FORECASTING

SERIES DE TIEMPO

LECCIÓN 5

Applied Mathematics and Actuary Training

AMAT

Applied Mathematics and Actuary Training

FORECASTING

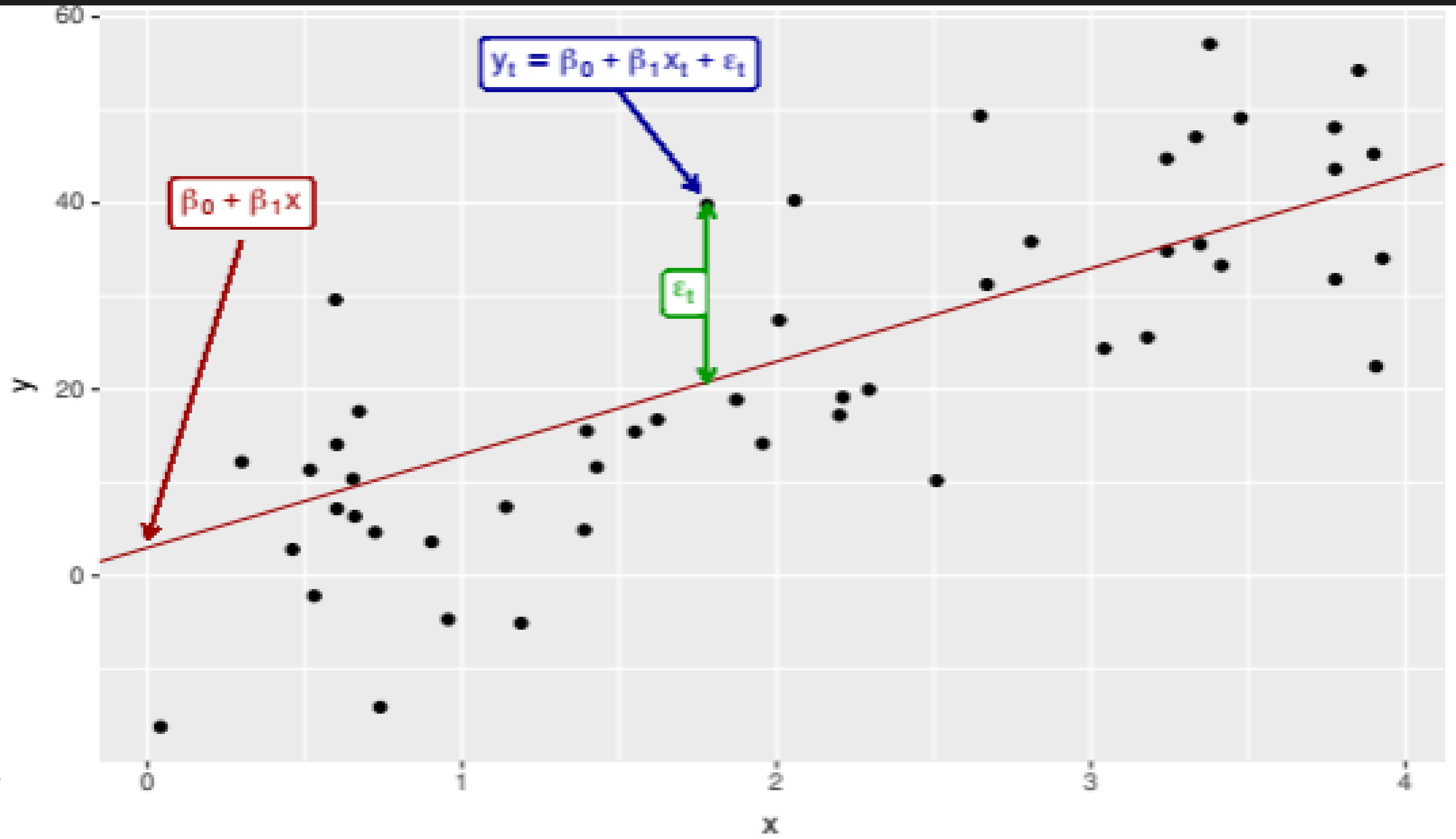
Lección 5 – Regresión

REGRESIÓN LINEAL
EN SERIES DE TIEMPO

ESTIMAR UNA ST $\{y_t\}$
Variable
Pronóstico
o Dependiente

ASUMIENDO

RELACIÓN LINEAL *con otra serie* $\{x_t\}$
Variable
Predictora
o Independiente



MODELO DE REGRESIÓN LINEAL SIMPLE

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$$

β_0: *Intercepto (valor de* $\boldsymbol{y}$ *cuando* $\boldsymbol{x} = \boldsymbol{0}$*)*
β_1 : *Pendiente (cambio en* $\boldsymbol{y}$ *cuando* $\boldsymbol{x}$ *aumenta una unidad)*
ε_t : *Error aleatorio (desviación del modelo lineal)*

