

Predicción

Series Cronológicas

Silvia Rodríguez Collazo

Facultad de Ciencias Económicas y de Administración

La predicción tal como la vamos a presentar en el curso está inevitablemente asociada a un modelo estocástico. En esta primer aproximación vamos a suponer que el modelo es conocido y es el correcto, que los parámetros son conocidos y están adecuadamente estimados.

En la realidad vamos a ver que la predicción va a resultar de un modelo que puede no ser el correcto y vamos a usar parámetros estimados a partir de una muestra de datos. Este modelo va a dar una caracterización estadística de la dinámica que vincula el pasado con el futuro.

Y la predicción es lo que en el presente se espera, condicionado al pasado que ocurrirá en el futuro.

Nos abocaremos a estudiar cómo de realizar predicciones sobre eventos futuros en base a la aplicación de modelos estadísticos.

Según Diebold,F.(2007), hay algunos elementos a tener en cuenta cuando se realiza la tarea de hacer predicciones mediante modelos.

- Definir la **función de pérdida** adecuada al problema planteado.
Esta decisión tiene implicancias, por ejemplo en términos de la evaluación de la predicción.

Las predicciones pueden ser usadas como guía de las decisiones a tomar. Consideremos el caso de una empresa que debe planear el inventario para un período. Este va a estar asociado a la demanda esperada.

Hay 4 posibles resultados: dos de ellas donde la predicción es correcta (la demanda subió o bajó y se pronosticó adecuadamente) y dos de ellas donde la predicción es errada (se pronostica aumento, cuando en realidad disminuye y viceversa). En los primeros dos casos la pérdida es 0. En los últimos dos casos hay una pérdida asociada, porque no se tuvieron los stocks requeridos para cubrir la demanda o porque fueron excesivos y se invirtió más de lo necesario.

¿Es el mismo el costo de equivocarse en un sentido que en otro (simétrico)?

No en todas las oportunidades es adecuado definir un costo simétrico, por lo que las pérdidas podrían ser diferentes en los dos errores.

- Definir el **objeto de la predicción**, si es una serie de tiempo (las ventas de una empresa, el producto de un país), la ocurrencia de un evento (pasar el nivel de 10% de inflación o entrar en el rango meta de inflación) o el momento de la ocurrencia de un evento (¿en qué mes o se pasa el 10% de inflación anualizada?, ¿en qué mes la inflación entra en el rango meta?). ¿Se predice una sola serie o muchas ?, ¿Contamos con una muestra larga?
- **En qué va a constar el pronóstico** que se realizará:
 - un valor “razonable”,
 - un rango de posibles valores futuros que reflejan la incertidumbre asociada a esa predicción
 - la distribución de probabilidad completa de los valores futuros.

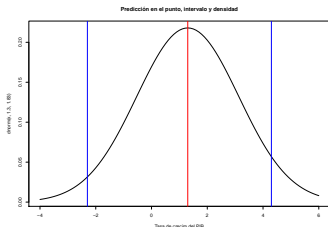
Si lo que queremos obtener es la **predicción en el punto**, un sólo número, por ejemplo caída del producto para este año.

Esto es un resultado sencillo de comprender a la hora de comunicar y para usar como guía de decisiones futuras pero van a haber shock futuros, impredecibles, que afecten la predicción.

Por tanto no podemos esperar un error = 0, por lo que es deseable poder saber el grado de incertidumbre asociada a esa predicción puntual.

- Un **intervalo de confianza**, es un rango de valores en los que se espera, que el valor de la series pueda ubicarse en el futuro, con cierta probabilidad. Es relevante la dimensión del intervalo, pues informa sobre la incertidumbre asociada.

Finalmente, se puede brindar información completa a través de la densidad de la predicción.

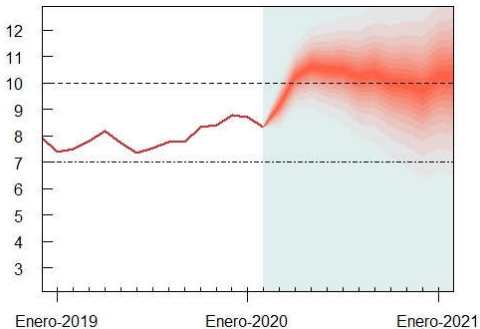


En el gráfico se representan tres tipos de predicción, en el punto, el intervalo y la densidad. La Densidad brinda mayor información, seguida por el intervalo y finalmente la predicción en el punto.

Pero puede ser más fácil de entender y comunicar la predicción puntual que un intervalo o una función de distribución.

Fan Chart

Proyección de la **inflación interanual** (en %) e **incertidumbre** y **Fan Chart**



- El **horizonte de predicción**: definir cuál es el horizonte de predicción, 1 mes, 1 año, 10 años? Esta definición condiciona el tipo de modelo a usar y la metodología a seleccionar.

El horizonte de predicción es el número de unidades de tiempo entre el último dato observado y la predicción a realizar. Si tenemos datos mensuales, estamos en t y queremos predecir hasta $t+12$, el horizonte de predicción es 12.

- Definir el conjunto de **información** en que se apoya la predicción. La calidad de la predicción está limitada por la calidad y cantidad de información que se disponga. La predicción es *condicional* a la información disponible. Esto refiere tanto a la cantidad de observaciones que se dispongan en la muestra como a si se cuenta con el pasado de la propia serie (modelo univariado), o si se construye un modelo en base a la información histórica de los determinantes (modelo multivariado). Veremos que la cantidad de datos con que contemos nos permitirá una mejor/peor evaluación de la predicción.
- **Método**. ¿Cuál es el mejor método de predicción para el problema que se tiene entre manos? ¿Cuál es el nivel de complejidad del mismo?. En los 70 cuando los macro modelos económicos dejaron de funcionar, los ARIMA entraron en acción.

En Ericsson (2001) se subraya el papel que la incertidumbre tiene en la predicción.

- **¿Qué es la incertidumbre?**

Es una medida que refleja la dispersión de los posibles resultados en base a la predicción realizada.

- **¿De qué depende la incertidumbre?**

Está asociada tanto a la ocurrencia de eventos futuros como al modelo asociado a la predicción.

- **¿Cómo debe ser medida la incertidumbre?**

Hay diversas medidas de incertidumbre que resumen las propiedades de las predicciones y sus errores de predicción.

- El sesgo de la predicción,
- la varianza del error de predicción ,
- la media del cuadrado de los errores (Error cuadrático medio, ECM) que combina las dos medidas anteriores.

La medida primaria de incertidumbre es el error cuadrático medio que se convierte en la varianza de los errores de predicción, cuando la predicción es insesgada.

