Combinación de pronósticos y variables predictoras con error

Elkin Castaño V.

Lecturas de Economía. No. 41.

Introducción

La primera referencia sobre la generación de un pronóstico como la combinación de otros se encuentra en Barnard (1963). Este autor publicó un artículo en el que comparaba los pronósticos basados en la metodología de Box-Jenkins (1976) con los obtenidos usando procedimientos clásicos como los métodos de suavización exponencial, y observó que, aunque la técnica propuesta por Box-Jenkins era mejor, un simple promedio de los dos pronósticos era superior. Este hallazgo promovió la investigación sobre la existencia de alguna otra combinación ponderada de los pronósticos que fuera aún mejor que dicho promedio. Los primeros en explorar esta posibilidad fueron Bates y Granger (1969) y encontraron bajo el supuesto de que cada predictor fuera insesgado, las ponderaciones óptimas para la combinación deberían sumar la unidad, produciendo así un pronóstico combinado insesgado. Lo más importante de los estudios de Granger y su grupo fue el señalar que se debe emplear toda la evidencia disponible en la construcción de un pronóstico. Esto implica que sería mejor usar varios modelos y combinarlos, y no seleccionar uno sólo.

A pesar de que el método propuesto era fácil de usar, su aceptación fue bastante lenta al principio. Entre otras cosas se argüía que si se necesitaba mezclar un modelo con otro para mejorar el pronóstico entonces el problema no había sido correctamente especificado. Afortunadamente, la creciente aceptación de la metodología Bayesiana de usar varios expertos y diferentes fuentes de evidencia para generar predicciones, reforzó la idea de usar múltiples modelos para construir pronósticos combinados. Morris (1974) señala que si los resultados de difentes modelos pueden ser tratados formalmente como pronósticos de diferentes expertos, entonces su combinación no debería generar ninguna controversia, pues equivaldría a combinar en la predicción distintos aspectos y fuentes de información.

Afortunadamente, los fuertes resultados estadísticos obtenidos, han permitido dejar atrás las críticas. Bajo el criterio de minimizar la varianza del error del predictor, la combinación de pronósticos individuales, ya sea con errores independientes o correlacionados, es óptima y nunca peor, en teoría, al pronóstico del mejor modelo individual. Además, los estudios empíricos generalmente reivindican la robustez en la práctica de la combinación.

I. Información y pronósticos

Para comprender más sobre las bases de la teoría de la combinación supongamos que tenemos N pronósticos, donde el j-ésimo individuo tiene el conjunto de información I_{jn}:(I_{on},J_{jn}) en el período n, donde I_{on} es la información disponible a todos y J_{jn} es la información disponible sólo al j-ésimo pronosticador. Asumimos que el contenido de J_{jn} es independiente de I_{on} y de J_{kn}, si $k \neq j$.

Concretamente, sea:

$$I_{on}: Z_{n-s}, \quad S \geq 0$$

Z_t puede ser un vector de series, pero aquí, por conveniencia, supondremos que es univariada.

$J_{jn}: X_{j,t-s}$, s | 0, y sea Y_t la serie que va a ser pronosticada. Denotemos el conjunto universal de información como:

$$U_n: \{ I_{on}, J_{jn}, j=1,2,...,N \}$$

el cual contiene la información disponible por todos los pronosticadores. El pronóstico óptimo -lineal- de mínimos cuadrados de Y_{n+1}, si U_n fuera conocido es:

$$\begin{aligned} E_u &= E\,[Y_{n+1}/U_n] \\ &= a(B)Z_n + \Sigma_j\, ß_j(B)X_{j,n} \qquad (1) \end{aligned}$$

donde a(B) y $ß_j(B)$ son polinomios del operador de rezagos B, definido como $Bw_t = w_{t-1}$. Bajo los supuestos hechos el pronóstico óptimo para la j-ésima persona es:

$$\begin{aligned} E_j &= E[Y_{n+1}/I_{jn}] \\ &= a(B)Z_n + ß_j(B)X_{j,n} \qquad (2) \end{aligned}$$

Si formamos un simple promedio de estos pronósticos individuales obtenemos:

$$\begin{aligned} \bar{E} &= \frac{1}{N}\Sigma_j E_j \\ &= a(B)Z_n + \frac{1}{N}\Sigma_j\, ß_j(B)X_{j,n} \qquad (3) \end{aligned}$$

