln\left(y_{t-1}\right) = Ln(\frac{y_t}{y_{t-1}}) = Ln\left(1 + \frac{y_t - y_{t-1}}{y_{t-1}}\right) \approx \frac{y_t - y_{t-1}}{y_{t-1}} \qquad (\text{pues } lim_{x\to 1}\, Ln\left(x\right) = x - 1)$

- Notar que esto solo vale para cambios pequeños

Modelos con tendencias estocástica

- Veremos que existen diferencias muy importantes con los modelos TS
- $y_t = \alpha + y_{t-1} + \psi(L)\varepsilon_t$
- Dicho proceso no es estacionario pues tiene una raíz unitaria. También se denominan procesos integrados. Se trata de un proceso integrado de orden 1 (pues solo una raíz del polinomio AR es 1)
- Se denota I(1). Notar que si es estacionaria la primera diferencia de la serie

Modelos con tendencias estocástica

- El modelo se puede reescribir de la siguiente forma considerando diferencias:

  $y_t^* = \Delta y_t = (1 - L)y_t = \delta + \psi(L)\varepsilon_t$ y este proceso es estacionario

- Si hay dos raíces igual es a 1: $(1 - L)^2 y_t = \alpha + \acute{\psi}(L)\varepsilon_t$ es estacionario

- Esto es, la segunda diferencia es estacionaria. Sin embargo, y_t e

  $(1 - L)y_t$ no son estacionarios

Comparación TS-DS

- i) Predicciones
- ii) Errores de predicción
- iii) Persistencia de los shocks
- iv) Transformación necesaria para lograr estacionariedad. Eliminación de la tendencia

Comparación TS-DS: i) predicción

- Para los procesos TS en tendencia podemos usar la representación de medias móviles

- $y_t = \alpha + \beta t + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + ..$

- $y_{t+h} = \alpha + \beta\,(t+h) + \varepsilon_{t+h} + \psi_1 \varepsilon_{t+h-1} + \psi_2 \varepsilon_{t+h-2} + .. + \psi_{h-1} \varepsilon_{t+1} + \psi_h \varepsilon_{t+h-h} + \psi_{h+1} \varepsilon_{t+h-(h+1)} + ..$

- Podemos escribir la predicción como:
  $\hat{y}_{t+h|t} = \alpha + \beta(t+h) + \psi_h \varepsilon_t + \psi_{h+1} \varepsilon_{t-1} + ..$

- La predicción tiene dos componentes:

  - i) el crecimiento tendencial de la serie $\beta(t+h)$ y

  - ii) el efecto de los shocks que aún están actuando en el instante t (ε_t , ε_{t-1}....) pero ponderado por el impacto que van a tener en el instante h (ψ_h, ψ_{h+1}....)

Comparación TS-DS: i) predicción

- Como se cumple la condición de absolute summability

  $$E\left[\hat{y}_{t+h|t} - \alpha - \beta(t+h)\right]^2 \rightarrow 0 \qquad si \;\; h \rightarrow \infty$$

- Esto es, en el largo plazo, cuando $h \longrightarrow \infty$ los shocks que están presentes en t van a desaparecer y la predicción de la variable es únicamente la parte determinística

Comparación TS-DS: i) predicción

- Para los procesos DS podemos usar la representación de MA para la primera diferencia

- $(1-L)y_t = \alpha + \psi(L)\,\varepsilon_t \Rightarrow y_t = \alpha + y_{t-1} + \psi(L)\,\varepsilon_t$

- $y_{t+h} = \alpha + y_{t+h-1} + \psi(L)\varepsilon_{t+h}$ expresamos la predicciómo como

- $\hat{y}_{t+h|t} = y_t + h\alpha + (\psi_h + \psi_{h-1} + ... + \psi_1)\varepsilon_t + (\psi_{h+1} + \psi_h + ... + \psi_2)\varepsilon_{t-1} + (\psi_{h+2} + \psi_{h+1} + ... + \psi_3)\varepsilon_{t-2} + ..$

- Si h=1: $\hat{y}_{t+1|t} = y_t + \alpha + \psi_1\varepsilon_t + \psi_2\varepsilon_{t-1} + ..$

Comparación TS-DS: i) predicción

- Por lo que la predicción de la serie en t+1 para el caso de variables integradas depende de la posición inicial de la serie, y_t (diferente en el modelo TS)

- Otra diferencia es que aquí el intercepto de la serie no es fijo como en el modelo estacionario en tendencia (α)

- Otra gran diferencia es que aquí la historia de las innovaciones afecta en forma permanente el nivel de las series,mientras que en el modelo estacionario esto no es así. Para los procesos estacionarios en diferencias podemos usar la representación de MA para la primera diferencia

Comparación TS-DS: ii) errores de predicción

- Para el caso de variables TS el error de predicción es la suma ponderada de las innovaciones futuras

- $y_{t+h} = \alpha + \beta\,(t+h) + \varepsilon_{t+h} + \psi_1 \varepsilon_{t+h-1} + \psi_2 \varepsilon_{t+h-2} + .. + \psi_{h-1}\varepsilon_{t+1} + \psi_h \varepsilon_{t+h-h} + \psi_{h+1}\varepsilon_{t+h-(h+1)} + ..$

- $\hat{y}_{t+h|t} = \alpha + \beta(t+h) + \psi_h \varepsilon_t + \psi_{h+1}\varepsilon_{t-1} + ..$y error de predicción es

- $ep^{y}_{t+h|t} = y_{t+h} - \hat{y}_{t+h|t} =$

  $\varepsilon_{t+h} + \psi_1 \varepsilon_{t+h-1} + \psi_2 \varepsilon_{t+h-2} + .. + \psi_{h-1}\varepsilon_{t+1}$

- Por lo que podemos calcular el error cuadrático medio (de las predicciones)

Comparación TS-DS: ii) errores de predicción

- $E\left[\, y_{t+h} - \hat{y}_{t+h} \right]^2 = \left[1 + \psi_1^2 + \psi_2^2 + ... + \psi_{h-1}^2\right] \sigma^2 < \infty \qquad \text{(TS)}$
- Por lo que el ECM aumenta con h, pero converge a un valor finito por la propiedad del absolute summability

Comparación TS-DS: ii) errores de predicción

- Para una variable DS el error de predicción es (probarlo):
$y_{t+h} - \hat{y}_{t+h|t} = \varepsilon_{t+h} + (1+\psi_1)\varepsilon_{t+h-1} + (1+\psi_1+\psi_2)\varepsilon_{t+h-2} + ..$
$+(1+\psi_1+\psi_2+..+\psi_{h-1})\varepsilon_{t+1}$

- Por lo que podemos calcular el error cuadrático medio (de las predicciones, DS):

$$E\left[y_{t+h} - \hat{y}_{t+h}\right]^2 =$$
$$\left[1 + (1+\psi_1)^2 + (1+\psi_1+\psi_2)^2 + .. + (1+\psi_1+\psi_2+..+\psi_{h-1})^2 ..\right]\sigma^2$$

- Como antes el ECM crece con h, pero la diferencia es que ahora no converge. Dado que el ECM crece linealmente con h, la desviación estándar de la predicción crece con $\sqrt{h}$. Además, si $\alpha > 0$ la serie crece de manera lineal con h

- Por lo que, un intervalo de confianza de la predicción al 95% crece más lento que la serie