- **¿Cómo esas medidas de incertidumbre deben ser usadas en la práctica?**

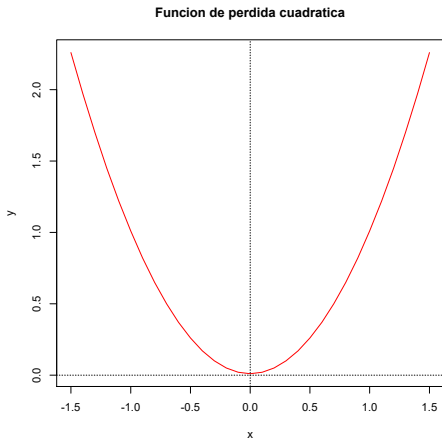
- Todo procedimiento de predicción está basado en un conjunto de información disponible hasta el momento t .
- La variable a ser predicha Y_{t+s} es una variable aleatoria, ella puede ser **completamente caracterizada** en términos de la función de densidad de probabilidad o alguna función equivalente.
- Dado que se utiliza un conjunto de información pasada y presente, I_t , es necesario trabajar con una función de densidad *condicional*. Si esta está disponible, entonces, otras propiedades de Y_{t+s} como la media condicional pueden ser determinadas.
- En general esta tarea es muy ambiciosa y **en lugar de obtener una completa caracterización de Y_{t+s} , se obtiene cierto intervalo de predicción para Y_{t+s} o un único valor, la predicción en el punto como el mejor «representante» de la variable en ese momento futuro.**

- Para obtener el valor de la predicción en el punto, se requiere definir un criterio a partir del cual otras alternativas puedan ser comparadas.
- Dado que hay que contar con la ocurrencia de errores de predicción vinculados a la aleatoriedad del proceso, podemos intentar identificar ese **costo**, es el efecto de incurrir en un error de predicción de tamaño e , que anotaremos como $C(e)$, tal que si no hay error, no es necesario asumir ningún costo, $C(0) = 0$.
- **Dada la función de costos, la predicción en el punto se elige de tal modo que se minimice el costo esperado y esa esperanza está condicionada al conjunto de información que se dispone.**
- La idea es incurrir en el mínimo costo posible, se seleccionará una función de predicción que haga mínimo ese costo. El objetivo es obtener errores de predicción pequeños.

- Consideraremos que esta función $C(e)$ es monótona no decreciente para $e > 0$ y monótona no creciente para $e < 0$, esto es, si hay un error distinto de cero, cuanto mayor es el error, mayor es el costo.
- Se pueden considerar distintas especificaciones para la función de pérdida: si nos es indiferente en términos de costos, que los errores sean positivos o negativos, implica considerar funciones **simétricas**, y por lo anterior prescindiremos del signo del error.
- Una función de costos particular muy utilizada es **la función de costos cuadrática**, $C(e) = a * e^2$ (donde a es una constante positiva).
- Esta función penaliza mucho más los errores de gran magnitud que los pequeños. Los *outliers* son muy costosos en esta función.

Tomaremos como objetivo minimizar el valor esperado de esta función simétrica y cuadrática del error: se minimiza el error cuadrático medio.

Función de costos cuadrática



Estamos interesados en predecir Y_{t+s} basados en un conjunto de variables Y_t , observadas en el momento t .

A la información disponible al momento t la anotamos como I_t .

En particular queremos predecir Y_{t+1} (predicción a 1 paso) en base a sus m valores más recientes.

Sea Y_t el conjunto de valores $Y_t, Y_{t-1}, \dots, Y_{t-m+1}$ y $Y_{t+1/t}^*$ el pronóstico de Y_{t+1} basado en la información presente y pasada de Y_t .

La Esperanza Condicional

Si se asume que la función de pérdida es cuadrática, se obtienen resultados muy convenientes.

Escogemos un predictor $Y_{t+1/t}^*$ de modo tal que $\min \left[E \left(Y_{t+1} - Y_{t+1/t}^* \right)^2 \right]$.

Esta expresión es el Error Cuadrático Medio (ECM) asociado con el pronóstico $Y_{t+1/t}^*$, y se escribe como:

$$ECM(Y_{t+1/t}^*) \equiv E(Y_{t+1} - Y_{t+1/t}^*)^2$$

El pronóstico con el menor ECM resulta ser la **Esperanza Condicional**, esto es, la **Esperanza** de Y_{t+1} **Condicional** a la información pasada y presente (disponible, en términos generales):

$$Y_{t+1/t}^* = E(Y_{t+1}/X_t)$$

OBS: La demostración está en las notas de clase

Restringimos ahora los pronósticos considerados a la clase de las funciones lineales de \mathbf{X}_t .

La predicción va a ser una combinación lineal de las observaciones pasadas. $Y_{t+1/t}^*$ será una función lineal de \mathbf{X}_t , una combinación lineal de los datos disponibles:

$$Y_{t+1/t}^* = \alpha' \mathbf{X}_t$$

Suponga que buscamos un α tal que el error de pronóstico $(Y_{t+1} - \alpha' \mathbf{X}_t)$ no está correlacionado con \mathbf{X}_t : $E[(Y_{t+1} - \alpha' \mathbf{X}_t) \mathbf{X}_t'] = 0'$.

Despejamos α de la expresión que lo define, $E[(Y_{t+1} - \alpha' \mathbf{X}_t) \mathbf{X}_t'] = 0'$:

$$\alpha = E(Y_{t+1} \mathbf{X}_t') [E(\mathbf{X}_t \mathbf{X}_t')]^{-1}$$

Cuando esta propiedad se cumple, entonces el pronóstico $\alpha' \mathbf{X}_t$ se denomina **Proyección Lineal** de Y_{t+1} en \mathbf{X}_t .

La Proyección Lineal produce el menor ECM entre la clase de los pronósticos lineales.

OBS: La demostración está en las notas.

La forma de realizar los pronósticos y los cálculos de los errores de predicción correspondientes que vamos a desarrollar a continuación se apoyan en un conjunto de supuestos:

- Se supone que se conoce toda la historia de la serie de tiempo Y_t
- El modelo ARMA para Y_t está correctamente identificado
- Conocemos los valores de los parámetros $\mu, \phi_i, \theta_j, \sigma^2$.

Teniendo presente lo anterior, consideremos procesos ARMA, estacionarios e invertibles. Si se cumplen estas dos condiciones, es posible expresar al proceso como un MA o un AR de orden infinito.

En lo que sigue se muestra como es posible plantear el pronóstico en función de los infinitos shocks pasados, en función de las infinitas observaciones pasadas. Estas representaciones son útiles para estudiar las propiedades estadísticas de las predicciones.

Considere un proceso ARMA, estacionario e invertible con una representación $MA(\infty)$, $(Y_t - \mu) = \psi(L)\varepsilon_t$

siendo ε_t un proceso Ruido Blanco y

$$\psi(L) \equiv \sum_{j=0}^{\infty} \psi_j L^j; \quad \psi_0 = 1 \text{ y } \sum_{j=0}^{\infty} |\psi_j| < \infty$$

Si tenemos información sobre ε_t hasta el momento t . Conocidos los parámetros μ y los $\{\psi_1, \psi_2, \psi_3, \dots\}$ podemos obtener la predicción.