El segundo término es la suma de N componentes independientes dividido por N lo que produce una desviación estándar de orden $N^{-1/2}$, y por tanto si N es grande

$$\bar{E} \cong a(B)z_n$$

Obsevemos que el lado derecho es el pronóstico usando sólo I_{on}. En general, si $\bar{E}$ es construído usando cualquier otro conjunto de ponderaciones que sumen uno, el resultado será similar. Por tanto, E_u no puede ser obtenido por la combinación de los pronósticos individuales óptimos. Sin embargo, si consideramos otro pronóstico usando la información I_{on}, obtenemos:

$$E_o = E\,[Y_{n+1}/I_{on}] = a(B)z_n \qquad (4)$$

y entonces el pronóstico óptimo E_u se puede optener como

$$E_u = \Sigma_j\, E_j \; - (N-1)E_o \qquad (5)$$

Este análisis simple ilustra varias observaciones generales:

(1) Agregar pronósticos no es lo mismo que agregar conjuntos de información. $\bar{E}$ está basada en toda la información, pero no es igual a E_u, pues la información no está siendo usada eficientemente.

(2) Ponderaciones iguales, como en la ecuación (3), son útiles cuando cada conjunto de información contiene componentes comunes y componentes independientes. Si la cantidad de información compartida varía a través de los pronósticadores, generalmente las ponderaciones usadas son diferentes.

(3) Un nuevo pronóstico puede mejorar la combinación aún si no está basado en un nuevo conjunto de información (p.e. E_o).

(4) Ponderaciones negativas pueden ser útiles, como en la ecuación (5).

(5) Es útil incluír muchos pronósticos, como en la ecuación (5).

En la próxima sección discutiremos 4 diferentes procedimientos para obtener las ponderaciones para el pronóstico combinado. Los tres primeros, se encuentran descritos en Granger y Ramanathan (1984) y el cuarto se encuentra en Holden y Peel (1989). La sección 3 presentan una aplicación y la sección 4 presenta algunas extensiones recientes a casos más generales.

II. Teoría de la combinación de pronósticos

Supongamos que una serie x_{t+1} va a ser pronosticada usando combinaciones de k pronósticos, $f_{j,t}$, para el período t=0,1,2,......,n-1. Usaremos la siguiente notación:

$x^T = (x_1, x_2, \ldots, x_n)$ es el vector de los valores de x que van a ser pronosticados.

$f_j^T = (f_{j0}, f_{j1}, \ldots, f_{j,n-1})$ es un vector de pronósticos usando el j-ésimo modelo.

$F = (f_1, f_2, \ldots, f_k)$ es una matriz de n∗k con los valores pronosticados por los diferentes modelos.

l = es un vector de unos de dimensión apropiada.

Método A

Sea $F\alpha = \alpha_1 f_1 + \alpha_1 f_1 + \ldots + \alpha_k f_k$ el pronóstico combinado sin restricciones, donde α es un vector de k∗1 con las ponderaciones de los f_j. El error del pronóstico es $e_A = x - F\alpha$. Suponga que el criterio para determinar las ponderaciones óptimas es elegir aquel α que minimice la suma de cuadrados de los errores de los pronósticos. Es decir, para obtener α debemos minimizar $(x-F\alpha)^T(x-F\alpha)$ con respecto a α. La solución es:

$$F^T(x-F\alpha) = 0, \text{ de donde, } \hat{\alpha} = (F^TF)^{-1}F^Tx$$

El pronóstico combinado es $\hat{x}= F\hat{\alpha} = F(F^TF)^{-1}F^Tx$. La suma de cuadrados de los errores de los pronósticos mínima es:

$$Q_A = (x-\hat{x})^T(x-\hat{x}) = x^Tx-x^TF\hat{\alpha}$$

Como se puede ver el procedimiento corresponde a regresar x contra $f_1, f_2, \ldots, f_k$ *sin término constante.* El valor de Q_A es simplemente la suma de cuadrados residuales disponible en todo programa de regresión.