MODELO DE REGRESIÓN LINEAL MÚLTIPLE

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \cdots + \beta_k x_{k,t} + \varepsilon_t$$

$x_1, \ldots, x_k$ *son las k variables predictoras*
$\beta_1, \ldots, \beta_k$: *Efecto marginale de c/predictor*
ε_t : *Error aleatorio*

SUPOSICIONES:

- La relación planteada por el modelo es verdadera
- Los predictores $\{x_{k,t}\}$ no son variables aleatorias
- Los errores $\{\varepsilon_t\}$:
  - Tienen media 0
  - No están autocorrelacionados
  - No tienen correlación con los predictores

PREDICTORES IMPORTANTES REGRESIÓN DE ST

TENDENCIA

Utilizamos:
$x_{1,t} = t$ *para modelar*

Es decir, $y_t = \beta_0 + \beta_1 t + \varepsilon_t$

Es posible utilizar potencias:
$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon_t$

ESTACIONALIDAD

Variable categórica, se crean variables dummy:

Ejemplo trimestral

	$s_{1,t}$	$s_{2,t}$	$s_{3,t}$
Trim1	1	0	0
Trim2	0	1	0
Trim3	0	0	1
Trim4	0	0	0

Y modelamos:
$y_t = \beta_0 + \beta_1 s_{1,t} + \beta_2 s_{2,t} + \beta_3 s_{3,t} + \varepsilon_t$

¡Podemos incluir ambas!
$y_t = \beta_0 + \beta_1 s_{1,t} + \beta_2 s_{2,t} + \beta_3 s_{3,t} + \beta_4 t + \varepsilon_t$

BONDAD DE AJUSTE

Múltiples Predictores

$\{\boldsymbol{x}_{1,t}\}$ ... $\{\boldsymbol{x}_{k,t}\}$

Seleccionar los mejores

Métricas de precisión predictiva

R^2	R^2 Ajustada	AIC	BIC
Coeficiente de Determinación $R^2 = \frac{\sum(\hat{y}_t - \bar{y})^2}{\sum(y_t - \bar{y})^2}$	***Coeficiente de Determinación Ajustado*** $R_a^2 = 1 - \frac{T-1}{T-k-1}(1-R^2)$	***Criterio de información de Akaike*** $AIC = 2(k-1) - 2\ln(L_k)$	***Criterio de información Bayesiano*** $BIC = \ln(L_k)(k-1) - 2\ln(L_k)$
Proporción de variación en variable pronóstico y_t explicada por la regresión	R^2 incrementa agregando predictores (incluso insignificantes) R_a^2 compensa	Penaliza por el número de parámetros k utilizado para el modelo	Penaliza más fuertemente el número de parámetros k utilizado para el modelo
$0 \leq R^2 \leq 1$ **Entre más grande mejor**	$0 \leq R_a^2 \leq 1$ **Entre más grande mejor**	Con pocos datos tiende a preferir más predictores **Entre más pequeño mejor**	Es más estricto con el sobreajuste en datasets grandes **Entre más pequeño mejor**

T: *Número de datos conocidos de la ST*
k: *Número de predictores del modelo*
L_k : *Logverosimilitud maximizada*

