

# Evaluación de pronóstico y comparación de modelos

Fernando Arias-Rodríguez

Banco Central de Bolivia

30 de agosto de 2024



- 1 ¿Por qué es importante?
- 2 Prueba de Diebold-Mariano
- 3 Prueba de Li-Liao-Quaedvlieg
- 4 Conjuntos de confianza de modelos

- 1 ¿Por qué es importante?  
Acercamiento teórico al problema
- 2 Prueba de Diebold-Mariano
- 3 Prueba de Li-Liao-Quaedvlieg
- 4 Conjuntos de confianza de modelos

- A lo largo del curso, hemos visto procedimientos de validación interna de cada uno de los modelos propuestos.
- Sin embargo, es importante evaluar cuál de todos los pronósticos es mejor si se evalúa *entre modelos*.
- Así, la evaluación de pronóstico se erige como una manera de comparar distintos modelos, tratando de escoger el más adecuado para la tarea.

- 1 ¿Por qué es importante?  
Acercamiento teórico al problema
- 2 Prueba de Diebold-Mariano
- 3 Prueba de Li-Liao-Quaedvlieg
- 4 Conjuntos de confianza de modelos

- Sea  $\mathcal{M}_h$  un conjunto de modelos de pronóstico alternativos para la variable objetivo  $y$ , en el horizonte  $t + h$ ,  $h = 1, 2, \dots, H$ .
- Cada modelo está definido en términos de un vector de parámetros  $\theta_m$ . Sea  $\theta_m^*$  los parámetros pseudo-ciertos y  $\hat{\theta}_m$  sus estimadores.
- Sea  $\hat{y}_{m,t+h|t} = \hat{y}_{m(\hat{\theta}_m,t+h|t)}$  el pronóstico de  $y_{t+h}$  producido por un modelo  $m \in \mathcal{M}_h$ .
- La función de pérdida asociada al  $m$  –esimo modelo se puede escribir como:

$$\mathcal{L}_{m(\hat{\theta}),t+h} = \mathcal{L}(y_{t+h}, \hat{y}_{m,t+h|t}), t = R, \dots, T - h \quad (1)$$

- La igualdad de pronósticos producidos por dos modelos diferentes,  $m_1$  y  $m_2$ , se pueden comparar a partir de la siguiente prueba de hipótesis:

$$\mathcal{H}_0 : E\left[\mathcal{L}_{m_1(\hat{\theta}_1),t+h}\right] = E\left[\mathcal{L}_{m_2(\hat{\theta}_2),t+h}\right]$$

- Nótese que las funciones de pérdida se encuentran en términos de los parámetros estimados para cada modelo a comparar. Así, se dice que esta prueba de hipótesis se aplica sobre muestras y no poblaciones.
- Los métodos que se revisarán se basan en esta prueba de hipótesis.

- 1 ¿Por qué es importante?
- 2 Prueba de Diebold-Mariano
- 3 Prueba de Li-Liao-Quaedvlieg
- 4 Conjuntos de confianza de modelos



- Propuesto por Diebold y Mariano (1995).
- Considere dos modelos  $m_1$  y  $m_2$ , los cuales producen dos secuencias de pronósticos  $\{\hat{y}_{m_1,t+h|t}\}_{t=R}^{T-h}$  y  $\{\hat{y}_{m_2,t+h|t}\}_{t=R}^{T-h}$ , con sus respectivas secuencias de *errores de pronóstico*  $\{\hat{z}_{m_1,t+h|t}\}_{t=R}^{T-h}$  y  $\{\hat{z}_{m_2,t+h|t}\}_{t=R}^{T-h}$ .
- La prueba comienza definiendo  $d_{(12),h,t} = \mathcal{L}_{m_1,t+h} - \mathcal{L}_{m_2,t+h}$ ,  $t = R, \dots, T-h$ , que se interpreta como la diferencia en pérdida entre los dos modelos.
- El estadístico de la prueba está dado por:

$$DM = \frac{\bar{d}_{(12),h}}{\hat{\sigma}(\bar{d}_{(12),h})} \quad (2)$$

con  $\bar{d}_{(12),h} = \frac{1}{P_h} \sum_{t=R}^{T-h} d_{(12),h,t}$  y  $\hat{\sigma}(\bar{d}_{(12),h})$  es un estimador de la varianza del promedio de  $d_{h,t}$ .

- El estadístico DM puede ser rápidamente calculado como el estadístico t de una regresión entre el diferencial de pérdidas ( $d_{(12)}$ ) y un intercepto.
- La prueba DM se puede extender al controlar por variables adicionales que puedan explicar la diferencia de errores (Giacomini y White (2006) discuten el tema con profundidad).
- Harvey, Leybourne y Newbold (1997) proponen un ajuste para el caso de muestra pequeña:

$$MDM = \sqrt{\frac{P_h + 1 - 2h + h(h-1)/P_h}{P_h}} x DM \quad (3)$$

- 1 ¿Por qué es importante?
- 2 Prueba de Diebold-Mariano
- 3 Prueba de Li-Liao-Quaedvlieg
- 4 Conjuntos de confianza de modelos

- DM se basa en el desempeño promedio *incondicional* de los pronósticos de los modelos que compiten.
- Li, Liao y Quaedvlieg (2021) proponen una prueba *condicional* para medir la habilidad predictiva.
- La hipótesis nula establece que la pérdida esperada condicionada de un modelo de referencia no es mayor que la de sus alternativas. Esto es cierto uniformemente a través de todos los estados condicionales.
- Tales estados condicionales se determinan a partir de una variable condicionante, escogida previamente por el investigador.