Los pronósticos generados por este procedimiento serán, en general, sesgados, pues el promedio de los errores del pronóstico no es necesariamente cero. El error del pronóstico es $\hat{e}_A = x-F\hat{\alpha}$, y no hay razón para pensar que $l^T\hat{e}_A= l^T(x-F\hat{\alpha})$ sea cero. Aún si cada f_j tiene error medio de pronóstico igual a cero, la afirmación anterior sigue siendo válida. En efecto, supongamos $l^T(x-f_j)=0$ para cada j, es decir cada pronóstico individual es insesgado. En notación matricial, esto significa que $(l^Tx)l^T= l^TF$. Por tanto, $(l^Tx)l^T\alpha= l^TF\hat{\alpha}$. Para que el pronóstico combinado sea insesgado es necesario que $l^TF\alpha=i^Tx$ De lo anterior esto significa que $(l^Tx)l^T\alpha= l^Tx$. Si, como es muy probable, $l^T\hat{x}\neq 0$, -es decir x no tiene media cero-, entonces el pronóstico combinado será insesgado solamente cuando $l^T\hat{\alpha}= 1$, es decir cuando las ponderaciones sumen la unidad. Por tanto las condiciones suficientes para que el pronóstico combinado tenga error medio cero son:

(a) Cada pronóstico tiene error medio cero ($l^Tx = l^Tf_j$)

(b) Las ponderaciones deben sumar la unidad.

En la práctica, no hay razón por la cual dichas condiciones se cumplan. En particular, la condición (b) será violada a menos que sea impuesta *a priori*.

Método B

Consideremos ahora el caso en el cual las ponderaciones son restringidas a que sumen la unidad, de manera que debemos minimizar $(x\text{-}F\beta)^T(x\text{-}F\beta)$ con respecto a β, sujeto a $l^T\beta = 1$.

$$\min_\beta \{(x\text{-}F\beta)^T(x\text{-}F\beta) + 2\lambda_B(l^T\beta\text{-}1)\}$$

donde λ_B es el multiplicador lagrangiano. La condición de primer orden es

$$F^T(x\text{-}F\beta) - \lambda_B i = 0$$

$$\beta = (F^TF)^{-1}F^Tx - \lambda_B(F^TF)^{-1}l = \alpha - \lambda_B(F^TF)^{-1}l \qquad (6)$$

usando la ecuación (5). Además,

$$\lambda_B = (l^T\alpha\text{-}1)/[l^T(F^TF)^{-1}l] \qquad (7)$$

El valor mínimo de la suma de cuadrados de los residuales es:

$$Q_B = Q_A + \lambda^2_B[l^T(F^TF)^{-1}l] \qquad (8)$$

Evidentemente, $Q_B \geq Q_A$ y por tanto hay pérdida en el error cuadrático medio cuando se impone la restricción $l^T\beta = 1$.

Computacionalmente, el método B es equivalente a regresar $x\text{-}f_k$ contra $f_1 - f_k, f_2 - f_k, \ldots, f_{k-1} - f_k$. La ponderación de f_k es simplemente 1-(suma de las ponderciones de f_1, f_2, $f_3, \ldots, f_{k-1}$). Esta regresión también es sin constante. Q_B es la suma de cuadrados residuales dada por un programa de regresión estándar. La ganancia obtenida en el error cuadrático medio al eliminar la restricción de que las ponderaciones deben sumar la unidad puede entonces hallarse facilmente.

En la práctica, es frecuente el uso de la restricción de que las sumas de las ponderaciones debe ser la unidad. Lo anterior muestra que no existe razón para imponer tal restricción. Además, tampoco la hay para pensar que cada pronóstico alternativo sea insesgado.

El siguiente método muestra como obtener un pronóstico combinado con error medio cero a partir de pronósticos individuales con los cuales pueden ser sesgados.

Método C

El tercer método de combinación, *que requiere que la serie sea estacionaria*, no tiene restricciones pero agrega un término constante. En este caso debemos considerar la minimización de $(x-\delta_o l-F\delta)^T(x-\delta_o l-\delta)$ sin restricciones. En este caso δ_o es el término constante y δ es el vector de ponderaciones de los k pronósticos. Las ecuaciones normales son:

$$F^T(x-\delta_o l-F\delta) = 0 \ \text{ y } \ l^T(x-\delta_o l-F\delta) = 0$$

y sus soluciones son:

$$\hat{\delta} = \hat{\alpha} - \hat{\delta}_o(F^TF)^{-1}F^Tl$$

$$\hat{\delta}_o = (l^Tx-l^TF\hat{\delta})/n = l^T e_A/(n-\theta)$$

donde $\hat{e}_A$ es el vector de errores usando el método A y

$\theta = l^TF(F^TF)^{-1}F^Tl$. El pronóstico combinado obtenido es

$$\hat{x}_c = \hat{\delta}_o i + F\hat{\delta}.$$

El error cuadrático medio de este pronóstico es:

