Modelos para la Media Condicional Univariados.

December 5, 2022

Contenido

- Series Financieras.
- Modelos Univariados (ARMA).

Series de tiempo Económicas y Financieras

- Datos Económicos:
  - baja frecuencia (PIB)
  - ciclos, tendencias.
- Datos Financieros:
  - mediana y alta frecuencia.
  - Media condicional:
  - Varianza medición y pronostico:

Media Condicional

- observables.
- precios a retornos.
- retornos diarios $E[r_t] = 0$.
- dependencia intertemporal.
- estacionarios, memoria larga o corta.
- heteroscedasticidad.
- Colas pesadas (no-normalidad).
- no-predecible (AOA).
- microestructura de los mercados (ruido).

Formalización de los datos

- Y variable aleatoria (rv)
- y realización de rv.
- Y generado por modelo estadístico $(\Omega, \mathcal{Y}, \mathcal{P})$.
- Y tiene proceso generador de datos (DGP).
- $Y_{\cdot}$ proceso estocástico (rv indexada).
- Y_t serie de tiempo (rv indexada por el tiempo).
- Dependencia a través del tiempo.
- No necesariamente predecible.

El muestreo de los datos

- Frecuencia: trimestre, mes, día, intradía.
- rv continuas y discretas (distribuciones).
- Los datos son discretos!

En estadística mas datos $\{y_t\}_{t=0}^{T}$ es mejor.

- Baja frecuencia (clásico).
    - Δ fijo $n \to \infty$.
    - índice $t \in (0, T := \frac{n}{\Delta})$.
    - Estacionariedad y ergodicidad es importante.
- Alta frecuencia en 1 día.
    - $\Delta = \Delta_n \to 0$ cuando $n \to \infty$.
    - índice $t \in (0, T := \frac{1}{\Delta_n})$.
    - ergodicidad es importante.
- Mixto $\Delta_n \to 0$ y varios días.

Proceso generador de datos: difusión

Proceso de precios $S_t = S_0 exp(X_t)$
en tiempo continuo donde X_t

$$X_t = X_0 + \int_0^t b ds + \int_0^t \sigma dW_s$$

donde W movimiento Browniano.
S_t es proceso Browniano geométrico
con retorno instantaneo.

$$\frac{\partial S_t}{S_t} = bdt + \sigma dW_t$$

Discretizado,

$$r_{t+1} := \ln S_{t+1} - \ln S_t = (b - \frac{\sigma^2}{2})(t + 1 - t) + \sigma \sqrt{t + 1 - t} Z_{t+1}$$

con $Z_{t+1} \sim N(0, 1)$

Proceso de Levy, mas general

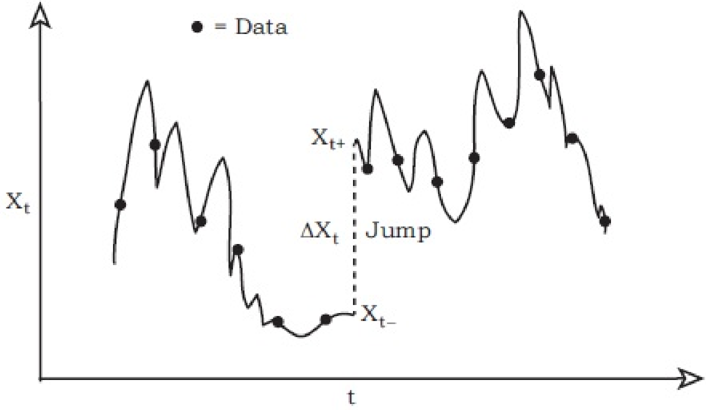
$$X_t = bt + \sigma W_t + (x\mathbf{1}_{||x|| \leq \varepsilon}) * (\mu - v)_t + (x\mathbf{1}_{||x|| > \varepsilon}) * \mu_t$$

donde W movimiento Browniano y μ medida de Poisson.
Descomposición canónica:

- drift (tendencia)
- continuous martingale (difusión)
- purely discontinuous martingale (saltos "pequeños" $< \varepsilon$).
- purely discontinuous martingale (saltos "grandes" $> \varepsilon$).

  $$X_t = bt + \sigma W_t + \sum_{s \leq t} \Delta X_s \mathbf{1}_{||x|| > \varepsilon}$$

variación continua vs saltos



Análisis de series de tiempo

- Objetivo pronostico: $E[y_{t+h} \mid \mathcal{F}_t]$
- Cual es la forma "funcional" de la media condicional?
- La función sera lineal utilizando proyección ortogonal.
- Descomponer los datos

Descomposición clásica

- determinístico v_t (finanzas no muy importante)
  - tendencia.
  - ciclo.
- estocástico
  - regular (estacionario): Descomposición de Wold.
  - irregular (ruido).