Comparación TS-DS: iii) persistencia de los shocks

- Analizar la persistencia de los shocks es otra manera de comprender la diferencia entre procesos integrado y procesos estacionarios
- Vemos la diferencia en la persistencia de las innovaciones
- Veremos que en una variable estacionaria en tendencia determiística los shocks no tienen efecto permante y si lo tienen para procesos estacionarios en diferencias o integrados

Comparación TS-DS: iii) persistencia de los shocks

- (TS):$y_{t+h} = \alpha + \beta\left(t + h\right) + \varepsilon_{t+h} + \psi_1 \varepsilon_{t+h-1} + \psi_2 \varepsilon_{t+h-2} + .. + \psi_{h-1}\varepsilon_{t+1} + \psi_h \varepsilon_t + \psi_{h+1}\varepsilon_{t+h-(h+1)} + ..$

- Multiplicador dinámico: $\frac{\partial y_{t+h}}{\partial \varepsilon_t} = \psi_h$

- Como el proceso es TS, se cumple absolute summability, por lo tanto:

$lim_{h\rightarrow\infty}\frac{\partial y_{t+h}}{\partial \varepsilon_t} = 0$

- (DS): $\hat{y}_{t+h|t} = y_t + h\alpha + (\psi_h + \psi_{h-1} + ... + \psi_1)\varepsilon_t + (\psi_{h+1} + \psi_h + ... + \psi_2)\varepsilon_{t-1} + (\psi_{h+2} + \psi_{h+1} + ... + \psi_3)\varepsilon_{t-2} + ..$

- Multiplicadores dinámicos: $\frac{\partial y_{t+h}}{\partial \varepsilon_t} = \frac{\partial y_t}{\partial \varepsilon_t} + \left(\psi_h + \psi_{h+1} + ...\psi_1\right) = 1 + \psi_1 + \psi_2 + \psi_3 + ... + \psi_h ..$

- Por lo que el efecto es permanente. Hay un 1 en la ecuación anterior que hace que el efecto sea permanente.

Comparación TS-DS: iv) eliminar tendencias

- Si es TS quitar la tendencia determinística. Si es DS tomar d diferencias tal que $(1-L)^d \, y_t$ sea estacionaria
- ¿Qué pasa si nos equivocamos? Es decir a una serie TS la diferenciamos y a una DS le quitamos la tendencia lineal
- Si es DS$\Rightarrow$ $y_t = \alpha + y_{t-1} + \varepsilon_t$
- Si hacemos detrending:
  $y_t - \overset{\wedge}{\beta} t = \alpha + y_{t-1} - \overset{\wedge}{\beta} t + \varepsilon_t = \alpha + \alpha + (y_{t-2} + \varepsilon_{t-1}) - \overset{\wedge}{\beta} t + \varepsilon_t = ..$
- $= \alpha t + y_0 + (\varepsilon_t + \varepsilon_{t-1} + .. + \varepsilon_1) - \overset{\wedge}{\beta} t$
- No estacionario pues esperanza y varianza varían con t

Comparación TS-DS: iv) eliminar tendencias

- ¿Qué pasa si diferencio un TS?
- $y_t = \alpha + \beta t + \psi(L)\varepsilon_t$
- $\Delta y_t = y_t - y_{t-1} = \alpha + \beta t + \psi(L)\varepsilon_t - [\alpha + \beta\,(t-1) + \psi(L)\varepsilon_{t-1}]$
- $\Delta y_t = \beta t - \beta t + \beta + \psi(L)\,(1-L)\,\varepsilon_t$
- Estacionario pero queda un MA(1) no invertible

Filtro de Hodrick Prescott (HP, 1980)

- Un tercer enfoque para remover tendencias es mediante el uso de filtros
- En general, un filtro es un procedimiento a través del cual se enfrentan los datos y sirve para aislar componentes de una serie
- A continuación veremos el filtro de Hodrick-Prescott

Filtro de Hodrick Prescott (HP, 1980)

- Método para descomponer una serie en tendencia y ciclo (previamente se extrajeron los demás componentes)
- El objetivo es observar una estimación suave del componente de tendencia de largo plazo de una serie
- Minimiza la varianza de Y_t alrededor de la tendencia sujeto a una penalidad que restringe la segunda diferencia de la tendencia

Filtro de Hodrick Prescott

- $Min_{\{l_t\}_{t=1}^{T}} \sum_{t=1}^{T} (y_t - l_t)^2 + \lambda \sum_{t=2}^{T-1} [(l_{t+1} - l_t) - (l_t - l_{t-1})]^2$
- El parámetro de penalidad es arbitrario y controla la suavidad de t_t
- Cuanto más grande λ más suave es la estimación de la tendencia
- H-P recomiendan: $\lambda = \left\{ \begin{array}{cc} 100 & \text{datos anuales} \\ 1600 & \text{datos trimestrales} \\ 14400 & \text{datos mensuales} \end{array} \right\}$
- ¿Qué pasa si $\lambda = 0$ o $\lambda \to \infty$?

Filtro de Hodrick Prescott

- $\lambda = 0 \ \Rightarrow l_t = y_t$. El componenente cíciclo es 0

- $\lambda \to \infty \Rightarrow l_t$ es una tendencia lineal pues el segundo término debe ser 0: $l_t = \beta_1 + \beta_2 t$

- Limitantes:

  - Sin justificación teórica

  - Harvey y Jaeger (1993, JAE) encuentran que HP puede distorsionar la estimación del componente cíclico. Puede conducir a conclusiones erróneas sobre la relación entre los movimientos de corto plazo de las variables macro

  - No permite distinguir entre ciclos de diferentes frecuencias

  - J. Hamilton (2017). Why You Should Never Use the Hodrick-Prescott Filter. NBER WP 23429

J. Hamilton 2017. Porque no usar HP

- 1. The HP filter produces series with spurious dynamic relations that have no basis in the underlying data-generating process
- 2 Filtered values at the end of the sample are very different from those in the middle, and are also characterized by spurious dynamics
- 3. A statistical formalization of the problem typically produces values for the smoothing parameter vastly at odds with common practice, e.g., a value for λ far below 1600 for quarterly data
- 4. There's a better alternative. A regression of the variable at date t+h on the four most recent values as of date t offers a robust approach to detrending that achieves all the objectives sought by users of the HP filter with none of its drawbacks

Estacionalidad

- Oscilaciones sistemáticas a lo largo del año (venta de helados, PIB, inflación, etc)
- Relevante para un adecuado análisis de la coyuntura
- Teníamos que $y_t = l_t + s_t + \varepsilon_t \qquad \Rightarrow \qquad y_t - s_t = l_t + \varepsilon_t$
- Por lo que la serie ajustada de estacionalidad puede ser vista como la suma de los componentes de tendencia e irregular
- Veremos 3 métodos de desestacionalizar:
  - i) determinísticos, variables dummy
  - ii) en base a modelos (TRAMO-SEATS). TRAMO significa "Time series Regression with ARIMA noise, Missing values and Outliers" y SEATS significa Signal Extraction in ARIMA Time Series
  - iii) X13-ARIMA, método empiricista, no se basa en modelos

Estacionalidad: i) variables dummy

- Si por ejemplo los datos son mensuales se podría plantear:

$$y_t = \alpha_0 + \sum_{i=1}^{12} \alpha_i D_{it} + \phi_p(L) + \varepsilon_t + \theta_q(L) \qquad \text{con } \sum_{i=1}^{12} \alpha = 0$$

- D_{it} es una variable dummy que toma el valor 1 si la observación t corresponde al mes i y 0 en otro caso

- α_i es el coeficiente estacional asociado al mes i y $\alpha_0 + \alpha_1$ es el nivel medio correspondiente al mes i

- Una limitante es que aumenta de manera importante la cantidad de parámetros a estimar

- $\phi_p(L) = \phi_1 y_{t-1} + .. + \phi_p y_{t-p}\ ; \theta_q(L) = \theta_1 \varepsilon_{t-1} + .. + \theta_q \varepsilon_{t-q}$

Estacionalidad: i) variables dummy

- Problema: como hay multicolinealidad exacta entre los doces impulsos estacionales y la constante, se opta por eliminar una de las variables, por ejemplo la del mes de diciembre y se estima

- $y_t = \beta_0 + \sum_{i=1}^{11} \beta_i D_{it} + \phi_p(L) + \varepsilon_t + \theta_q(L)$

- ¿Cómo se intepretan estos nuevos parámetros β's?