Queremos pronosticar el valor que Y tomará dentro de s períodos (pronóstico a s pasos), Y_{t+s} .

$$Y_{t+s} = \mu + \varepsilon_{t+s} + \psi_1 \varepsilon_{t+s-1} + \dots + \psi_{s-1} \varepsilon_{t+1} + \psi_s \varepsilon_t + \psi_{s+1} \varepsilon_{t-1} + \dots$$

El **pronóstico óptimo** es la esperanza condicional toma la forma:

$$\hat{E}[Y_{t+s}/\varepsilon_t, \varepsilon_{t-1}, \dots] = \mu + \psi_s \varepsilon_t + \psi_{s+1} \varepsilon_{t-1} + \psi_{s+2} \varepsilon_{t-2} + \dots$$

Si la esperanza condicional a la información hasta el momento t de los shocks futuros son iguales a 0.

$$\hat{E}[\varepsilon_{t+s}/t] + \psi_1 \hat{E}[\varepsilon_{t+s-1}/t] + \dots + \psi_{s-1} \hat{E}[\varepsilon_{t+1}/t] = 0$$

Por lo que la predicción es insesgada.

OBSERVACIÓN SOBRE LA NOTACIÓN:

En estas notas se sigue la notación usada en el capítulo 4 de Hamilton(1994), en la que anota a la esperanza condicional como $\hat{E}[Y_{t+s}/\varepsilon_t, \varepsilon_{t-1}, \dots]$. Siendo esa esperanza el predictor óptimo, que hace mínimo el ECM. Alternativamente en el mismo capítulo también se utiliza $Y_{t+s/t}^*$ esta notación.

Donde los futuros ε 's, shocks, se igualan a su valor esperado de 0.

El **error de pronóstico** a s pasos, considerando que el modelo es el correcto y conocemos los parámetros poblacionales es:

$$Y_{t+s} - \hat{E}[Y_{t+s}/\varepsilon_t, \varepsilon_{t-1}, \dots] = \varepsilon_{t+s} + \psi_1 \varepsilon_{t+s-1} + \dots + \psi_{s-1} \varepsilon_{t+1}$$

El **ECM asociado** a este pronóstico es la varianza del error de predicción:

$$E \left[Y_{t+s} - \hat{E}[Y_{t+s}/\varepsilon_t, \varepsilon_{t-1}, \dots] \right]^2 = (1 + \psi_1^2 + \psi_2^2 + \dots + \psi_{s-1}^2) \sigma^2$$

Pronóstico de un proceso MA(q)

Sea un proceso MA(q):

$$X_t = \sum_{j=0}^q -\theta_j \varepsilon_{t-j}; \text{ con } \varepsilon_t \rightarrow RB; V(\varepsilon_t) = \sigma^2; -\theta_0 = 1$$

$$X_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

En este caso, que el

$$X_t = (1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q) \varepsilon_t = \theta(L) \varepsilon_t$$

proceso es de orden finito, los parámetros se anotan como θ podemos escribir $\psi(L) = \theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q$

El pronóstico lineal óptimo es:

$$\hat{E}[Y_{t+s} / \varepsilon_t, \varepsilon_{t-1}, \dots] = \begin{cases} \mu - \theta_s \varepsilon_t - \theta_{s+1} \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q+s} & s = 1, 2, 3, \dots, q \\ \mu & s = q+1, q+2, \dots \end{cases}$$

El ECM o la varianza del error de predicción es:

$$\begin{cases} \sigma^2 & \text{para } s=1 \\ (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_{s-1}^2) \sigma^2 & \text{para } s=2, 3, \dots, q \\ (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) \sigma^2 & \text{para } s > q \end{cases}$$

Podemos ver que el error se incrementa con el horizonte del pronóstico, hasta $s = q$ (cuando el horizonte de predicción es menor o igual al orden del MA).

Si queremos pronosticar un MA(q) más allá de q períodos en el futuro (el horizonte de predicción es mayor al orden del MA), el pronóstico es simplemente la media incondicional del proceso, y la varianza del error es la varianza incondicional.

Estas propiedades caracterizan también al caso MA(∞). Cuando $s \rightarrow \infty$ el pronóstico tiende a μ , la media incondicional, y la varianza del error tiende a $\sigma^2 \sum_{j=0}^{\infty} \psi_j^2$, la varianza incondicional del proceso expresado en forma MA(∞).

A continuación se deriva una expresión más compacta para el pronóstico a s pasos, que nos permitirá operar con más parsimonia:

$$\hat{E}[Y_{t+s}/\varepsilon_t, \varepsilon_{t-1}, \dots] = \mu + \left[\sum_{j=0}^{\infty} \psi_{s+j} L^j \right] \varepsilon_t = \mu + \left[\sum_{j=0}^{\infty} \frac{\psi_{s+j} L^{s+j}}{L^s} \right] \varepsilon_t =$$
$$\mu + \left[\sum_{j=0}^{\infty} \frac{\psi_j L^j}{L^s} \right]_+ \varepsilon_t = \mu + \left[\frac{\psi(L)}{L^s} \right]_+ \varepsilon_t$$

Whittle (1963) diseñó una notación para este operador: *Operador Aniquilación*:

$$\left[\frac{\psi(L)}{L^s} \right]_+ = \psi_s + \psi_{s+1}L + \psi_{s+2}L^2 + \dots$$

Este operador elimina todos los términos en los que el exponente de L es negativo, lo que da lugar a eliminar los términos adelantados.

Por lo que como resultado de la aplicación del operador, se iguala a cero la esperanza condicional de las innovaciones futuras y se mantienen las pasadas.

Lo que equivale a :

$$\hat{E}[\varepsilon_{t+s}/t] + \psi_1 \hat{E}[\varepsilon_{t+s-1}/t] + \dots + \psi_{s-1} \hat{E}[\varepsilon_{t+1}/t] = 0$$

Antes supusimos que los ε_t se observaban directamente, en la realidad vamos a contar con información pasada sobre Y_t y no sobre ε_t .