ANÁLISIS DE RESIDUALES

REMANENTE TRAS AJUSTAR UN MODELO

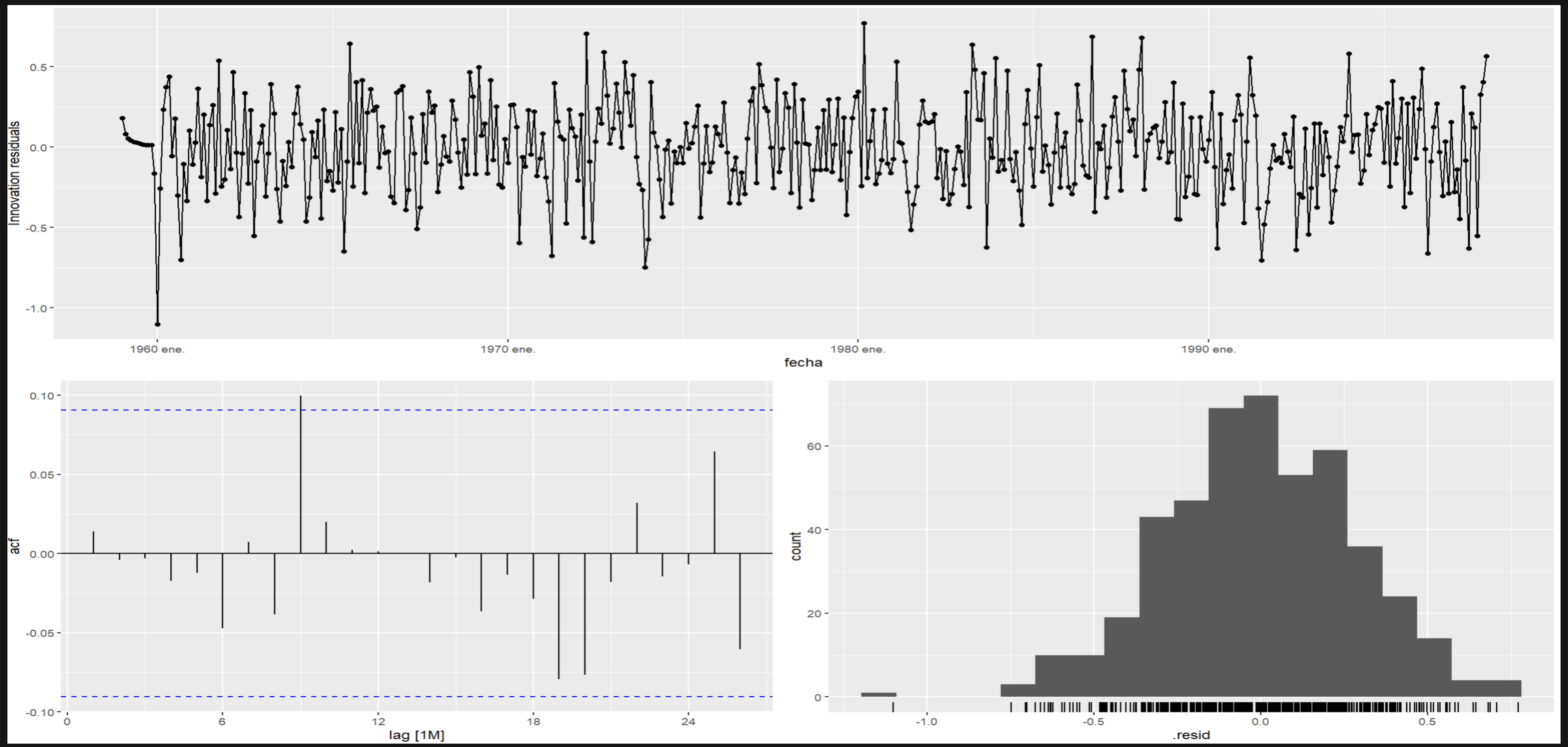
$$e_t = y_t - \widehat{y_t}$$

Son una Serie de Tiempo

Si el modelo es correcto.

Muestra de los errores aleatorios ε_t

BUSCAMOS	VERIFICAMOS
ALEATORIEDAD	Si existen patrones que nos indiquen lo contrario
MEDIA EN 0	Si oscilan arriba y abajo de 0 (entre más pequeños mejor)
NO AUTOCORRELACIÓN	Si presentan correlación serial



PRUEBAS FORMALES

H_0: *no hay autocorrelación* VS H_1: *hay autocorrelación*

Ljung-Box

$$Q = n(n+2)\sum_{k=1}^{h} \frac{\hat{\rho}_k^2}{n-k}$$

$$Q \sim \chi_h^2$$

Breusch-Godfrey

$$nR^2 \sim \chi_p^2$$

h: *número de lags que se prueban*

$\hat{\rho}_k^2$: *Autocorrelación de la muestra*

p: *Número de lags de errores (regresión aux)*

R^2: *Coef de determinación (regresión aux)*

FORECASTING

SERIES DE TIEMPO

Regresión| Lección 5

LECCIÓN 5