Estacionalidad: i) variables dummy

- $y_t = \alpha_0 + \sum_{i=1}^{12} \alpha_i D_{it} + \phi_p(L) + \varepsilon_t + \theta_q(L)$
- $y_t = \beta_0 + \sum_{i=1}^{11} \beta_i D_{it} + \phi_p(L) + \varepsilon_t + \theta_q(L)$
- Ahora veamos que pasa cuando i=1$\Rightarrow$el efecto en primera ecuación es $\alpha_0 + \alpha_1$y en la otra es $\beta_0 + \beta_1$, por lo tanto: $\alpha_0 + \alpha_1 = \beta_0 + \beta_1$.Además, $\alpha_0 + \alpha_{12} = \beta_0 \Rightarrow$
- $\alpha_0 + \alpha_1 = \beta_0 + \beta_1 = (\beta_0 - \alpha_{12}) + \alpha_1 = \beta_0 + \beta_1 \Rightarrow \beta_1 = \alpha_1 - \alpha_{12}$
- La constante β_0 es el nivel medio correspondiente al mes que se omitió (diciembre). El coeficiente del impulso estacional del mes i, β_i es la diferencia entre el coeficiente estacional del mes i y el mes omitido (que sirve de comparación)

Estacionalidad: ii) en base a modelos

- Se estiman modelos ARMA para captar el componente estacional
- Para simplificar se factoriza el polinomio AR de la siguiente forma: $\phi_p^*(L) = \phi_{p_1}(L)\,\gamma_{p_2}(L^s)$ donde ϕ es el polinomio del modelo AR
- Se factoriza en dos terminos: unos sobre L y otro sobre L^s. Si los datos son mensuales tendré L y L^{12}
- Ej. AR(3) con componentes estacionales:

  $$(1 - \phi_1 L - \phi_2 L^2 + \phi_3 L^3)(1 - \phi_{12} L^{12} - \phi_{24} L^{24})Y_t = \varepsilon_t$$

- Lo vamos a observar en los correlogramas como un AR(27)

Estacionalidad: ii) en base a modelos

- Esta ecuación es mucho más fácil de estimar. $p = p_1 + p_2\ s = 3 + 2{*}12 = 27$. En el ejemplo p=27, s es la periodicidad de la serie s=12. p_1 es el número de rezagos de la parte AR, p_2 es número de años para atrás de los que depende el componente estacional

- Así logramos simplificar la estructura dinámica y también el número de parámetros a estimar

- Se simplifica pues solo se incorporan los principales rezagos: los inmediatamente anteriores (1,2,3) y los estacionales iniciales (solo 2 en este caso)

Estacionalidad: ii) en base a modelos

- El tema es que tengo restricciones en los coeficientes: los rezagos 4 a 11 y los 16 a 23 están restringidos a tener coeficiente nulo

- Obtuvimos el SARIMA (p,d,q) $(P,D,Q)_s$ en donde $(P,D,Q)_s$ refiere a lo mismo que en un ARIMA(p,d,q) pero referido al componente estacional

- Lo bueno es que se incorporan solo los principales rezagos

  - Los inmediatamente anteriores
  - Los estacionales

- Y así nos aseguramos de cumplir con el principio de parsimonia

- Este es un ejemplo de desestacionalización en base a modelización econométrica. Se usa el software TRAMO-SEATS que están incorporados en las últimas versiones de Eviews y en Demetra

Estacionalidad: ii) en base a modelos

- En la estimación del modelo ARMA se realizar análisis de intervención (AI)

- La idea es incorporar variables dummies para capturar efectos de eventos exógenos o especiales como feríados, huelgas, desastres naturales, outliers o observaciones atípicas, cambios de nivel de la serie (quiebre)

- Cuando estimamos un ARMA puede existir información relevante en el componente irregular como por ejemplo:

a) Eventos que marcaron la historia de la serie: cambios estructurales, sequías, otros desastres naturales, huelgas, ciertos cambios en la nomenclatura, etc. Estos eventos son conocidos en teoría y pueden ser por lo tanto introducidos en el modelo

b) Otros efectos calendario no estacionales, como el número de número de días hábiles del mes, feríados, controlar en que trimestre-mes es semana santa, etc

Disgresión: análisis de intervención

- Enfoque general y no solamente aplicable cuando queremos desestacionalizar una serie
- Veremos un versión simplificada para datos mensuales
- Consideraremos: a) outliers aditivos, b) cambio de nivel, c) cambios transitorios y d) oscilaciones

Disgresión: análisis de intervención

- Relevante para el análisis de series de tiempo
- Los atípicos deben tener justificación
- ¿Cómo detectamos estos atípicos?

Disgresión: análisis de intervención

- Para elegir el modelo se exige que los residuos sean ruido blanco
- Media igual a 0, varianza constante y ausencia de correlación
- "Bajo" error de predicción
- Distribución normal
- La serie observada se puede descomponer en $y_t = l_l + s_t + d_t + \varepsilon_t$
- Donde d representa los eventos determinísticos
- Se filtra la serie para eliminar la distorsión ocasionada por los valores atípicos

Estacionalidad: iii) método empiricista: X13-ARIMA

- Este método se basa en el uso de medias móviles para remover la estacionalidad de una serie
- Utilizado por el US Census Bureau para desestacionalizar las series que releva
- Veremos un versión simplificada para datos mensuales
- En este método existen dos etapas que funcionan de manera independiente

Estacionalidad: iii) método empiricista: X13-ARIMA

- Etapa I: Se estima un modelo ARIMA con fines predictivos con análisis de intervención. Se usa como input la serie ya corregida de factores deterministas
- Se predice fuera de la muestra en ambas extremos, por lo general se predice 12 o 24 meses y se agregan dichos valores predichos a la serie original
- Ahora en la etapa II veremos cual es el objetivo de esta predicción

Estacionalidad: iii) método empiricista: X13-ARIMA

- 1. Una idea muy natural e intuitiva es tomar promedios móviles para eliminar el componente estacional. Por ejemplo:se calcula $x_t = \frac{0.5y_{t-6}+y_{t-5}+....+y_t+\hat{y}_{t+1}+....+0.5\hat{y}_{t+6}}{12}$ Se divide entre 4 si los datos son trimestrales. En este caso se tomaron promedios moviles de un año, pero puede hacerse de 2 años.