Estacionariedad

Estacionariedad estricta:

y_t es estacionario $\Leftrightarrow \forall t_1 < t_2 < \ldots < t_k$

su distribución conjunta:

$$L(y_{t_1}, y_{t_2}, \ldots, y_{t_k}) \propto L(y_{t_1+h}, y_{t_2+h}, \ldots, y_{t_k+h})$$

desplazamiento en el tiempo no afecta su DGP.

Estacionariedad débil:

y_t es débilmente estacionario (momentos no-condicionales):

- $E[y_t]$ independiente t.
- $Var[y_t]$ finito, positivo, independiente t.
- $Cov[y_t, y_s]$ función finita de $t - s$.

La relación entre dos variables con índices de tiempo diferente no depende del índice, solo de la distancia (temporal entre ellas).

Proporciona principio de intercambiabilidad (inferencia).

Estacionariedad en proceso autoregresivo

Considere AR(1)

$$y_t = \mu + \rho y_{t-1} + \epsilon_t$$

donde ϵ_t:

- $E[\epsilon_t] = 0$.
- $Var[\epsilon_t] = \sigma^2 \forall\ t$

Comportamiento del proceso AR(1)

- $\mid \rho \mid = 1$ no estacionario.
- $\rho \in (-1, 1)$ estacionario.
- $\mid \rho \mid > 1$ explosivo.

Estacionariedad en proceso autoregresivo

recursión AR(1):

$$\begin{aligned} y_t &= \mu + \rho y_{t-1} + \epsilon_t \\ y_{t-1} &= \mu + \rho y_{t-2} + \epsilon_{t-1} \\ &. \\ &. \\ y_{t-k} &= \mu + \rho y_{t-k-1} + \epsilon_{t-k} \end{aligned}$$

el resultado es descomposición de Wold ($y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} + v_t$)

$$\begin{aligned} y_t &= \mu(1 + \rho + \rho^2 + ... + \rho^k) + \rho^k y_{-0} + \epsilon_t + \rho\epsilon_{t-1} + \rho^2\epsilon_{t-2} + ... + \rho^k\epsilon_{t-k} \\ y_t &= \mu \sum_{i=0}^{k} \rho^i + \rho^k y_{-0} + \sum_{i=0}^{k} \rho^i \epsilon_{t-i} \end{aligned}$$

Valor de ρ determina comportamiento.

Ergodicidad

Ergodicidad es versión de LLN en análisis de series de tiempo
Ergodicity:
Sea $\{Y_t\}, t = 1, \ldots, T$ proceso estacionario ergódico,
entonces $E \mid Y_t \mid < \infty \Rightarrow \bar{Y}_n \rightarrow_{a.s.} \mu \equiv$E[Y$_t$].

- $\{Y_t\}$ realización tamaño T del proceso.
- Autocovarianza acotada (va a cero en la medida en que j incremente) $\sum_{j=0}^{\infty} \mid \gamma_j \mid < \infty$ entonces $\{Y_t\}$ es ergódico en la media.

Proceso autoregresivo de orden superior AR(p)

AR(P)

$$y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + ... + \rho_p y_{t-p} + \epsilon_t$$

Estacionariedad de AR(p)

$$\begin{aligned} y_t &= \gamma_1 y_{t-1} + \gamma_2 y_{t-2} + ... + \gamma_p y_{t-p} \\ y_t &= (\gamma_1 L + \gamma_2 L^2 + ... + \gamma_p L^p) y_t \end{aligned}$$

L operador de rezago $y_t L = y_{t-1}$
Ecuación característica:

$$C(z) = 1 - \gamma_1 z - \gamma_2 z^2 - .... - \gamma_p z^p$$

raíces (ceros) por fuera del circulo unitario.