Consideremos que el proceso tiene una representación $AR(\infty)$:

$$\pi(L) Y_t = \varepsilon_t$$

$$\text{con: } \pi(L) = \sum_{j=0}^{\infty} \pi_j L^j; \pi_0 = 1, \sum_{j=0}^{\infty} |\pi_j| < \infty$$

Recordemos que los polinomios $AR(\infty)$ y $MA(\infty)$, están relacionados.

$$\pi(L) = [\psi(L)]^{-1}$$

Bajo estas condiciones, los pronósticos pueden escribirse usando el operador Aniquiación de la siguiente forma:

$$E[Y_{t+s}/Y_t, Y_{t-1}, \dots] = \left[\frac{\psi(L)}{L^s} \right]_+ \pi(L) Y_t = \left[\frac{\psi(L)}{L^s} \right]_+ \frac{1}{\psi(L)} Y_t$$

Pronóstico para un proceso AR(1)

Sea un proceso AR(1) estacionario:

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

Escribo el polinomio $\psi(L)$

$$\psi(L) = \frac{1}{1-\phi L} = 1 + \phi L + \phi^2 L^2 + \dots + \phi^j L^j + \dots$$

Ahora escribimos como quedaría el operador:

$$\left[\frac{\psi(L)}{L^s} \right]_+ = \phi^s + \phi^{s+1} L + \phi^{s+2} L^2 + \dots + \phi^{s+j} L^j + \dots = \frac{\phi^s}{1-\phi L}$$

El pronóstico a s pasos del AR(1):

$$\begin{aligned} \hat{E}[Y_{t+s}/Y_t, Y_{t-1}, \dots] &= \left[\frac{\psi(L)}{L^s} \right]_+ \frac{1}{\psi(L)} Y_t = \\ &= +\phi^s 1 - \phi L (1 - \phi L) Y_t = \phi^s Y_t \end{aligned}$$

$$\hat{E}[Y_{t+s}/Y_t, Y_{t-1}, \dots] = \phi^s Y_t$$

El pronóstico decae geométricamente desde Y_t hasta la media incondicional, a medida que el horizonte temporal s se incrementa.

La varianza del error de predicción para s períodos es:

$$\left[1 + \phi^2 + \phi^4 + \dots + \phi^{2(s-1)}\right] \sigma_\varepsilon^2.$$

Note que el error aumenta a medida que nos alejamos del presente, y cuando s tiende a infinito, se aproxima asintóticamente a $\frac{\sigma^2}{1-\phi^2}$, la varianza incondicional de Y_t .

Si el pronóstico es a 1 paso y la media es cero:

Entonces $s=1$ y $\mu = 0$

La esperanza condicional, el predictor sería:

$$\hat{E}[Y_{t+1}/Y_t, Y_{t-1}, \dots] = \phi Y_t$$

Sea un proceso MA(1) con media distinta de cero

$Y_t - \mu = (1 - \theta L) \varepsilon_t$ Escribimos el pronóstico aplicando el operador aniquilador:

$$\hat{E}[Y_{t+s}/Y_t, Y_{t-1}, \dots] = \mu + \left[\frac{\psi(L)}{L^s} \right]_+ \frac{1}{\psi(L)} (Y_t - \mu)$$

$$\hat{E}[Y_{t+s}/Y_t, Y_{t-1}, \dots] = \mu + \left[\frac{1 - \theta L}{L^s} \right]_+ \frac{1}{1 - \theta L} (Y_t - \mu)$$

Si la **predicción es a un paso** y $s = 1$, entonces tenemos que $\left[\frac{1 - \theta L}{L} \right]_+ = -\theta$, y por tanto la predicción sería la combinación lineal de los Y_t pasados

$$\hat{Y}_{t+1/t} = \mu - \frac{\theta}{1 - \theta L} (Y_t - \mu) = \mu - \theta (1 - \theta L - \theta^2 L^2 - \theta^3 L^3 - \dots) (Y_t - \mu)$$

Pero podemos escribirlo como:

$$\varepsilon_t = \frac{1}{1 - \theta L} (Y_t - \mu)$$

Donde ε_t es el resultado de la recursión:

$$\hat{\varepsilon}_t = (Y_t - \theta \hat{\varepsilon}_{t-1})$$

La predicción a un paso sería:

$$\hat{Y}_{t+1} = \theta \hat{\varepsilon}_t$$

OBS: La notación $\hat{\varepsilon}_t$ se introduce acá como un adelanto de lo que se tratará más adelante cuando se aproxime ε_t .

Qué debemos considerar al calcular en la práctica las predicciones utilizando los predictores descritos previamente.

Para hacer el salto a un número finito de observaciones debemos considerar para calcular las esperanzas condicionales:

$$\hat{E}(Y_{t-j}/t) = Y_{t-j} \quad \text{con } j=0,1,2,\dots$$

$$\hat{E}(Y_{t+j}/t) = \hat{Y}_{t+j} \quad \text{con } j=1,2,\dots$$

$$\hat{E}(\varepsilon_{t-j}/t) = \hat{\varepsilon}_{t-j} = (Y_{t-j} - \hat{Y}_{t-j/t-j-1}) \quad \text{con } j=0,1,2,\dots$$

Los errores son errores de predicción a 1 paso.

$$\hat{E}(\varepsilon_{t+j}/t) = 0 \quad \text{con } j=1,2,\dots$$

Para obtener la predicción de \hat{Y}_{t+j} , se escribe el modelo para Y_{t+j} en algunas de las formas antes vistas y se trata a los términos del lado derecho de acuerdo a las siguientes reglas:

- Y_{t-j} con $j = 0, 1, 2, \dots$, en el momento t , es pasado, por lo que permanece incambiado (se toma la realización).
- Y_{t+j} con $j = 1, 2, \dots$, que no han ocurrido se reemplazan por las predicciones.
- ε_{t-j} con $j = 0, 1, 2, \dots$, en el momento t , ya han ocurrido, se reemplazan por los errores de predicción: $(Y_{t-j} - \hat{Y}_{t-j/t-j-1})$
- $\varepsilon_{t+j}/t = 0$ con $j = 1, 2, \dots$, que no ha ocurrido se reemplaza por cero.

La rutina de cálculo se puede realizar a partir de la expresión de la ecuación en diferencias:

$$(1 - \phi_1 L, \dots, -\phi_p L^p)(Y_t - \mu) = (1 - \theta_1 L, \dots, \theta_q L^q)\varepsilon_t$$

Ejemplo: Predicción de un proceso AR(1) con número finito de observaciones

Consideremos el siguiente proceso:

$$Y_t = 0,8Y_{t-1} + \varepsilon_t$$

Escribimos la **predicción a un paso**, usando la expresión de la ecuación en diferencias.

$$Y_{t+1} = 0,8Y_t + \varepsilon_{t+1}$$

$$\hat{Y}_{t+1}/t = \hat{E}(0,8Y_t + \varepsilon_{t+1}/t) = 0,8E(Y_t/t) + E(\varepsilon_{t+1}/t) = 0,8Y_t$$

$$\hat{Y}_{t+1}/t = 0,8Y_t$$

Escribimos la **predicción a dos pasos**

$$Y_{t+2} = 0,8Y_{t+1} + \varepsilon_{t+2}$$

$$\hat{Y}_{t+2}/t = \hat{E}(0,8Y_{t+1} + \varepsilon_{t+2}/t) = 0,8E(Y_{t+1}/t) + E(\varepsilon_{t+2}/t) = 0,8\hat{Y}_{t+1}/t =$$

$$0,8 * 0,8 * Y_t = 0,8^2 * Y_t$$

$$\hat{Y}_{t+2}/t = 0,8^2 * Y_t$$

Ejemplo: Predicción de un proceso MA(1) con número finito de observaciones

Consideremos el siguiente proceso:

$$Y_t = 0,5\varepsilon_{t-1} + \varepsilon_t$$

Escribimos la **predicción a un paso**

$$\hat{Y}_{t+1}/t = \hat{E}(0,5\varepsilon_t + \varepsilon_{t+1}/t) = 0,5E(\varepsilon_t/t) + E(\varepsilon_{t+1}/t) = 0,5\hat{\varepsilon}_t$$

$$\hat{Y}_{t+1}/t = 0,5\hat{\varepsilon}_t$$

Escribimos la **predicción a dos pasos**

$$\hat{Y}_{t+2}/t = \hat{E}(0,5\varepsilon_{t+1} + \varepsilon_{t+2}/t) = 0,5E(\varepsilon_{t+1}/t) + E(\varepsilon_{t+2}/t) = 0$$

$$\hat{Y}_{t+2}/t = 0$$

Escribimos la **predicción a tres pasos**

$$\hat{Y}_{t+3}/t = 0$$

Ejemplo: Predicción de un proceso ARMA(1,1) con número finito de observaciones

Consideremos el siguiente proceso:

$$Y_t = 0,8Y_{t-1} + 0,5\varepsilon_{t-1} + \varepsilon_t$$

Escribimos la **predicción a un paso**

$$Y_{t+1} = 0,8Y_t + 0,5\varepsilon_t + \varepsilon_{t+1}$$

$$\hat{Y}_{t+1}/t = \hat{E}(0,8Y_t + 0,5\varepsilon_t + \varepsilon_{t+1}/t) = 0,8E(Y_t/t) + 0,5E(\varepsilon_t/t) + E(\varepsilon_{t+1}/t) =$$

$$\hat{Y}_{t+1}/t = 0,8Y_t + 0,5\varepsilon_t$$

Escribimos la **predicción a dos pasos**

$$Y_{t+2} = 0,8Y_{t+1} + 0,5\varepsilon_{t+1} + \varepsilon_{t+2}$$

$$\hat{Y}_{t+2}/t = \hat{E}(0,8Y_{t+1} + 0,5\varepsilon_{t+1} + \varepsilon_{t+2}/t)$$

$$\hat{Y}_{t+2}/t = 0,8(\hat{Y}_{t+1}/t) + 0,5(\hat{E}(\varepsilon_{t+1}/t) + \hat{E}(\varepsilon_{t+2}/t))$$

$$\hat{Y}_{t+2}/t = 0,8(\hat{Y}_{t+1}/t)$$

Escribimos la **predicción a tres pasos**

$$\hat{Y}_{t+3}/t = 0,8(\hat{Y}_{t+2}/t)$$

Calcular intervalos de predicción es una parte importante del proceso de predicción, pues es una forma de indicar la incertidumbre asociada a la predicción en el punto.

El intervalo de predicción consiste en un límite inferior y otro superior, entre los cuales se ubica la predicción en el punto con una cierta probabilidad preestablecida.

La predicción como valor futuro puede ser visto como una variable aleatoria evaluada en el momento en que se realiza la predicción.

Para construir el **intervalo de confianza para las predicciones en el punto** se requiere:

El **error asociado** con el pronóstico a s pasos condicional a la información hasta el momento t es:

$$Y_{t+s} - \hat{E}[Y_{t+s}/\varepsilon_t, \varepsilon_{t-1}, \dots] = \varepsilon_{t+s} + \psi_1 \varepsilon_{t+s-1} + \dots + \psi_{s-1} \varepsilon_{t+1}$$

La **varianza del error** de predicción a s pasos:

$$E \left[Y_{t+s} - \hat{E}[Y_{t+s}/\varepsilon_t, \varepsilon_{t-1}, \dots] \right]^2 = \left(1 + \psi_1^2 + \psi_2^2 + \dots + \psi_{s-1}^2 \right) \sigma^2$$

Intervalos de predicción para $\hat{Y}_{(t+s)/t}$

Observemos que estamos usando los ψ_j , que se pueden obtener, de acuerdo a lo que vimos antes a partir de los ϕ_k y los θ_j .

En este momento es necesario suponer una distribución para los ruidos, supondremos que las perturbaciones son procesos Ruidos Blancos **Gaussianos**, esto es $\varepsilon_t \sim N(0, \sigma^2)$.

Por tanto la distribución de probabilidad del valor futuro del proceso será la de una normal con media $\hat{Y}_{(t+s)/t}$ y desvío estándar $\sqrt{\sigma^2(1 + \sum_{j=1}^{s-1} \psi_j^2)}$

El **intervalo de confianza** sería:

$$\hat{Y}_{(t+s)/t} \pm z_{\frac{\alpha}{2}} \sqrt{\sigma^2(1 + \sum_{j=1}^{s-1} \psi_j^2)}$$

Con $\hat{Y}_{(t+s)/t} = E(Y_{t+s}/t)$ y $z_{\alpha/2}$ el percentil de la distribución normal estándar.

Se interpreta como el intervalo en que se espera se ubique $Y_{(t+s)}$ con una probabilidad de $1 - \alpha$

Intervalos de predicción para $\hat{Y}_{(t+s)/t}$ continuación

Cuando se supone que la distribución del error de predicción no es normal, obtener esa distribución es algo más complejo. Un ejemplo donde esta situación se da en forma frecuente es cuando se quiere predecir la volatilidad.

En ocasiones no es posible escribir analíticamente la forma de la densidad de predicción y se recorren alternativas para estimar diversos percentiles de la distribución de probabilidad condicional de los valores futuros.

$$\hat{Y}_{(t+s)/t} \pm z_{\frac{\alpha}{2}} \sqrt{\sigma^2 \left(1 + \sum_{j=1}^{s-1} \psi_j^2\right)}$$

Observen que el **intervalo es simétrico** en torno a la predicción en el punto, pero podemos encontrar intervalos asimétricos.

Un ejemplo: Si el modelo es formulado aplicando la transformación logarítmica y se desea obtener las predicciones en el punto y los intervalos de predicción de la serie en niveles, es necesario realizar la transformación inversa, de modo de recuperar las unidades de medida originales de la serie.

En ese caso el intervalo de predicción de la serie en niveles será asimétrico.

- Si la función de costos es proporcional al cuadrado de los errores de predicción, tenemos una función de costos cuadrática, y la predicción que hace mínima la pérdida esperada (óptima), es **la esperanza condicional** o la media de la función de densidad. Asociado a esta función de pérdida cuadrática viene el uso del ECM o su raíz como medida de resumen del desempeño predictivo.
- Una alternativa es la función de pérdida lineal, que se supone simétrica, donde la pérdida es proporcional al error absoluto de predicción y en ese caso la predicción óptima es **la mediana** de la distribución condicional. La correspondiente medida de resumen utilizada para evaluar el desempeño predictivo, es el error medio absoluto.
- Otra alternativa, asociada **al modo**, como medida central es la función de pérdida «todo o nada». La función de pérdida toma valor 0, si la predicción está dentro de un ε del resultado y asume valores positivos y constantes, si está fuera de ese rango. Cuando el $\varepsilon \rightarrow 0$, la pérdida esperada se minimiza por **el modo** de la distribución.