  Notar que en esta etapa $y_{t+1}, y_{t+2}, ...y_{t+6}$ son sutituidos por las predicciones obtenidas previamente. Notar también que al tomar períodos de 12 meses estamos eliminado los componentes estacional e irregular. Por lo tanto x_t nos esta indicando la tendencia de la serie

- 2. Calcular la diferencia $\boxed{d_t = y_t - x_t}$

Estacionalidad: iii) método empiricista: X13-ARIMA

- 3. Calcular los índices de estacionalidad.

  Para datos mensuales, i_m =promedio de d_t de los diferentes años para el mes m. Por ejemplo para el primer mes, usar d_1, d_{13}, d_{25}, etc.

  O de manera más compacta: $$i_m = \frac{\sum_{i=1}^{T} d_i I(i=m)}{\sum_{i=1}^{T} I(i=m)}$$

  donde I es la función indicatriz que vale 1 si la condición entre paréntesis se cumple. $I\left(x = A\right)$ es una función que vale 1 si x es A

Estacionalidad: iii) método empiricista: X13-ARIMA

- 4. Ajustamos el índice para que valga 0 en el año

$$s_m = i_m - \bar{i} \qquad \bar{i} = \frac{i_1 + ... + i_{12}}{12}$$

- 5. Calcular la serie ajustada de estacionalidad. $SAy_t = y_t - s_t$

Estacionalidad: iii) limitantes del X13-ARIMA

- Puede no remover totalmente la estacionalidad. En este caso volver a aplicar el filtro y/o aplicar un promedio móvil con más años
- La serie ajustada por estacionalidad en los extremos de la muestra puede ser sensible a los valores predichos utilizados

Pruebas empíricas de raíces unitarias

- Existe un gran variedad de pruebas de raíces unitarias, que difieren en su enfoque, poder y robustez frente a desviós en las condiciones supuestas

- Considerando la regresión $y_t = \phi y_{t-1} + \varepsilon_t$, la hipótesis nula de raíz unitaria será $H_0) \phi = 1$

- La prueba t exige que tomemos el parámetro estimado, restemos 1 y dividir por la desviación estándar

- Dado que las series económicas son no explosivas no tiene sentido probar $\mid \phi \mid > 1$, por lo que es una prueba a una cola

- Probaremos
  - $H_0) \mid \phi \mid = 1$ y el proceso no es estacionario
  - $H_1) \mid \phi \mid < 1$ y el proceso es estacionario

Pruebas de RU

- La distribución válida no es la que calculamos anteriormente,existe una discontinudad en la distribución límite cuando pasamos del caso estacionario al no estacionario

- Por conveniencia práctica vamos a cambiar ligeramente la hipótesis nula sobre si existen RU

- Una manera más directa es restar y_{t-1}de cada lado de la ecuación para obtener: $\triangle y_t = (\phi - 1)y_{t-1} + \varepsilon_t$ con lo que $H_0)\,(\phi - 1) = \gamma = 0$

- La hipótesis alternativa es que el proceso es estacionario

- Desde un punto de vista práctico, tener presente persistencia en el proceso generador de los datos inválida inferencia tradicional

Prueba de Dickey-Fuller aumentada

- p rezagos de Δy_t:

$$
\begin{aligned}
&(a) \qquad \triangle y_t = \gamma y_{t-1} + u_t + \sum_{p=1}^{P} \gamma_p \triangle y_{t-p} \\
&(b) \qquad \triangle y_t = \alpha + \gamma y_{t-1} + u_t + \sum_{p=1}^{P} \gamma_p \triangle y_{t-p} \\
&(c) \qquad \triangle y_t = \alpha + \beta t + \gamma y_{t-1} + u_t + \sum_{p=1}^{P} \gamma_p \triangle y_{t-p}
\end{aligned}
\tag{8}
$$

H_0) $\mid \gamma \mid = 0$ y el proceso no es estacionario
H_1) $\mid \gamma \mid \neq 0$ y el proceso es estacionario

- Nuevamente la distribución del test es diferente a la de la prueba DF por la presencia de los γ_p

- Notar que como en otras ocasiones en econometría incluímos rezagos de la variable endógena para controlar autocorrelación serial de los residuos

Prueba de Dickey-Fuller aumentada: implementación

- Se debe definir el número de rezagos p. Si p=0 tenemos la prueba DF. Definir p según criterio de información de tipo Akaike

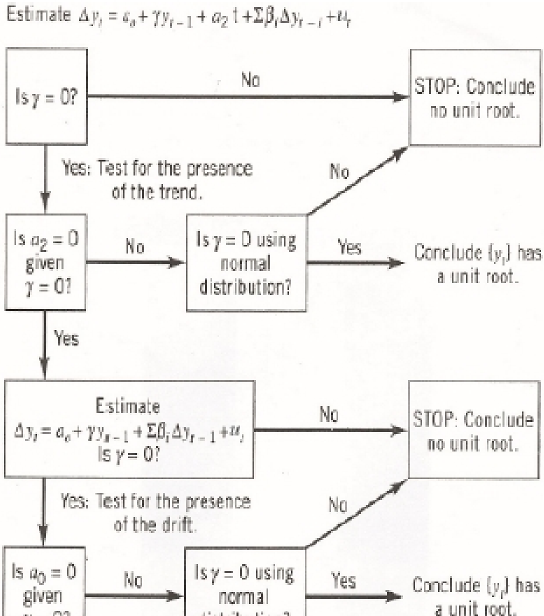
- De nuevo tenemos que probar incluyendo primero con constante y tendencia

- Finalmente hay que notar que los valores críticos reportados solo son válidos si se aplican directamente a la series de datos y no a una estimación de los mismos

- Es posible que la serie tenga más de una raíz unitaria por lo que si procedemos de lo general a lo particular deberíamos empezar con $H_0)$ 2 raíces unitarias, $y_t \sim I(2)$, $H_1)$ 1 raíz unitaria $(y_t \sim I(1))$

$$\triangle^2 y_t = \varphi \Delta y_{t-1} + \varepsilon_t + \sum_{p=1}^{P} \gamma_p \triangle^2 y_{t-p}$$

Prueba ADF


Estimate $\Delta y_t = a_0 + \gamma y_{t-1} + a_2 t + \Sigma\beta_i \Delta y_{t-i} + u_t$
Is $\gamma = 0$?
No
STOP: Conclude no unit root.
Yes: Test for the presence of the trend.
No
Is $a_2 = 0$ given $\gamma = 0$?
No
Is $\gamma = 0$ using normal distribution?
Yes
Conclude $\{y_t\}$ has a unit root.
Yes
Estimate $\Delta y_t = a_0 + \gamma y_{t-1} + \Sigma\beta_i \Delta y_{t-1} + u_t$ Is $\gamma = 0$?
No
STOP: Conclude no unit root.
Yes: Test for the presence of the drift.
No
Is $a_0 = 0$ given
No
Is $\gamma = 0$ using normal
Yes
Conclude $\{y_t\}$ has a unit root.