$$l^T(x-\hat{x}_c)/n = (l^Tx)/n - (l^TF\hat{\delta})/n - \hat{\delta}_o$$

Observemos que

$$\hat{e}_c = x-\hat{x}_c = x-\hat{\delta}_o l-F\hat{\delta} = \hat{e}_A - \hat{\delta}_o[I-F(F^TF)^{-1}F^T]l$$

y entonces el mínimo alcanzado por la suma de cuadrados de los errores es:

$$Q_c = Q_A - 2\hat{\delta}_o l^T[I-F(F^TF)^{-1}F^T]\hat{e}_A + \hat{\delta}^2_o l^T[I-F(F^TF)^{-1}F^T]l$$

Como $F^T\hat{e}_A = 0$, podemos simplificarla como:

$$Q_c = Q_A - (l^T\hat{e}_A)^2/(n-\theta) \leq Q_A$$

donde $\theta = l^TF(F^TF)^{-1}F^Tl$ y $\hat{e}_A$ es el vector de errores usando el método A. Comparando los tres métodos, el método C es evidentemente el mejor debido a que posee el menor error cuadrático medio y el pronóstico combinado producido es insesgado *aunque los pronósticos individuales sean sesgados*. "Por tanto la práctica común de obtener un promedio ponderado de los pronósticos individuales debería ser abandonada en favor de una combinación lineal no restringida incluyendo un término constante" (Granger y Ramanathan, 1984).

Nota: un test convencional para contrastar insesgamiento *ex-post* de una serie particular de pronósticos se basa en estimar la ecuación (Holden y Peel, 1985).

$$x_t = a + bf_{jt} + \mu_t \qquad (9)$$

donde f_{jt} es el t-ésimo pronóstico empleando el j-ésimo modelo, y chequear si a=0, b=1 y si μ es ruido blanco. Si cualquiera de estas condiciones no se cumple el pronóstico no será insesgado -un pronóstico combinado, posiblemente superior, puede ser obtenido por medio del método C y el pronóstico sesgado debería ser empleado en la combinación-.

Sin embargo, Holden y Peel (1989) han señalado que el test anterior puede producir resultados erróneos. La razón es la siguiente: la ecuación (9) puede ser escrita como

$$x_t = a\bar{x}/(\bar{x}) + bf_{jt} + \mu_t$$

donde $\bar{x}$ es la media no condicional de la serie. De aquí se ve claro que la estimación de la ecuación (9) combina dos pronósticos: el pronóstico de interés f_j y la media no condicional de la serie, $\bar{x}$, el cual es, en general, un pronóstico pobre pero insesgado de x_t. Dichos autores proponen un test alternativo para estudiar las propiedades del pronóstico, el cual se basa en ajustar la ecuación:

$$x_t - f_{jt} = a + \mu_t \tag{10}$$

o,

$$x_t - f_{jt} = a' + \lambda x_{t-1} + v_t$$

donde x_{t-1} es el conjunto de información rezagada y μ_t y v_t son los errores del proceso.

Si a es cero, el pronóstico será insesgado. Si μ_t no está serialmente correlacionado -dado el horizonte de pronóstico-, el pronóstico es debilmente eficiente. Si el pronóstico es así y λ y a' son cero entonces el pronóstico es fuertemente eficiente.

Es claro que la ecuación (10) es apropiada para contrastar si un error de pronóstico tiene una media constante diferente de cero -lo que se conoce como "sesgo de localización" en el pronóstico-. Es menos claro lo que sucede cuando el pronóstico está sujeto a un "sesgo de escala", el cual ocurre cuando el pronóstico, en promedio, sub o sobre estima el valor en algún porcentaje constante. Una discusión sobre este tema se encuentra en Holden y Peel (1989).

Método D

En su articulo, Holden y Peel (1989) discuten el método C sugerido por Granger y Ramanathan.

Suponga que los pronósticos candidatos para una combinación son insesgados y debilmente eficientes. Entonces una regresión no restringida, con constantes, de los datos de la serie x sobre los pronósticos puede producir que la constante sea estadisticamente significativa. Si el investigador usa esta combinación para obtener pronósticos *ex-ante* fuera de la muestra, implicitamente está también proyectando, con ponderación $\alpha/\bar{x}$, la media condicional de la serie. Si esto es o no razonable dependerá de la naturaleza de la variable que está siendo pronosticada y de las características de los pronósticos que son empleados. Para series económicas en las cuales se espera que la media incondicional cambie -por ejemplo, la inflación-, algunos pronósticos en la combinación estarán basados en modelos macroeconómicos y se espera que ellos incorporen tal cambio. En estos casos la media incondicional no debería ser incluída.