Intervalos de predicción , funciones de distribución y variantes

Hay instrumentos interesantes de visualización y muy útiles en la predicción económica, como por ejemplo los gráficos **fan charts**, que pueden ser vistos como una representación de un caso intermedio entre el intervalo de predicción y la densidad de predicción. Que es útil para comunicar la magnitud de la incertidumbre.

Los fan chart representan a la densidad de predicción, la distribución de probabilidad de la predicción para cada paso de la predicción. Cubre no solo el valor más probable del resultado que tomará la variable (modo), sino que ilustra el rango de posibles resultados a distintos intervalos de predicción.

Se construyen diversos intervalos de predicción para diferentes probabilidades, por ejemplo 10%, 30%, 50%, 70% y 90% y todos se incluyen en el mismo gráfico usando un degradé de colores para representar las diferentes probabilidades.

Los colores más fuertes son usados para los valores centrales y los más claros para las bandas más extremas. Se representa la predicción a varios pasos.

Parámetros clave en el fan chart:

- Proyección central central: modo, mediana, media
- Incertidumbre: Ancho de las bandas de confianza
- Balance de riesgo: simetría o asimetría. ¿Es igual el riesgo de que la inflación por ejemplo, se ubique por encima o por debajo del valore central?

La varianza de los errores de los modelos son más grandes cuando se calcula sobre la predicción que sobre el período de estimación del modelo.

Los errores de predicción tienden a ser más grandes que los residuos del modelo ajustado y por tanto los intervalos de confianza calculados sobre estos últimos son más estrechos de lo que deberían.

Más del 5% de las observaciones futuras, en promedio, caerán por fuera del intervalo de confianza construido al 95%, contruídos a partir de los residuos o errores en la muestra tomada para la estimación del modelo.

Es importante considerar que **el ajuste dentro de la muestra, es una mala guía respecto del desempeño futuro de la predicción**. La precisión de la predicción suele ser peor a lo esperado si se compara con el ajuste del modelo dentro de la muestra.

En Chatfield (2000) se enumeran algunas posibles razones para que estos intervalos sean tan estrechos:

- Porque los parámetros del modelo deben ser estimados
- En el caso de los modelos multivariados, se incluyen las predicciones de las exógenas.
- Cuando las innovaciones pueden no tener una distribución normal pero si una distribución con colas pesadas, o asimétrica, hay errores en los datos que pueden contaminar la distribución de los errores.
- Si el modelo identificado es incorrecto
- En caso que el modelo requiera cambios, ya sea durante el período de ajuste, cuando se inicia el período de predicción o durante el período de predicción.

Este tema es especialmente tratado en Ericsson, N. (2001) y menciona posibles fuentes de los errores de predicción basados en Clements y Hendry (1998) que los categorizan.

- ❶ Errores de predicción muy grandes producto de cambios ocurridos en la estructura de la economía
- ❷ Resultado de una mala especificación del modelo
- ❸ Problemas de medición en los datos en el período próximo al punto de inicio de la predicción
- ❹ Imprecisión en la estimación de los parámetros
- ❺ La acumulación de los errores futuros, los shocks futuros que impactan en la economía.

Estas fuentes de error son importantes cuando se analiza la incertidumbre de la predicción.

Las primeras tres fuentes representan **lo que no sabemos que no sabemos** . En cambio las últimas dos, si son predecibles, en el sentido que pueden ser anticipados e incluso calculados. La incertidumbre proviene de **lo que sabemos que no conocemos**.

- La incertidumbre de la predicción depende de la variable que ha de ser predicha (algunas implican mayores dificultades que otras)
- el tipo de modelo usado para predecir
- el proceso económico que determina la variable a ser predicha
- la información disponible
- el horizonte de predicción, la incertidumbre aumenta a medida que nos alejamos del momento t

¿Qué parte de la incertidumbre es predecible?

De acuerdo a Ericsson (2000), la discrepancia entre la predicción y el valor observado refleja la incertidumbre de la predicción. Dependiendo del grado de incertidumbre de la predicción, la predicción puede ser informativa o no.

Previo a la publicación del dato observado, la medida de la incertidumbre de la predicción ayuda a evaluar la incertidumbre a través de los errores de predicción, dando un panorama de los posibles valores en que puede ubicarse el dato observado.

Luego de que se publica el dato, el error de predicción y la incertidumbre de la predicción puede ser usada para evaluar el modelo del que proviene la predicción. Si el valor observado cae fuera del intervalo estimado, indica posibles problemas en la especificación del modelo

Vamos a distinguir entre la incertidumbre que es predecible de la que no es predecible.

Fuente: Ericsson (2000)

Fuentes de incertidumbre predecible: Lo que sabemos que no sabemos. Esto significa que la incertidumbre que proviene de estas fuentes es posible ser anticipada y calculada.

- Errores futuros que se acumulan a lo largo del período de predicción (“shocks”).
- Imprecisiones en las estimaciones de los parámetros del modelo, ya que en la práctica se utilizan coeficientes estimados y no los verdaderos parámetros.

Fuentes de incertidumbre impredecible: Lo que no sabemos que no sabemos. Estas fuentes no son predecibles ni se pueden anticipar. Si su magnitud y naturaleza fueran posibles de conocerse, se podrían incorporar en el modelo y la incertidumbre que los acompaña podría ser predicha.

- Futuros y desconocidos cambios estructurales que pueden ocurrir.
- Modelo/s usado/s para la predicción mal especificados.
- Datos con problemas de medición en el período base del modelo.

El cálculo de la incertidumbre predecible cumple roles importantes, tanto desde el punto de vista estadístico como desde la interpretación.

Desde el punto de vista estadístico, el cálculo de la incertidumbre predecible sienta las bases para el cálculo de intervalo de confianza o predicción (de acuerdo a la denominación de los diversos autores).

Como dijimos antes, previo a conocer el dato observado, los intervalos de confianza nos permiten valorar la incertidumbre esperada, permitiendo tener un panorama del rango de valores entre los que puede ubicarse el dato observado.

También nos sirve para evaluar el modelo del que proviene la predicción, si el dato observado, cae fuera del intervalo de predicción, esto puede ser por dos razones, o ocurrió un hecho que no era posible anticipar con el modelo o el modelo puede tener problemas de especificación.

El fan chart representa esa incertidumbre predecible.

Correlación entre los errores de predicción de distinto origen pero igual horizonte

Los errores de predicción a un paso están incorrelacionados. Pero esto no se cumple con los errores de predicción a varios pasos.