Prueba de Dickey-Fuller aumentada

- 1. Empezar por el modelo más general con constante y tendencia y probar $\gamma = 0$. Si rechazo H_0 listo. No hay raíz unitaria. Pero si no rechazo puede ser por haber incluido demasiados regresores
- 2. Probar la significancia de la variable de tendencia lineal (β)
- 3. Estimar sin tendencia. Solo con constante
- 4. Estimar sin tendencia ni constante

Prueba de Dickey-Fuller aumentada

- Una limitante de la prueba es cuando existe cambio de régimen. Si por ejemplo la serie es estacionaria pero existe cambio de régimen es posible que la prueba la detecte como I(1)

- Un problema común de la prueba ADF es el bajo poder (probabilidad de rechazar una hipótesis nula falsa) derivado de la *equivalencia observacional*

Prueba de Dickey-Fuller aumentada

- Supongamos que el verdadero modelo es no estacionario: $y_t = y_{t-1} + \varepsilon_t$ pero en una muestra finita de tamaño T estimamos un AR(1) estacionario $y_t = \phi y_{t-1} + \varepsilon_t$

- Consideremos la predicción y el MSE para ambos modelos

- Para el random walk: $E_t\left(y_{t+h}\right) = y_t \Rightarrow$ $E_t\left(y_{t+h} - E_t\left(y_{t+h}\right)\right)^2 = h\sigma^2$

- Para el proceso AR(1) estacionario: $E_t\left(y_{t+h}\right) = \phi^h y_t \Rightarrow$ $E_t\left(y_{t+h} - E_t\left(y_{t+h}\right)\right)^2 = \left(1 + \phi^2 + \phi^4 + ...\right)\sigma^2$

- Para un valor ϕ próximo a 1 ambos procesos se verán casi iguales en una muestra de tamaño finito (equivalencia observacional)

Prueba Kwiatkowoski, Phillips, Schmidt y Shin (KPSS, 1992)

- Bajo poder de la prueba ADF
- Puede que no rechacemos la nula de RU pues si la hay o porque la información no es suficiente para rechazar
- Una solución puede ser usar una prueba con $H_0)$ estacionaria
- KPSS difiere de ADF en que bajo $H_0)$ la serie es estacionaria

Prueba KPSS

- Partimos de un RW:
  $y_t = \alpha + \beta t + \gamma_t + \varepsilon_t \qquad \gamma_t = \gamma_{t-1} + v_t \qquad v_t \sim iid\left(0, \sigma_v^2\right)$

- Si $\text{VAR}(v_t) = 0$ $(\gamma$ constante$) \Rightarrow \gamma_t = \gamma_0$ y la serie y_t es estacionaria

- Usamos una regresión lineal para estimar el componente estocástico:
  $y_t = \overset{\wedge}{\alpha} + \overset{\wedge}{\varepsilon}_t$

- Notar que bajo $H_0)$ (de estacionariedad) $\overset{\wedge}{\varepsilon}_t$ es estacionario

- $H_0)\sigma_v^2 = 0 \qquad H_1)\sigma_v^2 > 0$

Prueba KPSS

- $KPSS = \frac{1}{T^2 \overset{\wedge}{\sigma}_u^2} \sum\limits_{t=1}^{T} S_t^2$
- $S_t = \sum\limits_{i=1}^{t} \overset{\wedge}{\varepsilon}_i$(suma parcial de los residuos de y_t contra constante y/o tendencia determinística)
- $\overset{\wedge}{\sigma}_u^2$estimador asintótico de la varianza de $\overset{\wedge}{\varepsilon}_t$
- Es una prueba de parámetros constantes contra parámetros de un RW
- Valores críticos simulados por KPSS

Comparación ADF-KPSS

- Podemos comparar los resultados de las pruebas ADF y KPSS
- ADF: $H_0)\ y_t \sim I(1)$. KPSS: $H_0)\ y_t \sim I(0)$. Cuatro posibles resultados
- 1. ADF rechaza $H_0)$. KPSS no rechaza $H_0)$
- 2. ADF no rechaza $H_0)$. KPSS rechaza $H_0)$
- 3. ADF rechaza $H_0)$. KPSS rechaza $H_0)$
- 4.ADF no rechaza $H_0)$. KPSS no rechaza $H_0)$

Comparación ADF-KPSS

- En 1 y 2 los resultados son robustos
- En 3 y 4 los resultados son diferentes
- Esto se llama confirmatory data analysis

Raíces unitarias estacionales

- *Definición*: El proceso no estacionario y_t,observado en S períodos de igual tiempo por año es integrado estacional de orden d, denotado por $y_t \sim SI\,(d)$ si $\Delta_S^d = \left(1 - L^S\right)^d y_t$ es estacionario

- Prueba de Dickey-Hasza-Fuller (DHF). Generalización de la prueba ADF para un AR(1) no estacional que ya vimos

- Supongamos que el proceso es un SAR(1)
  $\Rightarrow \Delta_{12} y_t = \gamma y_{t-12} + \varepsilon_t \qquad H_0)\gamma = 0$ (raíz unitaria estacional)
  $\Rightarrow \Delta_{12} y_t = \varepsilon_t \Rightarrow y_t = y_{t-12} + \varepsilon_t$ RW estacional

  $H_1)\gamma < 0$ proceso estocástico estacional estacionario

- La distribución del estadístico t no es estándar, pero sin embargo es similar a las simuladas por DF. DHF también simulan valores críticos

Prueba de Hylleberg-Engle-Granger-Yoo (HEGY)

- Parte de notar que el operador de diferencias estacionales se puede descomponer como:

$\Delta_S = 1 - L_S = (1 - L)\left(1 + L + L^2 + ... + L^{S-1}\right)$ por lo que tenemos una raíz unitaria regular o convencional y S-1 raíces estacionales

- La prueba HEGY prueba justamente las anteriores S raíces

- Por ejemplo, para datos trimestrales S=4:
  $\Delta_4 = \left(1 - L^4\right) = (1 - L)\left(1 + L + L^2 + L^3\right)$

$\Delta_4 = (1 - L)(1 + L)(1 - iL)(1 + iL) = (1 - L)(1 + L)(1 - i^2 L^2) = (1 - L)(1 + L)(1 + L^2)$ por lo que hay 4 raíces en el circulo unitario: $\pm 1, \pm i$

- Por lo que aparte de la raíz unitaria regular hay tres raíces unitarias estacionales que son: -1,+i,-i

Prueba de Hylleberg-Engle-Granger-Yoo (HEGY)

- Aproximando por Taylor Δ_4 HEGY llegan a:

$$\Delta_4 y_t = \pi_1 y_{1t-1} + \pi_2 y_{2t-1} + \pi_3 y_{3t-2} + \pi_4 y_{3t-1} + \varepsilon_t \qquad \text{con}$$

$$y_{1t} = (1+L)(1+L^2) y_t = y_t + y_{t-1} + y_{t-2} + y_{t-3}$$
$$y_{2t} = -(1-L)\left(1+L^2\right) y_t = -y_t + y_{t-1} - y_{t-2} + y_{t-3}$$
$$y_{3t} = -(1-L)(1+L)\, y_t = -y_t + y_{t-2}$$

- $H_0)$ $\quad \pi_1 = \pi_2 = \pi_3 = \pi_4 = 0 \quad \Rightarrow \Delta_4 y_t = \varepsilon_t \Rightarrow y_t = y_{t-4} + \varepsilon_t$ RW estacional