Ahora bien, si los pronósticos no son insesgados dichos autores sugieren introducir una constante en la regresión y corregir las ponderaciones de los pronósticos a que sumen la unidad. Esto corregirá los sesgos sin incluír implicitamente la media incondicional.

Para demostrar esto, supongamos, por simplicidad, que se quieren combinar dos pronósticos f_{1t} y f_{2t} los cuales siguen la forma general:

$$x_t - f_{1t} = \beta_1 + \mu_t$$

$$x_t - f_{2t} = \beta_2 + v_t$$

donde β_1 y β_2 son constantes -los sesgos de f_{1t} y f_{2t}, respectivamente- y μ_t, v_t son los errores de los procesos.

Ahora, consideremos la combinación

$$x_d = \lambda_o + \lambda_1 f_1 + \lambda_2 f_2 \quad (13)$$

donde las ponderaciones λ son determinadas de forma tal que minimicen el error cuadrático medio de los errores

$$E(x-x_d)^2 = E[x(1-\lambda_1-\lambda_2) - \lambda_o + \lambda_1 \beta_1 + \lambda_2 \beta_2 + \lambda_1 \mu_t + \lambda_2 v_t]^2$$

Bajo las condiciones dadas, las condiciones para un mínimo son:

$$1 = \lambda_1 + \lambda_2$$

$$\lambda_o = \lambda_1 \beta_1 + \lambda_2 \beta_2 \quad (14)$$

$$\lambda = var(\mu)-cov(\mu,v)/[var(\mu)+var(v)-2cov(v,\mu)]$$

Obdervando la ecuación (14), el papel de la constante λ_o en la ecuación (13) es el de corregir los pronósticos f_1 y f_2 de sus sesgos (β_1 y β_2). Además, existe la restricción de que las ponderaciones sumen la unidad. Por tanto para mejorar la eficiencia, se debería correr la regresión sobre los dos pronósticos en la forma restringida:

$$x_t - f_{1t} = \lambda_o + (1 - \lambda_1)(f_{2t} - f_{1t}) + w_t$$

donde λ_o corrige los pronósticos de sesgos, el coeficiente de $f_2 - f_1$ proporciona las ponderaciones y w es el término de error. Obviamente, el procedimiento puede generalizarse a k pronósticos.

En resumen: en la regresión sin restricción sugerida por el método C, la inclusión de una constante corregirá los sesgos de los pronósticos usados en la combinación pero implicitamente combinará esos pronósticos con la media no condicional de la serie, el cual es un pronóstico, en general muy pobre. Si las ponderaciones de los pronósticos son restringidas de forma que sumen la unidad y se incluye una constante en la regresión entonces dicha constante corregirá los sesgos en el pronóstico, pero la media no condicional no estará implicitamente incluída como un pronóstico adicional. Si no se incluye una constante en la regresión y hay algún pronóstico sesgado entonces el pronóstico combinado será sesgado.

III. Extensiones

Una extensión del procedimiento de combinación cuando la serie no es estacionaria se encuentra en Hallman y Kamstra (1989).

Como vimos, el método C exige que la serie que se va a pronosticar sea estacionaria, puesto que si no lo es, x_t no tiene media. Ahora bien, frecuentemente es posible estacionarizar una serie a través de diferencias. Cuando una serie x_t debe ser diferenciada d veces antes que aparezca ser estacionaria se dice que es integrada de orden d, y se denota por $xt \sim I(d)$. Un vector $y_t = (x_{1t}, x_{2t},....,x_{nt})$ es I(d) si cada serie componente lo es. En general las combinaciones lineales de series I(d) serán I(d), pero es posible que exista una o más combinaciones lineales que sean integradas de orden más bajo que d. Cuando esto sucede, se dice que y_t está integrada y el vector α de los coeficientes de una de tales combinaciones es llamado vector cointegrante (Engle et al.., 1987).

