

¿Cómo medir la precisión de los pronósticos?

Por Tomás Gálvez Martínez

Presidente y Director de CELOGIS

Education Partner de ENAE Business School

A la fecha de la publicación de este documento usted podrá encontrar, en la mayoría de artículos y libros especializados en el tema y en algunas soluciones tecnológicas dedicadas al cálculo de pronósticos, más de 25 modelos y más de 15 formas para medir la efectividad de ellos. El propósito de este artículo es exponer las principales herramientas de precisión de los pronósticos considerando tres interrogantes pertinentes en la evaluación del modelo.

1. ¿Los modelos que mejor se ajustan a los datos históricos son igual de efectivos al pronosticar?
2. ¿Cuántas medidas de precisión existen y en qué casos aplicar cada una de ellas?
3. ¿Qué otros criterios existen para la selección de los modelos de pronóstico?

Al final es posible que no dispongamos de un sólo criterio de selección, además del costo que ello implica, pero tendremos más información que nos ayudará a la evaluación del resultado de los pronósticos. Se presentan recomendaciones para medir la efectividad del modelo de pronóstico durante el proceso de selección y después de él.

1. ¿Los modelos que mejor se ajustan a los datos históricos son igual de efectivos al pronosticar?

Existe evidencia significativa que un modelo que ajusta bien a los datos históricos, no necesariamente pronostica bien. ¿Entonces para qué evaluar la precisión de un modelo de pronóstico en función de su capacidad de ajuste a los datos históricos? ¿Existen alternativas para evaluar un modelo en función de su capacidad de pronosticar?

Una de las alternativas es dividir la información en dos conjuntos. El primer conjunto lo podemos denominar *datos de control* y utilizarlo para encontrar el mejor modelo de pronóstico. Y al segundo conjunto etiquetarlo como *datos de predicción* y no incluirlos en el análisis inicial. La cuestión es ¿qué cantidad de datos tomar para cada conjunto de información? Por ejemplo si usted dispone de 4 años de información histórica divididos en forma mensual, puede tomar los tres primeros años para los *datos de control* y el resto para los *datos de predicción*. O bien dejar los últimos 6 o 3 meses para esta última actividad. Existe una gran variedad de combinaciones que siempre estarán sujetas a las necesidades y resultados esperados de cada organización.

Para los *datos de control*, es necesario seleccionar el modelo de pronóstico que mejor ajusta a los datos históricos, pero ¿Cuántas formas de medición existen? ¿Todas tienen la misma funcionalidad? ¿El que tiene menor error o el de mayor complejidad y variables? Se sabe que a medida que aumentan las variables en el modelo, el error puede disminuir. Con relación a los *datos de predicción*. ¿Cómo determinar aquél que pronosticó mejor? O ¿qué medida utilizar para este fin? En las siguientes páginas se expondrán alternativas para solucionar estas interrogantes.

2. ¿Cuántas medidas de precisión existen y cómo aplicar cada una de ellas?

Para cada uno de los escenarios que usted genere es necesario medir el desempeño de ellos mediante indicadores de precisión. Se requiere definir un criterio para la precisión del pronóstico (*datos de predicción*) y otra para la selección del modelo (*datos de control*). No todos tendrán el mismo significado ni el mismo uso, pero ambos están basados en la siguiente fórmula del error (e_t):

$$e_t = (Y_t - \hat{Y}_t)$$

Donde e_t es el error del pronóstico, Y_t el valor observado o real en el tiempo t de la serie de tiempo, y \hat{Y}_t es igual al valor pronosticado en el tiempo t de la serie de tiempo.

El error también puede ser representado en términos relativos y/o absolutos (%) utilizando la siguiente formulación:

$$ea_t(\%) = \frac{|Y_t - \hat{Y}_t|}{Y_t} * 100$$

O bien expresarlo en forma cuadrática:

$$e_t^2 = (Y_t - \hat{Y}_t)^2$$

En la Tabla 1 se muestran las medidas de error más utilizadas donde F_t es igual a \hat{Y}_t , notación utilizada para identificar el valor pronosticado.