  $H_1)$ Al menos un $\pi_i \neq 0$ para i=1,2,3,4

- Notar que se trata de una prueba F. Se usan los valores críticos de HEGY

No estacionariedad en varianza

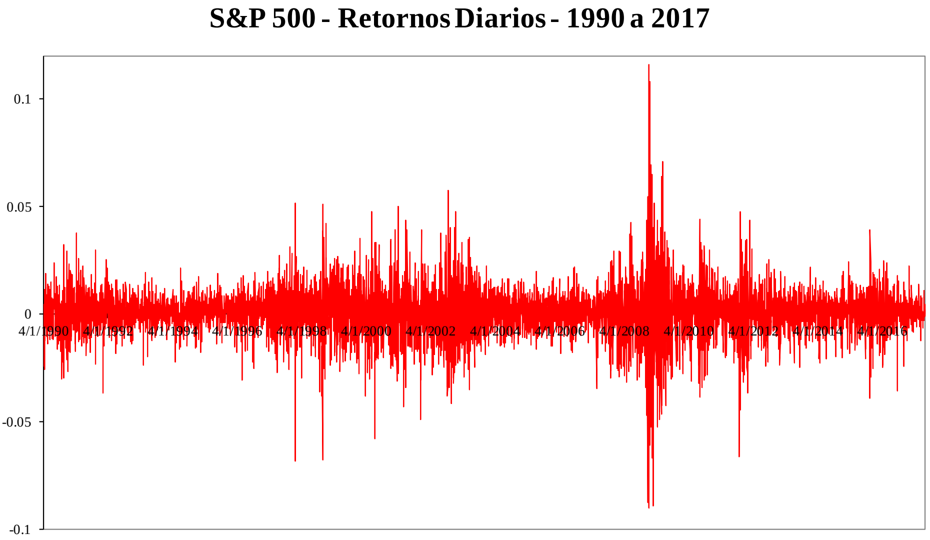
- La estructura lineal de los datos no permite captar una serie de características importantes de los datos financieros, principalmente su:

  - no linealidad
  - no normalidad

- El modelo "tradicional" era $y_t = X_t\beta + \varepsilon_t$ o $y_t = ARMA(p, q)$

- También asumiamos $\varepsilon_t \sim N\left(0, \sigma^2\right)$

Ejemplo de una serie de tiempo de activos financieros



Introducción

- La no linealidad es relevante es finanzas
- La varianza de un activo es usada para cuantificar riesgo. Esta medida de la volatilidad no condicional no toma en cuenta que puede existir un patrón predecible de regularidad en la volatilidad
- Analizaremos modelos de volatilidad condicional (a información en el período t-1)
- En estos modelos se puede predecir el riesgo. Pueden explicar que los precios de activos tengan largos períodos de alta o de baja volatilidad
- Si los agentes pueden predecir volatilidad entonces cuando esta es alta pueden requerir un premio para compensar dicho alto riesgo o directamente retirarse del mercado
- Son modelos donde la variable no es estacionaria pues la varianza no es constante sino que varía con t

Introducción

- Los modelos de varianza constante no solo nos permiten cacular predicciones de la serie sino también de la varianza condicional
- Es decir, nos permiten predecir la varianza
- De igual manera que los ARMA tenían una parte sistemática y un error, en estos modelos podemos descomponer la varianza en un componente predecible y otro que no lo es

Introducción

- El supuesto que la varianza del error es constante es homocedasticidad, esto es $V\left(\varepsilon_t\right) = \sigma^2$

- ¿Cuáles son las consecuencias del no cumplimiento del supuesto de homocedasticidad?

  - Los estimadores de MCO siguen siendo insesgados
  - Los errores estándares estimados son incorectos por lo que las pruebas t,F no son aplicables

- Definitivamente esperamos que para datos financieros no se cumpla el supuesto de homocedaticidad

Introducción: media y varianza condicional y no condicional

- Asumiendo y_t RW: $y_t = y_{t-1} + \varepsilon_t \Rightarrow y_t = y_0 + \sum_{i=1}^{t} \varepsilon_i$

- Media y varianza no condicional:
  - $E(y_t) = y_0$
  - $V(y_t) = t\sigma^2$

- Media y varianza condicional:
  - $E(y_t \mid y_{t-1}) = y_{t-1}$
  - $V(y_t \mid y_{t-1}) = E\left[y_t - E(y_t \mid y_{t-1})\right]^2 = E(y_{t-1} + \varepsilon_t - y_{t-1})^2 = E\left(\varepsilon_t^2\right) = \sigma^2$

- Por lo que un RW presenta media no condicional constante pero varianza no condicional no constante

- La varianza no condicional tiende a ∞ al crecer t pero la varianza condicional es constante

Modelos ARCH: AR con heterocedasticidad condicional

- Engle (1982) :"Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation", Econometrica

- Se planeta un modelo que no asuma que la varianza es constante

- En particular se asume que la varianza condicional es una función lineal de las r innovaciones pasadas al cuadrado

- ARCH(r): $\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + .. + \alpha_r \varepsilon_{t-r}^2 = \alpha_0 + \sum_{i=1}^{r} \alpha_i \varepsilon_{t-i}^2$

- Notar que aparte tenemos la ecuación para la media que viene dada por el ARMA(p,q) para y_t

- Es decir, tenemos dos ecuaciones, una para la media y otra para la varianza

Modelos ARCH

- La varianza debe ser positiva. Una condición suficiente es $\alpha_0 > 0, \alpha_1 > 0, .., \alpha_r > 0$

- Sea $u_t \equiv \varepsilon_t^2 - \sigma_t^2 \Rightarrow$ARCH(r): $\varepsilon_t^2 = \alpha_0 + \sum_{i=1}^{r} \alpha_i \varepsilon_{t-i}^2 + u_t$

- Es un modelo AR(r) para las innovaciones al cuadrado $\left(\varepsilon_t^2\right)$

- $\sigma_t^2 = E\left(\varepsilon_t^2 \mid \varepsilon_{t-1}^2, \varepsilon_{t-2}^2, .., \varepsilon_{t-r}^2\right)$

- Si el modelo es estacionario, la varianza no condicional viene dada por

- $V\left(\varepsilon_t\right) = \frac{\alpha_0}{1-\alpha_1-\alpha_2-....-\alpha_r}$

ARCH(1)

- Modelo ARCH más simple
- $y_t = \alpha + \varepsilon_t$ (o ARMA(p,q), o $X_t\beta + \varepsilon_t$)
- $\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 \qquad (\alpha_0 > 0, \alpha_1 > 0)$
- Alternativamente podemos escribir:
- $\varepsilon_t = v_t \sigma_t \qquad \text{con} \qquad \sigma_t = \sqrt{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2} \qquad v_t \sim N(0,1)$
- Son dos maneras diferentes de expresar el mismo modelo. La primera forma es más clara. La segunda es útil para rezalir simulaciones
- Esto es: $\varepsilon_t = v_t \sqrt{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2}$

ARCH(1)

- $y_t = \alpha + \varepsilon_t$
- $\varepsilon_t = v_t \sqrt{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2} \qquad \varepsilon_t = v_t \sigma_t \qquad v_t \sim N(0,1)$
- $E(\varepsilon_t \mid \varepsilon_{t-1}) = E\left(v_t \sqrt{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2} \mid \varepsilon_{t-1}\right) =$
  $E(v_t \mid \varepsilon_{t-1}) \sqrt{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2} = 0$
- $V(\varepsilon_t \mid \varepsilon_{t-1}) = V\left(v_t \sqrt{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2} \mid \varepsilon_{t-1}\right) =$
  $\left(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2\right) V(v_t \mid \varepsilon_{t-1}) = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$
- Notar que al tomar esperanza de una variable en t-1 en el período t-1 es una constante conocida