- Partiendo de la misma idea de prueba de hipótesis ya revisada, sea  $d_{(1m),h,t}$  el diferencial de pérdida entre el modelo  $m \in \mathcal{M}_m$  y el modelo de referencia,  $m_1$ .
- A partir de una variable condicionante, definida por el usuario,  $C_t$ , se define:

$$h_{m,h}(c) = E(d_{(1m),h,t} | C_t = c) \quad (4)$$

- $h_{m,h}(c) \geq 0$  indica que se espera que el método de referencia se desempeñe mejor que el competidor, condicional a  $C_t = c$ . En otras palabras, la hipótesis nula de la prueba se puede reescribir así:

$$\mathcal{H}_0 : h_{m,h}(c) \geq 0, \quad \forall c \in C \text{ \& } m \in \mathcal{M}_h \quad (5)$$

- Por la ley de expectativas iteradas, que no se rechace la hipótesis nula implica que la pérdida esperada incondicional del modelo de referencia es menor que la de los modelos competidores.
- Es más, el criterio de dominancia condicionada puede ayudar al investigador a diferenciar modelos que incondicionalmente puedan lucir similares (mejora a DM).

- 1 ¿Por qué es importante?
- 2 Prueba de Diebold-Mariano
- 3 Prueba de Li-Liao-Quaedvlieg
- 4 Conjuntos de confianza de modelos

- Conocido como *Model Confidence Sets* (MCS).
- MCS comienza con un conjunto de modelos  $\mathcal{M}^0$  de dimensión  $M$ , que incorpora todas las especificaciones de modelos disponibles para el investigador y arroja, para un nivel de confianza  $1 - \alpha$ , un conjunto más pequeño  $\mathcal{M}_{1-\alpha}^*$  de dimensión  $M^* \leq M$ .
- $\mathcal{M}_{1-\alpha}^*$  es un conjunto de todos los modelos superiores.
- El mejor escenario es cuando el este conjunto incluye un solo modelo, es decir,  $M^* = 1$ .



- Sea  $d_{mn,t+h}$  la diferencia de pérdida entre los modelos  $m$  y  $n$ :  
 $d_{mn,t+h} = \mathcal{L}(m, t+h) - \mathcal{L}(n, t+h)$ ,  $m, n = 1, \dots, M$  y  
 $t = R, \dots, T-h$ .
- Definimos:

$$d_{m,\cdot,t+h} = \frac{1}{M-1} \sum_{n \in \mathcal{M}} d_{m,n,t+h}, m = 1, \dots, M \quad (6)$$

la pérdida del modelo  $m$  relativo a cualquier otro modelo  $n$  en el tiempo  $t+h$ .

- La hipótesis de misma habilidad predictiva (EPA) para un conjunto de modelos  $\mathcal{M}$  se puede formular de dos maneras alternativas:

$$\mathcal{H}_{0,\mathcal{M}} : c_{mn} = 0 \quad \forall m, n = 1, \dots, M$$

$$\mathcal{H}_{A,\mathcal{M}} : c_{mn} \neq 0 \quad \text{para algunos } m, n = 1, \dots, M$$

o

$$\mathcal{H}_{0,\mathcal{M}} : c_{m\cdot} = 0 \quad \forall m = 1, \dots, M$$

$$\mathcal{H}_{A,\mathcal{M}} : c_{m\cdot} \neq 0 \quad \text{para algunos } m = 1, \dots, M$$

donde  $c_{mn} = E(d_{mn,t+h})$  y  $c_{m\cdot} = E(d_{m\cdot,t+h})$

- Para evaluar las hipótesis propuestas, se construyen los siguientes estadísticos:

$$t_{mn} = \frac{\bar{d}_{mn}}{\hat{\sigma}(\bar{d}_{mn})} \quad \text{y} \quad t_{m.} = \frac{\bar{d}_{m.}}{\hat{\sigma}(\bar{d}_{m.})}$$

$$\bar{d}_{mn} = \frac{1}{P_h} \sum_{t=R}^{T-h} d_{mn,t+h} \quad \text{y} \quad \bar{d}_{m.} = \frac{1}{M-1} \sum_{n \in \mathcal{M}} \bar{d}_{mn}$$

$\hat{\sigma}(\bar{d}_{mn})$  y  $\hat{\sigma}(\bar{d}_{m.})$  son estimaciones de las desviaciones estándar de  $\bar{d}_{mn}$  y  $\bar{d}_{m.}$ , respectivamente.