Sea $e_t(s)$, el error de predicción a s pasos a partir del momento t

Sea $e_{t-j}(s)$, el error de predicción a s pasos a partir del momento $t-j$

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

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

Si $j < s$, la autocovarianza es:

$$E(e_t(s), e_{t-j}(s)) = \sigma_\varepsilon^2 \sum_{i=j}^{s-1} \psi_i \psi_{i-j}$$

Ejemplo: $j = 2, s = 3$

$$e_t(3) = Y_{t+3} - \hat{Y}_{t+3/t} = \varepsilon_{t+3} + \psi_1 \varepsilon_{t+2} + \psi_2 \varepsilon_{t+1}$$

$$e_{t-2}(3) = Y_{t+1} - \hat{Y}_{t+1/t-2} = \varepsilon_{t+1}$$

$$COV(e_t(3), e_{t-2}(3)) = \psi_2 \sigma_\varepsilon^2$$

Correlación entre los errores de predicción de igual origen pero distintos horizontes

Sea $e_t(s)$, el error de predicción a s pasos a partir del momento t

Sea $e_t(s+j)$, el error de predicción a $s+j$ pasos a partir del momento t

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

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

$$E(e_t(s), e_t(s+j)) = \sigma_\varepsilon^2 \sum_{i=0}^{s-1} \psi_i \psi_{i+j}$$

Ejemplo: $j = 2, s = 3, s+j = 5$

$$e_t(3) = Y_{t+3} - \hat{Y}_{t+3/t} = \varepsilon_{t+3} + \psi_1 \varepsilon_{t+2} + \psi_2 \varepsilon_{t+1}$$

$$e_t(5) = Y_{t+5} - \hat{Y}_{t+5/t} = \varepsilon_{t+5} + \psi_1 \varepsilon_{t+4} + \psi_2 \varepsilon_{t+3} + \psi_3 \varepsilon_{t+2} + \psi_4 \varepsilon_{t+1}$$

$$\boxed{COV(e_t(3), e_t(5)) = \psi_2 \sigma_\varepsilon^2 + \psi_1 \psi_3 \sigma_\varepsilon^2 + \psi_2 \psi_4 \sigma_\varepsilon^2}$$

La transformación logarítmica es usualmente empleada con el propósito de obtener una varianza más homogénea en la serie transformada.

Cuando se estima un modelo ARMA a la serie transformada, se obtienen las predicciones de la variable transformada y es necesario aplicar la operación inversa para deshacer la transformación y obtener las predicciones de las variables en niveles.

En Granger y Newbold (1976) se advierte que la transformación no lineal aplicada a la predicción óptima de la variable puede no resultar en una predicción óptima de la variable transformada.

En particular cuando la serie de la que se obtienen las predicciones es la serie expresada en logaritmos y a ésta se le aplica la función exponencial, la predicción resultante (de la variable en niveles) en general no es óptima.

Sea $X_t = \log(Y_t)$, el logaritmo natural de la serie de tiempo univariada Y_t . X_t es generada por un proceso ARMA(p,q), estacionario e invertible. Sea $\hat{X}_{t+s/t}$,

el predictor óptimo de la serie transformada,

$$\hat{X}_{t+s/t} = E(X_{t+s}/X_t, X_{t-1}, X_{t-1}, \dots)$$

La predicción de la serie en niveles se puede obtener aplicando la operación inversa:

$$\hat{y}_{t+s/t} = \exp(\hat{x}_{t+s/t})$$

Granger y Newbold (1976) muestran que cuando se aplica la transformación logarítmica, la predicción óptima se obtiene :

$$\hat{y}_{t+s/t} = \exp(\hat{x}_{t+s/t} + \frac{1}{2}\sigma_x^2(s))$$

Lutkepohl y Xu (2009) investigan en qué condiciones aplicar la transformación logarítmica es beneficioso en términos de predicción.

Concluyen que si la transformación logarítmica verdaderamente estabiliza la varianza, la predicción mejora, se obtienen errores de predicción menores (en términos del ECM).

Por el contrario, si la transformación logarítmica es aplicada aunque no resulte en una varianza más homogénea, la precisión de la predicción se ve afectada. Adicionalmente muestra cómo no se obtienen ganancias sustanciales en términos de predicción de la serie en niveles si se aplica la transformación «ingenua», en lugar de la transformación que da lugar a la predicción óptima propuesta en Granger et,al(1976).

Con la actualización de las predicciones se incorporan las novedades que traen los nuevos datos y con ello se fortalece la capacidad de **interpretación temprana de los potenciales cambios de rumbo** en la evolución de la variable.

A través de los gráficos que vemos a continuación se representa el proceso de corrección y actualización de los pronósticos.

Esta representación toma la forma de un “cabello florecido”, donde cada florecimiento es la nueva predicción con datos hasta un cierto período y “cada nuevo florecimiento” muestra cómo se corrige la trayectoria esperada de la variable al incorporar los nuevos datos observados.

La corrección incorpora la nueva información que trae consigo cada dato observado al ser incorporado al modelo.

La información nueva desencadena esa corrección y con ello permite detectar tempranamente los cambios que se están procesando y que podrán consolidarse en el correr de los días o meses...

Este sistema de predicción incorpora otras de las ventajas que trae la actualización cada pocas unidades de tiempo. Se reducen los errores y se corrigen rápidamente las trayectorias pronosticadas. En momentos donde el cambio se está procesando, esto constituye una señal a atender e interpretar.

Proceso de actualización de las predicciones del acumulado de casos positivos para Uruguay MARZO 2021

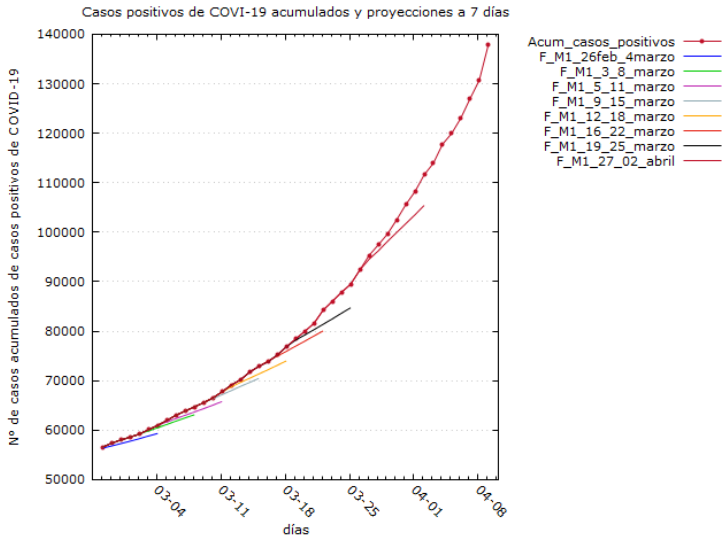


Figura: Fuente: SINAIE y www.doornik.com/COVID-19. Datos observados y Confirmed count forecast Latin America