Media Movil MA(q)

Decomposición de Wold

$$y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$

truncando numero q parámetros

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}$$

donde $\varepsilon_t \sim WN(0, \sigma^2)$

Modelo autoregresivo y media móvil ARMA(p,q)

$\{y_t\}$ es ARMA(p,q)

$$y_t - \phi_1 y_{t-1} - \ldots - \phi_p y_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}$$

polinomio ARMA(p,q),

$$\Phi(L) y_t = \Theta(L) \varepsilon_t$$

con $\Phi(L) = 1 - \sum_{i=1}^{p} \phi_i L^i$ y $\Theta(L) = 1 - \sum_{i=1}^{q} \theta_i L^i$

Como determinar (p,q)?

- Descomposición de Wold
  - Filtro genera ruido blanco.

$$\begin{aligned} \frac{\Phi(L)}{\Theta(L)} y_t &= \varepsilon_t \\ \Psi(L) y_t &= \varepsilon_t \end{aligned}$$

- Metodo de Box-Jenking
  1. Identificación (ACF, PACF)
  2. Estimación
  3. Diagnostico
- Automatización
  - Auto-arima
  - Criterios Información

Función de autocorrelacion (ACF)

Autocovarianza

$$\lambda_k = Cov[y_t, y_{t-k}]$$

Autocorrelación λ_0, varianza

$$\hat{\rho}_k = \frac{\lambda_k}{\lambda_0}, \quad -1 \leq \rho_k \leq 1$$

- correlación bruta entre y_t y y_{t-k}.
- dependencia intertemporal.
- función decae en procesos estacionarios.
- decae menor ritmo, procesos larga memoria.

ACF, AR(1)

$$y_t = \gamma y_{t-1} + e_t$$

Autocovarianza

$$\begin{aligned}
Cov(y_t, y_{t+k}) &= E[y_t y_{t+k}] = E[y_t(\gamma y_{t+k-1} + e_{t+k})] \\
&= \gamma E[y_t y_{t+k-1}] + E[y_t e_{t+k}] \\
&= \gamma E[y_t(\gamma y_{t+k-2} + e_{t+k-1})] \\
&= \gamma^2 E[y_t y_{t+k-2}] + E[y_t e_{t+k-1}] \\
&\quad . \\
&\quad . \\
&= \gamma^k E[y_t y_t] \\
&= \gamma^k \lambda_0
\end{aligned}$$

ACF, AR(1)

Autocorrelación

$$\rho_k = \gamma^k$$

- serie geometrica.
- $\rho_0 = 1$
- decae monotónicamente $0 < \gamma < 1$
- cambio signos $-1 < \gamma < 0$

ACF MA(1)

$$y_t = e_t - \theta e_{t-1}$$

Autocovarianza

$$\lambda_0 = (1 + \theta^2)\sigma_e^2$$

$$\lambda_1 = -\theta\sigma_e^2$$

Autocorrelación

$$\rho_k = \frac{-\theta}{(1 + \theta^2)}$$

No depende de k trunca en $k = 1$.

Autocorrelación parcial PACF

Medir correlación neta entre y_t
y y_{t-2} eliminar efecto de y_{t-1}.
Función de autocorrelación parcial:

$$\rho_k^* = Corr(y_t - E^*(y_t|y_{t-1,\dots,}y_{t-k+1}), y_{t-k}),$$

donde

$$E^*(y_t|y_{t-1,\dots,}y_{t-k+1}) = \min_{\Phi} E[(y_t - \phi_1 y_{t-1} + \dots + \phi_{k+1} y_{t-k+1})^2]$$

Minimizar error/proyección ortogonal

Identificación Box-Jenkins

A partir de DGP

	ACF	PACF
AR(p)	$\downarrow$ Monotonically $\mapsto$ p	Truncated $\mapsto$ p
MA(q)	Truncated $\mapsto$ p	$\downarrow$ Monotonically $\mapsto$ p
ARMA(p,q)	$\downarrow$ Monotonically $\mapsto$ p	$\downarrow$ Monotonically $\mapsto$ p

Método Box-Jenkins

1. Test de Raíz Unitaria (ADF, KPSS).
2. Garantizar estacionariedad (retornos).
3. Identificación
   - gráfica (ACF, PACF)
   - automatizada AutoArima
4. Estimación.
5. Diagnostico: error estimado es ruido blanco?.
   - No, regresar a 2.
   - Si, proceder 5.
6. Pronóstico.