Como se puede observar en dicha tabla existen más de 15 herramientas para medir el error del pronóstico, y su cantidad complica la selección de la medida de precisión más adecuada para el modelo de predicción. Por ejemplo, si seleccionamos aquellas de errores absolutos en lugar de los cuadráticos, éstos penalizan en mayor medida los errores grandes. La elección dependerá de la importancia que se les dé a los grandes errores. El coeficiente de desigualdad U de Theil, que no se muestra en la tabla y que se detalla más adelante, presenta una solución para estos escenarios. Si el valor de U es cercano a cero, supone una predicción perfecta, lo cual es muy difícil que suceda.

Medida de error	Fórmula
MSE Mean Square Error	$Media\{e_t^2\}$
RMSE Root Mean Square Error	\sqrt{MSE}
MAE Mean Absolute Error	$Media\{e_t\}$
MdAE Median Absolute Error	$Mediana\{e_t\}$
MAPE Mean Absolute Percentage Error	$Media\{p_t\}$
MdAPE Median Absolute Percentage Error	$Mediana\{p_t\}$
sMAPE Symmetric Mean Absolute Percentage Error	$Media\left\{2 \cdot \frac{ Y_t - F_t }{Y_t + F_t}\right\}$
sMdAPE Symmetric Median Absolute Percentage Error	$Mediana\left\{2 \cdot \frac{ Y_t - F_t }{Y_t + F_t}\right\}$
MRAE Mean Relative Absolute Error	$Media\{r_t\}$
MdRAE Median Relative Absolute Error	$Mediana\{r_t\}$
GMRAE Geometric Mean Relative Absolute Error	$MediaG\{r_t\}$
ReMAE Relative Mean Absolute Error	MAE / MAE^*
ReRMSE Relative Root Mean Squared Error	$RMSE / RMSE^*$
LMR Log Mean Squared Error Ratio	$\log(ReRMSE)$
PB Percentage Better	$100 \cdot Media\{I\{r_t < 1\}\}$
PB(MAE) Percentage Better (MAE)	$100 \cdot Media\{I\{MAE < MAE^*\}\}$
PB(MSE) Percentage Better (MSE)	$100 \cdot Media\{I\{MSE < MSE^*\}\}$

Tabla 1 – Medidas de error de pronóstico. Adaptado de De Gooijer y Hyndman (2005)

Una medida muy utilizada en la práctica de los pronosticadores es el MAPE, pero ésta presenta sesgos que favorece a los pronósticos que están por debajo de los valores reales. Para evitar ese problema, puede utilizarse el sMAPE. Pero éste indicador también tiene los propios, ya que tiene un comportamiento no deseado cuando el valor real o el pronóstico están muy cerca de cero. Otra de las medidas para seleccionar el mejor modelo de pronóstico es el MSE o la RMSE pero a pesar su efectividad probada, y aunque aparece en primer lugar en la Tabla 1, ésta no es muy utilizada entre los pronosticadores.

Con este breve antecedente y la estructura básica del error definida, he dividido las diferentes alternativas de medición del error en tres: a) Medidas de selección y b) Medidas de interpretación, y c) Medidas de precisión. En esta división no se incluyen todas las que aparecen en la mencionada tabla, pero si las más significativas que son producto de más de 15 años de experiencia en el tema y en más de 200 empresas. Como siempre el lector tendrá el último comentario con base en la información que dispone y los datos de la empresa en la que participa.

a) Medidas de selección.

En la primera sección de este artículo, se sugirió dividir la serie de tiempo de dos partes. Una parte la denominamos *datos de control* y otra *datos de predicción*. En esta sección haremos referencia a la primera, y con ella encontrar el mejor modelo de pronóstico que presente el mínimo error. Las herramientas que nos ayudan a cumplir este objetivo son:

MSE (Mean Square Error): Es el promedio de los cuadrados de las diferencias de cada artículo en el periodo t y se utiliza para comparar la precisión entre diferentes métodos o criterios de pronóstico. Según mi experiencia esta medida o la siguiente son las más recomendadas para seleccionar el mejor método de pronóstico.

$$MSE = \frac{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}{n}$$

RMSE (Root Mean Square Error): Esta medida es la raíz del promedio de los cuadrados del error de cada artículo en el periodo t y también se utiliza para comparar la precisión de diferentes métodos de pronóstico. La diferencia con la anterior es que el resultado está en las unidades originales de la información histórica.

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}{n}}$$

b) Medidas de Interpretación.