La autocorrelación parcial poblacional m -ésima $\alpha_m^{(m)}$ se define como la correlación lineal entre X_t , X_{t+m} , habiendo quitado el efecto de las correlaciones lineales intermedias, es el último coeficiente de una proyección lineal de X sobre sus m más recientes valores:

$$\hat{X}_{t+1/t,t-1,\dots,t-m+1} = \alpha_1^{(m)} X_t + \alpha_2^{(m)} X_{t-1} + \dots + \alpha_m^{(m)} X_{t-m+1}$$

$$\hat{X}_{t+1/t,t-1,\dots,t-m+1} = \begin{pmatrix} \gamma_1 & \gamma_2 & \dots & \gamma_m \end{pmatrix} \begin{pmatrix} \gamma_0 & \gamma_1 & \dots & \gamma_{m-1} \\ \gamma_1 & \gamma_0 & \dots & \gamma_{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m-1} & \dots & \dots & \gamma_0 \end{pmatrix}^{-1} \begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-m+1} \end{pmatrix}$$

Función de Autocorrelación Parcial (FACP)

Entonces, $\alpha_m^{(m)}$ surge es el último elemento del vector:

$$\begin{pmatrix} \alpha_1^{(m)} \\ \alpha_2^{(m)} \\ \vdots \\ \alpha_m^{(m)} \end{pmatrix} = \begin{pmatrix} \gamma_0 & \gamma_1 & \cdot & \gamma_{m-1} \\ \gamma_1 & \gamma_0 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \gamma_{m-1} & \cdot & \cdot & \gamma_0 \end{pmatrix}^{-1} \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_m \end{pmatrix}$$

La autocorrelación parcial $\alpha_m^{(m)}$ puede entonces pensarse como la correlación existente entre X_{t+1} y X_{t-m+1} , una vez ajustada por el efecto de $X_t, X_{t-1}, \dots, X_{t-m+2}$.

Note que $\alpha_1^{(1)} = \rho_1$

En general

$$\alpha_m^{(m)} =$$

$$\text{Corr}(X_{t+1} - E(X_{t+1} / \{X_t, \dots, X_{t-m+2}\}); X_{t-m+1} - E(X_{t-m+1} / (X_t, \dots, X_{t-m+2}))); \quad k \geq 2$$

La FACP $\alpha_m^{(m)}$ para $m \geq 2$ es entonces la correlación de los 2 residuos obtenidos luego de regresar X_{t+1} y X_{t-m+1} sobre las observaciones intermedias $X_t, X_{t-1}, \dots, X_{t-m+2}$.

Si los datos X_t fueron generados por un $AR(p)$, entonces $\alpha_m^{(m)} = 0 \ \forall m > p$.

Si los datos X_t fueron generados por un $MA(q)$, entonces $\alpha_m^{(m)}$ converge asintóticamente a 0.

El estimador natural de $\alpha_m^{(m)}$ es el último coeficiente de una Regresión Mínimo Cuadrática (MICO) de X_{t+1} sobre una constante y sus m valores más recientes.

Cuando se tienen varios tipos de modelos y se realizan las predicciones con ellos, es útil poder compararlos en términos de su precisión. En Hyndman *et al* (2006) se presentan algunas medidas para ese fin.

Medidas dependientes de la escala de los datos

Estas medidas, pueden ser útiles cuando se compara la predicción de diferentes modelos para una misma serie, pero no entre predicciones de datos que tienen distinta escala.

- Media de los errores al cuadrado (MSE) = $\text{mean}(e_t^2)$
- Raíz de la media de los errores al cuadrado ($RMSE$) = \sqrt{MSE}
- Media del error absoluto (MAE) = $\text{mean}|e_t|$
- Mediana del error absoluto ($MdEA$) = $\text{median}|e_t|$

$RMSE$ y MSE son más sensibles frente a la existencia de puntos raros que MAE y $MdEA$, por ello muchos autores no recomiendan sus uso.

Medidas basadas en los errores porcentuales:

El porcentaje de error es: $p_t = 100 * e_t / Y_t$, es una medida independiente de la escala de medida de la serie.

- Media absoluta del porcentaje de error (MAPE) = $mean|p_t|$
- Mediana absoluta del porcentaje de error (MdAPE) = $median|p_t|$
- Raíz de la media al cuadrado del porcentaje de error (RMSPE) = $\sqrt{mean(p^2)}$
- Raíz de la mediana del cuadrado del porcentaje de error (RMdSPE) = $\sqrt{median(p^2)}$

Medidas basadas en los errores relativos:

El error relativo se define por ejemplo como: $r_t = e_t / e_t^*$, siendo e_t^* el error del modelo elegido como *benchmark*

- Media relativa del error absoluto (MRAE) = $mean|r_t|$
- Mediana relativa del error absoluto (MdRAE) = $median|r_t|$
- Raíz de la media al cuadrado del porcentaje de error(RMSPE) = $\sqrt{mean(p^2)}$
- Raíz de la mediana del cuadrado del porcentaje de error(RMdSPE) = $\sqrt{median(p^2)}$

Se recomienda la lectura de Hyndman et al (2006)

- Chatfield,Ch. (2000) "Time Series Forecasting". Chapman Hall/CRC
- Dieblod,F. (2007) "Elements of Forecasting". Thomson.South-Western. Fourth Edition
- Ericsson,N.R. (2000) "Predictable Uncertainty in Economic Forecasting".Board of Governors of the Federal Reserve System. International Finance Discussion Papers. Number 695.
- Ericsson,N.R. (2001) "Forecast Uncertainty in Economic Modeling". Board of Governors of the Federal Reserve System. International Finance Discussion Papers. Number 697.
- Granger,C.W.J and Newbold,P. (1976) - « Forecasting transformed series». *Journal of the Royal Statistical Society B* **38**.
- Hamilton, J. (1994) " *Time Series Analysis*". Princeton University Press
- Hyndman,R.J.;Koejler,A.B. (2006) " Another look at measures of forecast accuracy ". *International Journal of Forecasting*
- McLeod,A. Li,W. (1983) "Diagnostic Checking ARMA Time Series Models Using Squared Residual Autocorrelations" *Journal of Time Series Analysisi Vol4. N° 4*

- Lutkepohl,H. y Xu,F.(2009) - «The Role of the log Trnasformation in Forecasting Economic Variables». CESinfo Working Paper N° 2591. Marzo.
- Peña, D. (2005) *"Análisis de series temporales"* Alianza Editorial.
- Rodríguez-Collazo,S. (2023) Notas del curso de Series Cronológicas. Licenciatura de Estadística. Facultad de Ciencias Económicas y Administración.
- Rodríguez-Collazo,S. (2021) "Modelos de series temporales adaptativos y robustos para predecir contagios y muertes por COVID-19". Presentación en el Seminario del IESTA, 14 de abril de 2021.En base a las Notas de Reflexión elaboradas, con igual título.
https://iesta.fcea.udelar.edu.uy/wp-content/uploads/2021/02/Nota_tecnica_Reflexiones_en_tiempos_de_Pandemia_9.pdf