Estimación Máxima Verosimilitud

ARMA(p,q), dado valores p y q.

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}$$

$\varepsilon_t \sim i.i.d.N(0, \sigma^2)$, $\delta = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)$.
Si ε_t Gaussiana $\Rightarrow$ Máxima Verosimilitud.
Si ε_t no Gaussiana $\Rightarrow$ Quasi-Máxima Verosimilitud.

Máxima Verosimilitud, Gaussiana

Sea $\varepsilon_t \sim N(0, \sigma^2)$.

$$L_{\delta,\sigma^2}(\mathbf{y}) = \left(\frac{1}{2\pi}\right)^{n/2} \frac{1}{(\Sigma(\delta,\sigma^2))^{n/2}} \exp\{\frac{-1}{2}\mathbf{y}'(\Sigma(\delta,\sigma^2))^{-1}\mathbf{y}\}$$

Estimación de parámetros

$$\hat{\delta}_{ML} = argmin_{\delta \in T} \ln(L(\delta)) = argmin_{\delta \in T} \frac{1}{n}\mathbf{y}'(M(\delta))^{-1}\mathbf{y} \mid M(\delta) \mid^{1/n}$$

$$\hat{\sigma}_{ML} = \mathbf{y}'(M(\hat{\delta}_{ML}))^{-1}\mathbf{y}$$

Concentración de la función de verosimilitud: $\ln(L(\delta, \sigma(\delta)))$

Criterios de Información

Tamaño ARMA(p,q): trade-off ajuste dentro de la muestra y flexibilidad.

- Akaike: $AIC_{p,q} = -2 \ln L_{\hat{\delta},\hat{\sigma}^2;\text{Gaussian}} + 2(p+q+1)$, sobrestima (p,q)
- $AIC_{p,q} = -2 \ln L_{\hat{\delta},\hat{\sigma}^2;\text{Gaussian}} + \frac{2(p+q+1)T}{T-p-q-2}$, alta penalidad, para $AR(p)$ eficiente $T \to \infty$
- Bayesian $BIC_{p,q} = (T-p-q)\ln \frac{\hat{\sigma}^2}{T-p-q} + (p+q)\ln \frac{(\sum_{i=1}^{T} y_i^2 - T\sigma^2)}{p+q}$ consistente para (p,q), $T \to \infty$.

Escoger ARMA(p,q) minimiza criterios

Diagnostico

Modelo poblacional $\Phi(L)y_t = \Theta(L)\varepsilon_t$, $\varepsilon_t \sim i.i.d.N(0,\sigma^2)$.
Representación de Wold,

$$\varepsilon_t = \frac{\Theta(L)}{\Phi(L)}y_t = \Psi(L)y_t = \sum_{i=0}^{\infty}\psi_i y_{t-i}$$

Para unos valores (p,q) revisar $\varepsilon_t \sim i.i.d.N(0,\sigma^2)$:

- **Homoscedastisidad**.
- Distribución Normal.

Homoscedasticidad error

- Test Portmanteau, Rechazar $H_0 : \varepsilon_t \sim N(0, \sigma^2) \Leftrightarrow T \sum_{k=1}^{h} (\hat{\rho}_{\hat{\varepsilon}}(k))^2 > \lambda^2_{h-p-q,1-\alpha}$
- Mejorar desempeño en muestra finita.
- Test Q (Ljung–Box) $T \sum_{k=1}^{h} \frac{T+2}{T-k} (\hat{\rho}_{\hat{\varepsilon}}(k))^2 > \lambda^2_{h-p-q,1-\alpha}$

Pronostico

Pronostico 1 periodo adelante $y_{t+1|t}$ de y_t información en t.
Mean square error: Función de perdida (cuadrática)

$$MSE(y_{t+1|t}) = E(y_{t+1} - y_{t+1|t})^2$$

Métricas

- RMSE $\sqrt{\frac{1}{T}\sum_{i=t_1}^{T}(y_{i+1} - y_{i+1|i})^2}$
- MAE $\frac{1}{T}\sum_{i=t_1}^{T} \mid y_{i+1} - y_{i+1|i} \mid$