Suponga que x_t es integrada de orden 1 de manera que los cambios de la serie son estacionarios. Kamstra y Hallman (1989) discuten cómo obtener un pronóstico combinado para x_t. Si $x_t \sim I(1)$, cualquier pronóstico razonable f_j debería ser cointegrado con ella, a través del vector cointegrante (1,-1). Si esto no ocurre, el error del pronóstico x_t-f_j no será estacionario y la serie y su pronóstico se alejarán cada vez más en el tiempo. Sea $f^i_{t,h}$ el i-ésimo pronóstico de x_{t+h} realizado en el período t, i=1, 2, 3,....,k. Si x_t-$f^i_{t-h,h}$ y x_{t-h}-x_t son I(0), entonces también lo es

$$Z_{i,t-h} = f^i_{t-h,h} - x_{t-h}.$$

Si h=1, el pronóstico combinado para la serie estacionaria y_t, surge de la relación:

$$\Delta y_t = m + \beta_1 Z_{1,t-1} + \beta_2 Z_{2,t-1} + ...+ \beta_r Z_{r,t-1} + \varepsilon_t \qquad (15)$$

o,

$$y_t = m + y_{t-1} + \beta_1(f^1_{t-1,1} - y_{t-1}) + \beta_2(f^2_{t-1,1} - y_{t-1}) + ...+ \beta_r(f^r_{t-1,1} - y_{t-1}) + \varepsilon_t \qquad (16)$$

Puesto que $f^{i}_{t-1,1}$ puede interpretarse como el pronóstico del modelo del cambio en y_t hecho por el i-ésimo modelo, la ecuación (15) puede ser considerada la explicación del cambio en y_t como una combinación lineal de los pronósticos del nivel. Si reescribimos la ecuación (16) como

$$y_t = m + (1 - \Sigma_i\, \beta_i) y_{t-1} + \Sigma_i \beta_i f^{i}_{t-1,1} + \varepsilon_t \tag{17}$$

observemos varias diferencias:

primero, cuando se combinan pronósticos de una serie integrada, se debe incluír un rezago de la serie pronosticada. Segundo, los coeficientes de la regresión en (15), exeptuando m, deben ser restringidos a que sumen uno. Tercero, si los coeficientes β_i suman uno la variable dependiente rezagada en la ecuación (17) desaparece. Cuarto, puesto que todas las variables en la ecuación (13) son I(0), los estadísticos t de los regresores, incluyendo la constante m, deben ser empleados para decidir si todos los pronósticos deben ser retenidos en la combinación.

Diebold (1985) muestra que si los pronósticos individuales se encuentran correlacionados, una combinación dinámica de ellos puede generar pronósticos mejorados.

Guerard y Clemen (1989) señalan que frecuentemente los pronósticos individuales poseen un alto grado de dependencia y las ponderaciones estimadas son inestables. Para obtener estimaciones más eficientes recomiendan usar el método de la regresión de raíz latente.

Un uso más intensivo de los datos se obtiene cuando las ponderaciones de las combinaciones varían con el tiempo. En este caso es posible hacer una regresión de parámetros cambiantes usando el filtro de Kalman.

IV. Una aplicación

Frecuentemente, al usar modelos econométricos para pronosticar los valores futuros de una variable, los valores usados para las variables predictoras contienen errores. Estos errores pueden generarse, por ejemplo, por la elaboración de escenarios imprecisos, predicción de dichos valores usando modelos de series de tiempo, etc. Estamos interesados en estudiar, el comportamiento a corto, a mediano y a largo plazo, de las predicciones generadas cuando se mezclan los pronósticos generados por dicho modelo y un modelo basado exclusivamente en el pasado de la variable dependiente -no hace uso de las predictoras-, como un modelo ARIMA, frente al comportamiento de los pronósticos individuales de cada modelo. A través de un ejercicio de simulación podemos visualizar esta situación.

Considere el modelo

$$y_t = 3 + .3y_{t-1} + .4x_{1t} - .3x_{2t} + \varepsilon_t \tag{18}$$

donde e ~ N(0,0.5), y

$$x_{1t} = 2 + .4_t + \mu_{1t}, \tag{19}$$

$$x_{2t} = 1 + .15_t + \mu_{2t}, \tag{20}$$

con $\mu_{1t} \sim N(0,0.4)$ y $\mu_{2t} \sim N(0,0.2)$.

Para este modelo se simularon 300 observaciones de las cuales 180 se emplearon para estimar el modelo y las restantes 120 para hacer las predicciones de y_t en las siguientes situaciones:

(i) Cuando las predictoras no contienen errores, es decir usando los últimos 120 valores generados por (19) y (20).