MPE (Mean Percentage Error): Es la media del error porcentual. Es una métrica simple, que sirve para ver si el error del pronóstico tiene un sesgo (*bias*) positivo o negativo. También se dice que el pronóstico está subestimado o sobrestimado.

$$MPE = \frac{\sum_{t=1}^n \left(\frac{Y_t - \hat{Y}_t}{Y_t} \right)}{n}$$

Ambas formas de error incrementan el costo y reducción en la utilidad, por lo que entre más cercano a cero es mejor. Un pronóstico sobre estimado puede generar: exceso de inventario, sobre costos, rechazos por obsolescencia, costos por mermas, entre otros. Un pronóstico subestimado puede producir: reducción en los niveles de servicio, costo por pérdida de ventas, incremento en el costo por el reprocesamiento de órdenes, entre otros.

MAPE (Mean Absolute Percentage Error): Es la media de los errores porcentuales en valor absoluto, no considera el signo del error sólo la magnitud. El MAPE es una de las medidas más utilizadas a nivel mundial, pero no se recomienda para la selección de un

método de pronóstico puesto que ésta presenta sesgos que favorece a los pronósticos que están por debajo de los valores reales.

$$MAPE = \frac{\sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|}{n}$$

WMAPE (Weighted Mean Absolute Percentage Error): Es el MAPE ponderado por el peso de las ventas, compras o servicios ofrecidos. Es un indicador muy recomendado ya que la ponderación del total minimiza los efectos de productos con grandes variaciones pero con poco impacto en los valores reales. Pero presenta los mismos problemas que el MAPE.

$$WMAPE = \frac{\sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| * Y_t}{\sum Y_t}$$

c) Medida de Precisión

En esta sección se hace referencia a la segunda porción de información denominada *datos de predicción* que se describe en el número 1 de este artículo. Recuerde que tenemos que probar que tan efectivo fue el modelo o los modelos seleccionados para pronosticar. En seguida se muestran varias alternativas para evaluar la efectividad del modelo seleccionado para pronosticar tales como el FA (forecast accuracy), el coeficiente *U* de Theil, el AIC y el BIC. La selección dependerá siempre de los resultados esperados en la organización y de la experiencia del planeador.

FA (forecast accuracy)

La forma más común de medir la precisión de un pronóstico (*forecast accuracy*) es comparar los resultados del pronóstico contra los valores reales del siguiente periodo. El objetivo es encontrar valores cercanos a 1 para emitir juicios favorables sobre el modelo de pronóstico seleccionado. La fórmula utilizada para este efecto es:

$$FA \text{ (forecast accuracy)} = 1 - e_t a(\%)$$

Aun cuando ésta es la una de las medidas más utilizada entre los pronosticadores, sólo se recomienda utilizarla en el corto plazo (no más de tres periodos) ya que existen otros criterios que nos pueden ayudar a encontrar cuál modelo pronostica mejor a periodos más largos. Recuerde que si el modelo seleccionado se ajustó muy bien a los datos históricos no necesariamente es igual de efectivo en el momento de pronosticar.

U de Theil

El coeficiente de desigualdad U de Theil es otra medida que permite analizar la efectividad del modelo seleccionado en la predicción. Recuerde que las medidas de errores absolutos en lugar de los cuadráticos, suelen presentar sesgos y éstos últimos penalizan en mayor medida los errores grandes. La elección dependerá de la importancia que se les dé a los grandes errores. El coeficiente de desigualdad U de Theil presenta una solución para estos escenarios. Si el valor de U es cercano a cero, supone una predicción perfecta. Su formulación está basada en la diferencia cuadrática que existe entre las tasas de crecimiento de la variable real y la estimada.

Este coeficiente se puede utilizar para evaluar la efectividad del pronóstico a mediano plazo

$$U_t = \sqrt{\frac{\sum_{t=1}^n e_t^2}{\sum_{t=1}^n (Y_t - Y_{t-1})^2}}$$

AIC (Akaike Information Criterion)

Esta herramienta penaliza la complejidad del modelo tomando en cuenta el número de variables y se utiliza para seleccionar el mejor modelo dentro del conjunto de los mismos datos. Los métodos de Box & Jenkins tiene esta característica, ya que utilizan valores reales y anteriores de la variables independiente para producir pronósticos precisos a corto plazo. La solución dada por Akaike es elegir como función de pérdida (o criterio de especificación) el mínimo del criterio de información.

$$AIC = \sqrt{\frac{\sum_{t=1}^n e_t^2}{n}} * \exp\left(\frac{2 * k}{n}\right)$$

BIC (Bayesian Information Criterion)

El criterio BIC de Schwarz penaliza con mayor intensidad modelos más complejos y con mayor número de variables, por lo que se dice que es más consistente. Por el contrario el criterio de Akaike es asintóticamente eficiente, ya que a medida que aumenta el número de variable su eficiencia para medir la efectividad de la precisión no mejora.