Combinación de pronósticos, Newbold and Granger, 1974

Sea $f_{1,t+s|t}$ y $f_{2,t+s|t}$ pronósticos de y_{t+s} con información en t. Si pronósticos son insesgados, $E(e_{i,t+s}) = 0$ para $i = 1, 2$, donde $e_{i,t+s} = y_{t+s} - f_{i,t+s|t}$ y σ_i^2 es error de pronostico y la varianza.
Pronostico combinado,

$$f_{c,t+s|t} = (1-\lambda)f_{1,t+s|t} + \lambda f_{2,t+s|t}$$

$\lambda \in [0, 1]$. Error pronostico combinado

$$e_{c,t+s} = y_{t+s} - f_{c,t+s|t} = (1-\lambda)e_{1,t+s} + \lambda e_{2,t+s}$$

$E(e_{c,t+s}) = 0$, y varianza

$$E(e^2_{c,t+s}) = (1-\lambda)^2\sigma_1^2 + \lambda^2\sigma_2^2 + 2\lambda(1-\lambda)\rho\sigma_1\sigma_2$$

Combinación de pronósticos

$E(e^2_{c,t+s})$ se minimiza si λ

$$\lambda_{opt} = \frac{\sigma_1^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2}$$

$(\sigma_1^2, \sigma_2^2, \rho)$ no observado, pero podemos tener el desempeño histórico de modelos $(e_{1,t}, e_{2,t})$, $t = 1, \ldots, T$, calcular equivalente muestral.

$$\hat{\lambda}_{opt} = \frac{\sum_{t=1}^{T} e^2_{1,t} - \sum_{t=1}^{T} e_{1,t}e_{2,t}}{\sum_{t=1}^{T} e^2_{1,t} + \sum_{t=1}^{T} e^2_{2,t} - 2\sum_{t=1}^{T} e_{1,t}e_{2,t}}$$

Test comparación desempeño pronostico, Morgan-Granger-Newbold

- observados $y_t, t = 1, \ldots, T$
- pronósticos $\tilde{y}_{1,t}$, $\tilde{y}_{2,t}$ $t = 1, \ldots, T$
- errores pronostico $e_{i,t} = \tilde{y}_{i,t} - y_t$
- perdida por pronostico $g(y_t, \tilde{y}_{i,t}) = g(\tilde{y}_{i,t} - y_t) = g(e_{i,t})$
- comparación desempeño $d_t = g(e_{1,t}) - g(e_{2,t})$

Hipótesis de igual desempeño $Ho : E(d_t) = 0 \; \forall t$, $Ha : E(d_t) = \mu, \mu \neq 0$
Supuestos del test,

1. función de perdida cuadrática.
2. pronostico 1 periodo adelante.
3. errores pronostico $e_{i,t} \sim N(0, \sigma_i^2)$.

Test comparación desempeño pronostico, Morgan-Granger-Newbold

bajo los anteriores supuestos determinamos ortogonalidad entre los errores,

$$x_t = e_{1,t} + e_{2,t}$$

$$z_t = e_{1,t} - e_{2,t}$$

$\Rightarrow Cov(x_t, z_t) = E(e^2_{1,t} - e^2_{2,t})$

Estadístico

$$MGN = \frac{r}{((1 - r^2)/(T - 1))^{1/2}}$$

donde $r = \frac{x'z}{((x'x)(z'z))^{1/2}}$, bajo la nula $MGN \sim student - t_{T-1}$

Test comparación desempeño pronostico, Diebold-Mariano

Test mas flexible (no sujeto a supuestos).
Calcular proceso desempeño diferenciado $\{d_i\}_{i=1}^{T}$, si proceso es débilmente estacionario.

$$\sqrt{T}(\bar{d} - \mu) \to N(0, 2\pi f_d(0))$$

donde $\bar{d}$ diferencia muestral promedio y $f_d(.)$ densidad espectral

$$f_d(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_d(k) exp(-ik\lambda), -\pi \leq \lambda \leq \pi$$

$$\bar{d} = \sum_{t=1}^{T} \frac{(g(e_{1,t}) - g(e_{2,t}))}{T}$$

Test comparación desempeño pronostico, Diebold-Mariano

Estadístico

$$DM = \frac{\bar{d}}{(2\pi \hat{f}_d(0)/T)^{1/2}}$$

Rechaza $Ho : E(d_t) = 0$ si $\mid DM \mid > Q_{N(0,1),1-\alpha}$,
$\hat{f}_d(0)$ estimada de manera no-paramétrica.