(ii) Cuando las predictoras poseen pequeños errores; se contamina a (19) con una uniforme en el intervalo (-1,1) y a (20) con otra uniforme en el intervalo (-.5,.5).

(iii) Cuando las predictoras tienen errores moderados: se contamina a (19) con una uniforme en el intervalo (-2,2) y a (20) con otra uniforme en el intervalo (-1,1).

(iv) Cuando las predictoras poseen errores grandes: se contamina a (19) con una uniforme en el intervalo (-5,5) y (20) con otra uniforme en el intervalo (-3.5,3.5).

Cada una de las situaciones anteriores se simuló 1100 veces y para cada simulación se calculó el error cuadrático del pronóstico del modelo econométrico. Además, para cada serie simulada de y_t se ajustó un modelo ARIMA y se calculó también el error cuadrático de los pronósticos generados por este modelo.

Además se diferenció el horizonte del pronóstico en la siguiente forma: se consideró como pronóstico a corto plazo los primeros doce períodos; a mediano plazo los siguientes 48 períodos y a largo plazo los últimos 60 períodos.

Puesto que la naturaleza del proceso que genera a y_t es no estacionario el procedimiento empleado para la combinación de los pronósticos es aquel sugerido por Hallman y Kamstra (1989) en la sección anterior. Los cuadros 1, 2 y 3 resumen los resultados obtenidos. Para resumir el comportamiento de los diferentes pronósticos se emplearon la Mediana (*Me*), el cuartil inferior (Q_1) y el superior (Q_3) y la moda (M_o) de los errores cuadráticos.

Tabla 1
Pronóstico a corto plazo

	sin error 1	error 1 2	error 2 3	error 3 4
Econom: Me	.269	.330	.529	2.412
Q1,Q3	.196 .362	.244 .442	.393 .706	1.877 3.161
Mo	.268	.330	.529	2.410
ARIMA: Me	.428	.424	.427	.428
Q1,Q3	.270 .673	.277 .668	.281 .671	.277 .616
Mo	.427	.424	.426	.427
COMBIN: Me	.250	.288	.398	1.537
Q1,Q3	.185 .327	.212 .372	.297 .535	1.089 1.980
Mo	250	.288	.398	1.536

Tabla 2
Pronóstico a mediano plazo

	sin error 1	error 1 2	error 2 3	error 3 4
Econom: Me	.288	.380	.648	3.302
Q1,Q3	.241 .336	.318 .447	.541 .773	2.572 4.124
Mo	.288	.380	.647	3.301
ARIMA: Me	.554	.572	.544	.551
Q1,Q3	.367 1.071	.359 1.074	.364 1.079	.369 1.090
Mo	.553	.569	.542	.550
COMBIN: Me	.262	.315	.464	1.980
Q1,Q3	.223 .300	.269 .369	.391 .561	1.520 2.606
Mo	.262	.268	.464	1.979

Tabla 3
Pronóstico a largo plazo

	sin error 1	error 1 2	error 2 3	error 3 4
Econom: Me	.296	.446	.940	5.867
Q1,Q3	.253 .347	.380 .529	.768 1.162	4.299 7.808
Mo	.283	.446	.940	5.847
ARIMA: Me	.978	1.014	.909	.925
Q1,Q3	.454 2.303	.452 2.224	.428 2.488	.457 2.290
Mo	.972	1.015	.900	.925
COMBIN: Me	.268	.356	.648	3.483
Q1,Q3	.233 .309	.309 .407	.524 .791	2.455 4.684
Mo	.268	.356	.645	3.477

De los resulatados anteriores se puede concluir:

1) Para el corto plazo (Tabla 1):

a) Las columnas 1, 2 y 3 -predictoras sin errores o con errores pequeños y moderados- muestran que el pronóstico combinado tiene claras ventajas de precisión sobre los pronósticos individuales. Sin embargo cuando los errores en las predictoras son grandes -columna 4- los resultados sugieren el empleo del modelo ARIMA para el corto plazo. Se observa también una ganancia en precisión del pronóstico combinado frente al econométrico, a medida que el error crece.

b) Cuando los errores en las variables predictoras son grandes la columna 4 sugiere que es preferible usar un modelo basado en el pasado de la variable para predecir su futuro. También se observa que el pronóstico combinado supera al econométrico, aunque es inferior el generado por el modelo ARIMA.

2) Para el mediano plazo (Tabla 2).