ARCH(1)

- $E\left(y_t \mid y_{t-1}\right) = \alpha \qquad V\left(y_t \mid y_{t-1}\right) = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$

- Si el proceso es estacionario:

  $$V\left(y_t\right) = V\left(\varepsilon_t\right) = E\left(\varepsilon_t^2\right) = E\left[v_t^2\left(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2\right)\right] = \frac{\alpha_0}{1-\alpha_1}$$

- $V\left(y_t \mid y_{t-1}\right) = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$

- La varianza condicional de y_t varía con t pero no la no condicional

AR(1) con ARCH(1)

- $y_t = \phi y_{t-1} + \varepsilon_t$
- $\varepsilon_t = v_t \sigma_t$
- $\sigma_t = \sqrt{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2} \qquad (\alpha_0 > 0, \alpha_1 > 0) \qquad v_t \sim N(0,1)$
- $E\left(\varepsilon_t \mid \varepsilon_{t-1}\right) = E\left(v_t \sqrt{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2} \mid \varepsilon_{t-1}\right) = 0$ (como antes)
- $V\left(\varepsilon_t \mid \varepsilon_{t-1}\right) = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 \qquad V\left(\varepsilon_t\right) = \frac{\alpha_0}{1-\alpha_1}$ (como antes)
- $E\left(y_t \mid y_{t-1}\right) = \phi y_{t-1} \qquad V\left(y_t \mid y_{t-1}\right) = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$

AR(1) con ARCH(1)

- Para hallar la varianza no condicional de y_t estacionario usamos la siguiente propiedad de la varianza:
  $V\left(y_t\right) = E\left[V\left(y_t \mid y_{t-1}\right)\right] + V\left[E\left(y_t \mid y_{t-1}\right)\right]$

- $E\left[V\left(y_t \mid y_{t-1}\right)\right] = E\left(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2\right) = \alpha_0 + \alpha_1 E\left(\varepsilon_{t-1}^2\right) = \alpha_0 + \alpha_1 V\left(\varepsilon_{t-1}^2\right) = \alpha_0 + \frac{\alpha_1 \alpha_0}{1-\alpha_1}$

- $E\left[V\left(y_t \mid y_{t-1}\right)\right] = V\left(\phi y_{t-1}\right) = \phi^2 V\left(y_{t-1}\right) = \phi^2 V\left(y_t\right)$

- $V\left(y_t\right) = \alpha_0 + \frac{\alpha_1 \alpha_0}{1-\alpha_1} + \phi^2 V\left(y_t\right) \Rightarrow \left(1-\phi^2\right) V\left(y_t\right) = \frac{\alpha_0(1-\alpha_1)+\alpha_1\alpha_0}{1-\alpha_1} = \frac{\alpha_0}{(1-\alpha_1)\left(1-\phi^2\right)}$

- La varianza no condicional es constante

Probando efectos ARCH

- ¿Cómo determinamos si debemos incluir efectos ARCH al modelo?

- 1. Estimar la ecuación para la media: ARMA(p,q) o un modelo lineal y guardar los residuos muestrales $\overset{\wedge}{\varepsilon}_t$.La hipótesis alternativa es que los errores siguen un ARCH(r)

- 2. Elevar los residuos al cuadrado y regresarlo contra r de sus rezagos para comprobar la hipótesis de un ARCH(r).

  $\overset{\wedge}{\varepsilon}{}_t^2 = \gamma_0 + \gamma_1 \overset{\wedge}{\varepsilon}{}_{t-1}^2 + \gamma_2 \overset{\wedge}{\varepsilon}{}_{t-2}^2 + .. + \gamma_r \overset{\wedge}{\varepsilon}{}_{t-r}^2 + e_t$ e_t ruido blanco.
  Guardar R^2

- 3. $H_0)\gamma_1 = \gamma_2 = 0, .. = \gamma_r = 0$ (No hay efectos ARCH)
  $H_1)\gamma_1 \neq 0$ o $\gamma_2 \neq 0, ..$ o $\gamma_r \neq 0$

Probando efectos ARCH

- 4. La prueba estadística es $TR^2 \sim \chi^2_r$
- 5. Rechazamos la hipótesis nula si el valor de la prueba estadística es mayor que el valor de tablas
- También es posible hacer una prueba F de significación conjunta del modelo
- Tener cuidado con la prueba pues si el modelo esta mal especificado es probable rechazar la nula simplemente porque la correlación serial en el tiempo de los residuos induce correlación en los residuos al cuadrado

ARCH y cambio de régimen

- Diebold (1986) mostro que cuando existe un cambio de régimen en la varianza y esta cambia de un nivel a otro en los datos esto se ve como un modelo con efectos ARCH para todo el período
- En estos casos, dividir la muestra y probar efectos ARCH en cada subperíodo
- Sino hay efectos ARCH para cada subperíodo pero si los hay para toda la muestra es indicio de cambio de régimen en la varianza no condicional. En estos caso es incorrecto modelar un ARCH

Problemas con los modelos ARCH(r)

- ¿Cómo determinamos el valor de r? Mismas técnicas que vimos para determinar p,q en el ARMA(p,q). Prueba LR (Likelihood ratio)

- El valor requerido de r puede ser grande (pérdida de grados de libertad y parsimonia)

- Pueden no cumplirse las restricciones de no negatividad. Cuando estimamos un modelo ARCH requerimos $\alpha_i > 0$ para i=0,1,2,..,r (pues las varianzas no pueden ser negativas)

- Una extensión natural de los modelos ARCH(r) muy utilizada en la práctica y que considera alguno de estos inconvenientes en el GARCH (generalized ARCH)

Modelos GARCH

- Bolerslev (1986). La varianza condicional depende de sus rezagos
- $\sigma_t^2 = \alpha_0 + \sum_{i=1}^{r} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{s} \beta_i \sigma_{t-i}^2$ GARCH(r,s)
- GARCH(1,1): $\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$
- Es "como" un ARMA(1,1) para la ecuación de la varianza
- Notar que también podemos definir

  - $\sigma_{t-1}^2 = \alpha_0 + \alpha_1 \varepsilon_{t-2}^2 + \beta \sigma_{t-2}^2$
  - $\sigma_{t-2}^2 = \alpha_0 + \alpha_1 \varepsilon_{t-3}^2 + \beta \sigma_{t-3}^2$ y así sucesivamente
- Sustituyendo en la ecuación del GARCH(1,1)
- $\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta \left(\alpha_0 + \alpha_1 \varepsilon_{t-2}^2 + \beta \sigma_{t-2}^2 \right) =$
  $\alpha_0 (1 + \beta) + \alpha_1 \varepsilon_{t-1}^2 + \alpha_1 \beta \varepsilon_{t-2}^2 + \beta^2 \sigma_{t-2}^2$