La eficiencia asintótica tiene que ver con la hipótesis de que la realidad es mucho más *complicada que cualquier modelo considerado* por lo que, al aumentar el número de observaciones también debiera aumentarse el número de modelos en el conjunto considerado. Poler et al (2007)

$$BIC = \sqrt{\frac{\sum_{t=1}^n e_t^2}{n}} * n^{\frac{k}{n}}$$

Tanto en el criterio AIC y el BIC, k son el número de variables del modelo, n el número de observaciones y e_t el error de pronóstico en el período t . Cuando se utilizan estos criterios, se selecciona el modelo que presenta el valor más pequeño.

Otros criterios de evaluación:

Si bien la precisión es una cualidad importante en la selección de un modelo de pronóstico, no es la única que a considerar en la selección y evaluación de un modelo de pronóstico. Las características de la información como: 1) El tamaño o el horizonte de tiempo de los datos históricos, 2) el comportamiento de la información y 3) el tipo de relación del pronóstico también deben considerarse en la elección.

El horizonte de tiempo limita la búsqueda del modelo. Algunos modelos se utilizan para calcular pronósticos a corto plazo (de uno a tres meses) como los modelos de atenuación exponencial. Los modelos de descomposición y ARIMA son útiles para el mediano plazo, mientras que los econométricos funcionan bien a mediano y largo plazo.

El comportamiento de la información también es útil para limitar la búsqueda del modelo. Por ejemplo si los datos históricos presentan estacionalidad no todos los modelos de pronóstico tienen esta funcionalidad. Winters o Box & Jenkins serían los más recomendables.

Por último existen tres tipos de relaciones que limitan también esta búsqueda: 1) Si la serie es única y el pronóstico se obtiene a partir de su propio pasado, 2) Si la serie es no causal y el pronóstico se obtiene a partir de otras bases históricas, y 3) Si la serie es causal en donde los pronósticos son obtenidos a partir de otras variables que están relacionadas causalmente.

Conclusiones y recomendaciones

Como se expuso en este artículo, si usted busca seleccionar el modelo que mejor ajusta a sus datos históricos, no use el MAPE, ni el WMAPE estos son indicadores de interpretación relativos y tienden a presentar sesgos. El MSE o la RMSE presentan mejores resultados para dicho efecto. Al seleccionar un modelo de pronósticos compare precisión contra pertinencia y no olvide el costo que esto representa. Recuerde que no siempre los modelos más complicados son más precisos que los más simples y por lo general menos costosos. Tal vez no posea un criterio único para juzgar la selección de un modelo en ciertas situaciones, en cualquier caso es necesario contar con varios de ellos y siempre dependerán de lo que se va a pronosticar y de la experiencia del pronosticador.

Dependiendo del uso que se le dé en la empresa a las predicciones a corto, a medio y a largo plazo, la importancia del horizonte de predicción debe ser tomada en cuenta a la hora de definir un criterio de selección de modelos. Es posible que sea más adecuado un modelo de pronóstico que tenga menor efectividad en el corto y mayor al mediando plazo y viceversa.

TG

Bibliografía de apoyo

Box, G.E., Jenkins, G.M., 1970. Time Series Analysis: Forecasting and Control, Holden-Day, San Francisco.

Croston, J.D., 1972. Forecasting and stock control for intermittent demands, Operational Research Quarterly, 23, pp. 289–303.

Fildes, R., & Makridakis, S., 1995. The impact of empirical accuracy studies on time series analysis and forecasting, International Statistical Review 63, 289-30.

Makridakis, S., Hibon, M., 2000. The M3-Competition: results, conclusions and implications International Journal of Forecasting, 16, pp. 451–476.

Makridakis, S., Wheelright, S.C., Hyndman, 1998. Forecasting: Methods and Applications. Wiley.

Poler, R., Mula, J., Peidro D., Tomás, J.V., 2006. PROGNOS: Un sistema de soporte al proceso de previsión de demanda. X Congreso de Ingeniería de Organización.