Se tienen resultados semejantes a los de corto plazo: las predictoras no contienen errores o cuando poseen errores pequeños y/o moderados, el pronóstico combinado es superior a los pronósticos individuales y para grandes errores de nuevo es preferible usar el modelo ARIMA. Sin embargo, la precisión del combinado es menor que en el caso del corto plazo.

3) Para el largo plazo (Tabla 3).

Las conclusiones son análogas a las obtenidas para el corto y mediano plazo, aunque surge el problema adicional que a pesar de que es mejor el pronóstico ARIMA, éste no es recomendable para el largo plazo.

Los resultados muestran también que hay poca pérdida en la precisión del pronóstico econométrico y del combinado al pasar del corto, mediano y largo plazo, cuando conocemos exactamente los valores futuros de las predictoras, perdiendo menos precisión el pronóstico combinado; no ocurre lo mismo cuando los valores futuros de las predictoras contienen errores, pues la pérdida de precisión tanto en el pronóstico econométrico como el combinado es mayor y creciente con la magnitud de los errores, siendo más dramático en el pronóstico del modelo estructural

Resumiendo, se concluye que con excepción al caso en el cual los valores futuros de las variables predictoras tienen grandes errores, el procedimiento de la mezcla de pronósticos generó predicciones más precisas en el corto, mediano y largo plazo.

Conclusiones

El desarrollo de los métodos de combinación de pronósticos permite obtener pronósticos que, en general, son superiores a los pronósticos individuales. Las técnicas de combinación cubren cada vez una gama más ámplia de situaciones -ponderaciones cambiantes, cambios estructurales en los modelos [Diebold et al, 1987], horizonte del pronóstico mayor de un período [Engle et al, 1987]-, permitiendo obtener reglas de combinación más precisas en cada situación.

Finalmente, como señala Granger (1989) en sus conclusiones:

> Ocasionalmente se ha sugerido que la combinación no produce pronósticos mejores. Si esto es cierto, tanto los econometristas como los analistas de series de tiempo deberían sentirse muy preocupados, puesto que esto significa que métodos muy simples para pronosticar, basados en pequeños conjuntos de información ineficientemente usada, son difíciles de mejorar.

Referencias

Barnard, G.A., (1963) *"New Methods of Quality Control"*, Journal of the Royal Statistical ociety A, Vol. 126, 255-259.

Bates, J.M., and Granger, C.W.J., (1969) *"The Combination of Forecasts"*, Operational Research Quaterly, Vol. 20, 451-468.

Box, G.E.P., and Jenkins, G.M., (1976) *"Time Series Analysis:Forecasting and Control"*, revised edition, San Francisco, Holden-Day.

Diebold, F.X., (1988) *"Serial Correlation and the Combination of Forecasts"*, Journal of Bussiness and Economics Statistics, Vol. 6, 105-111.

Diebold, F.X., and Pauly, P., (1987) "*Structural Change and the Combination of Forecasts*", Journal of Forecasting, Vol. 6, 21-40.

Engle, R.F., and Granger, C.W.J., (1987) "*Cointegration and Error Correction: Representation, Estimation and Testing*", Econometrica, Vol. 55, 251-276.

Engle, R.F., Granger, C.W.J., and Hallman, R.J., (1989) "*Merging Short-and Long Run Forecasts: and Application of seasonal Co-integration to Monthly Electricity Sales Forecasting*", Journal Of Econometrics, Vol. 40, 45-62.

Granger, C.W.J., (1989) "*Combining Forecasts-Twenty Years Later*", Journal of Forecasting, Vol. 8, 167-173.

Granger, C.W.J., and Ramanathan, R., (1984) "*Improved Methods for Combining Forecasts*", Journal of Forecasting, Vol. 3, 197-204.

Guerard, J.B., and Clemen, R.T., (1989) "*Collinearity and the use of Latent Rot Regresion for Combining GNP Forecasts*", Journal of Forecasting, Vol. 8, 231-238.

Holden, K., and Peel, D.A., (1989) "*Unbiasedness, Efficiency and the Combination of Economic Forecasts*", Journal of Forecasting, Vol. 8, 175-188.

Hallman, J. and Kamstra, M., (1989) "*Combining Algorimthms Based on Robust Estimation Techniques and Co-Integrating Restrictions*", Journal of Forecasting, Vol. 8, 189-198.

Morris, P.A., (1974) "*Decision Analysis Expert Use*", Management Science, Vol. 20, 1233-1241.