Modelos GARCH

- $\sigma_t^2 = \alpha_0 (1+\beta) + \alpha_1 \varepsilon_{t-1}^2 + \alpha_1 \beta \varepsilon_{t-2}^2 + \beta^2 \left(\alpha_0 + \alpha_1 \varepsilon_{t-3}^2 + \beta \sigma_{t-3}^2\right)$
- $= \alpha_0 \left(1+\beta+\beta^2\right) + \alpha_1 \varepsilon_{t-1}^2 + \alpha_1 \beta \varepsilon_{t-2}^2 + \alpha_1 \beta^2 \varepsilon_{t-3}^2 + \beta^3 \sigma_{t-3}^2 ....$
- $\sigma_t^2 = \alpha_0 \left(1+\beta+\beta^2+..\right) + \alpha_1 \left(1+\beta L+\beta^2 L^2+..\right) \varepsilon_{t-1}^2 + \beta^\infty \sigma_0^2$
- Esto es un ARCH(∞) . Por lo que un ARCH(∞) puede representarse de manera más parsimoniosa como un GARCH(1,1)

Modelos GARCH

- En general un modelo GARCH(1,1) o un GARCH(r,s) con r,s pequeños es suficiente para capturar la volatilidad de los datos

- Ventajas del GARCH sobre el ARCH

  \- Más parsimonioso, evita un sobre ajuste del modelo

  \- Menos probable de incumplir las restricciones de no negatividad

Varianza no condicional del GARCH

- GARCH(1,1): $\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$
- Recordar: $\varepsilon_t = v_t \sigma_t = v_t \sqrt{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2} \qquad v_t \sim N(0,1)$
- $V(\varepsilon_t) = E(\varepsilon_t^2) = E(\alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2) =$
  $\alpha_0 + \alpha_1 E(\varepsilon_{t-1}^2) + \beta E(\sigma_{t-1}^2)$
- $= \alpha_0 + \alpha_1 E(\varepsilon_{t-1}^2) + \beta E(\varepsilon_{t-1}^2) \Rightarrow V(\varepsilon_t) = \frac{\alpha_0}{1-(\alpha_1+\beta)}$
- $(\alpha_1 + \beta) < 1$ GARCH estacionario en varianza
- $(\alpha_1 + \beta) > 1$ GARCH no estacionario en varianza
- $(\alpha_1 + \beta) = 1$ GARCH integrado (no estacionario)
- En los GARCH no estacionarios en varianza, la predicción de la varianza condicional no converge al valor no condicional al aumentar el horizonte de predicción

Estimación del ARCH-GARCH

- Tenemos dos ecuaciones, una para la media y otra para la varianza. Por lo tanto el modelo no es lineal y no podemos aplicar MCO
- La estimación es por MV asumiendo $\varepsilon_t \sim N\left(0, \sigma_t^2\right)$
- Notar que ahora la varianza depende del tiempo por lo que no es una constante
- MV selecciona los valores más probables de los parámetros dado los datos verdaderos
- Más especificamente, calculamos el Ln de la función de verosimilitud y lo maximizamos

Prediciendo varianzas condicionales con el ARCH

- Los ARCH-GARCH sirven para modelar volatilidad pues la varianza condicional es autoregresiva. Dichos modelos se pueden usar para predecir la volatilidad
- Las predicciones de la varianza son aditivas en el tiempo
- Para predecir las varianzas en modelos GARCH se usa un enfoque similar al usado para realizar predicciones en base a modelos ARMA
- De nuevo, se trata de un ejercicio de iterar el operador de expectativas condicionales

Comparando diferentes modelos GARCH

- ¿Cómo comparamos diferentes modelos GARCH?

- De igual manera que en los ARMA podemos compararlos en base a los criterios de información (Akaike etc) y a su capacidad para predecir la varianza condicional

- Criterios: error cuadrático medio (MSE), error cuadrático absoluto (MAE), etc

- A pesar de que el GARCH captura mejor los datos que el ARCH igual es posible que que las predicciones de la varianza sean negativa. Por lo que se desarrollaron un conjunto de extensiones que buscan explicar mejor los hechos estilizados. A continuación presentamos las extensiones más utilizadas en el análisis empírico

Extensiones: i) Modelos Threshold ARCH-GARCH (TARCH)

- ARCH-GARCH asimétricos

- En datos financieros, las malas noticias suelen tener un efecto más importante en la volatilidad que las buenas noticias

- Gloster, Jagannathan y Runkle (1993)

- $\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-1}^2 I_{t-1} + \beta \sigma_{t-1}^2$

- $I_{t-1} = 1$ si $\varepsilon_{t-1} < 0$, $I_{t-1} = 0$ si $\varepsilon_{t-1} \geqslant 0$

- Si $\alpha_2 > 0 \Rightarrow$los shocks negativos tienen un impacto mayor en la volatilidad que los positivos

- Por heterocedasticidad la prueba de hipótesis $H_o) \alpha_2 = 0$ se basa en los residuos estandarizados: $\hat{s}_t = \frac{\overset{\wedge}{\varepsilon}_t}{\overset{\wedge}{\sigma}_t}$

Extensiones: ii) GARCH exponencial (EGARCH)

- Nelson (1991)

- Efectos asimétricos pero se asegura que las varianzas sean siempre positivas
- $Ln\left(\sigma_t^2\right) = \alpha_0 + \alpha_1 \left(\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right) + \alpha_2 \mid \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \mid + \beta Ln\left(\sigma_{t-1}^2\right)$

- σ_t^2 no negativo

- Asimetría. Ante un shock positivo $\left(\frac{\varepsilon_{t-1}}{\sigma_{t-1}} > 0\right)$ el efecto en el log de la varianza condicional es $\alpha_1 + \alpha_2$. Si el shock es negativo el efecto es $\alpha_2 - \alpha_1$

- En la práctica suele ser difícil de obtener la predicción de la varianza

Extensiones: iii) GARCH en medias y ARCH en medias (ARCH-M)

- Engle, Lilien y Robins (1987). La media depende de la volatilidad del proceso. Varías teorías en finanzas predicen una relación entre la media de los retornos y su varianza

- Por ej. se requiere una prima de riego por mantener un activo más volatil (riesgoso). La varianza condicional que afecta la media se interpreta como una prima por riesgo variable en el tiempo. Capta riesgo sistemático

- $y_t = \delta x_t + \gamma \sigma_t^2 + \varepsilon_t$

- $\sigma_t^2 = \alpha_0 + \sum_{i=1}^{r} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{s} \beta_i \sigma_{t-i}^2$

- La solución del GARCH-M es numericamente inestable y en la práctica se suele utilizar el ARCH-M

Extensiones: iv) Modelo GARCH integrado (IGARCH)

- En estos modelos el componente AR de los residuos al cuadrado tiene una raíz unitaria $\alpha_1 + \beta = 1$
- En otras palabras, la volatilidad condicional es persistente
- $\sigma^2_{t+h|t} = \sigma^2_t + h\alpha_0$
- Si ε_t sigue un IGARCH la varianza condicional no existe y el proceso no es estacionario débil
- Nelson (1990) mostro que a pesar que el proceso no es estacionario en covarianza si es estrictamente estacionario
- En la práctica esta raíz unitaria puede estar indicando cambio de régimen y no un IGARCH

Extensiones: v) GARCH con variables explicativas exógenas

- Es posible incorporar variables explicativas exógenas a la ecuación de la varianza del GARCH

- Glosjer, Jagannathan y Runkel (1993) agregan una tasa de interés nominal de corto plazo en varias especificaciones para permitir que tenga efectos en la volatilidad

- $\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta \sigma_t^2 + \gamma x_{t-1}$ Si x_t endógena

- $\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta \sigma_t^2 + \gamma x_t$ Si x_t exógena

- $x_t > 0$