

**Running head: PROPAGACIÓN DE LOS ERRORES DE PROYECCIÓN DE LAS
SERIES DE TIEMPO AJUSTADAS CON MODELOS DE ESPACIO DE ESTADO**

Errores de proyección

Adriana Fátima Panico de Bruguera

apanico@herrera.unt.edu.ar

María Angélica Pérez de Del Negro

mperez@herrera.unt.edu.ar

Facultad de Ciencias Económicas

Universidad Nacional de Tucumán (Argentina)

RESUMEN

Uno de los problemas más importantes en el análisis de las series de tiempo es la *proyección* de valores futuros. Pero en todo momento se debe tener en cuenta que no se puede predecir el futuro, solamente se puede analizar el pasado, esperar que el modelo persista y en base a ello realizar las estimaciones de valores futuros mediante una proyección hacia delante del modelo generador de las observaciones. Cuando la serie representa alguna variable económica, la importancia práctica de la proyección es evidente.

El objetivo en el presente trabajo es estudiar los errores que se producen en el proceso de estimación y predicción, explicando la forma en que ellos se propagan, mediante una expresión matemática simple, y cómo influyen cuando una serie es la acumulada de otras.

Para el análisis y modelado estadístico de las series de tiempo tratadas, se aplica el enfoque de espacio de estado, el cual, como se demostró en diversos trabajos publicados, resulta muy satisfactorio a la hora de hacer proyecciones y además muy útil en la aplicación de diversos problemas reales.

Palabras Claves: Errores; Estimación y proyección; Modelos de Espacio de Estado.

1. Introducción

En un proceso estocástico $\{y_t\}$, donde su media, $E(y_t) = 0$ y su función de covarianza $E(y_t y_s) = \sigma(t-s)$, es importante considerar las proyecciones futuras de y_t en términos de su pasado. Si \hat{y}_t es la proyección, el criterio es hacer pequeña a $E(y_t - \hat{y}_t)^2$, esto significa que el error cuadrático medio de predicción se minimiza si \hat{y}_t se toma como la esperanza condicional de y_t dadas las observaciones pasadas, $y_{t-1}, y_{t-2}, \dots, y_{t-n}$. Si la distribución es normal o el proceso es Gaussiano, la esperanza condicional será una función lineal, bajo otras circunstancias podría no serlo.

Suponiendo que el conjunto de observaciones y_1, \dots, y_n satisface el modelo de espacio de estado (2), se desea predecir y_{n+r} , $r = 1, 2, \dots, j$ ¹.

$$\begin{aligned} y_t &= Z_t \alpha_t + \varepsilon_t, & \varepsilon_t &\sim N(0, H_t), \\ \alpha_t &= T_t \alpha_{t-1} + R_t \eta_t, & \eta_t &\sim N(0, Q_t), \end{aligned} \quad t = 1, \dots, n \quad (1)$$

Siguiendo con el argumento anterior, la predicción debe ser con *mínimo error medio cuadrático de predicción* dada la muestra completa Y_n . Es decir que se busca aquella estimación $\hat{y}_n(r)$ de y_{n+r} que minimice

$$\bar{F}_{n+r} = E \left[(\hat{y}_n(r) - y_{n+r})(\hat{y}_n(r) - y_{n+r})' \mid Y_n \right] \quad (2)$$

para todas las estimaciones de y_{n+r} . La predicción con mínimo error cuadrático medio de y_{n+r} dada la muestra completa Y_n es la media condicional

$$\hat{y}_n(r) = E(y_{n+r} \mid Y_n) \quad (3)$$

El propósito de esta investigación es analizar qué relación existe entre los errores de las proyecciones de cada una de las series de los impuestos que se recaudan en la Provincia de Tucumán, y el error de proyección de la serie del total acumulado de esos impuestos.

¹ HARVEY, A. C., (1989)

2. Datos

Los datos analizados corresponden a los valores totales mensuales de la recaudación de impuestos provinciales de la Provincia de Tucumán² y cada uno de los gravámenes que componen el total. Las series de tiempo así conformadas comprende el período: enero de 1996 a enero de 2006. Los datos en cuestión fueron suministrados por la Dirección General de Rentas de la Provincia de Tucumán en pesos corrientes, es decir que están expresados en pesos correspondientes al año en que fueron recaudados. Para poder realizar el análisis los mismos debieron ser transformados a pesos constantes, o sea, expresados en pesos correspondientes a un mes o a un año base determinado para eliminar la inflación durante el período. En la elección del período base deben tenerse en cuenta diversos factores que permitan considerar ese período como *normal* o sin grandes cambios o sobresaltos en la actividad económica. Se consideró más representativo de la realidad provincial deflactar las series, a precios de diciembre de 2001, con el IPC (Índice de Precios al Consumidor, Nivel General, de Tucumán), calculado mensualmente por la Dirección Provincial de Estadística, base 1984=100.

Los impuestos incluidos en la serie corresponden a los denominados Impuestos Provinciales conocidos como: *Ingresos Brutos, Inmobiliario, Automotor; Salud Pública, Sellos, Autotransporte, Sellos Judiciales, Riego, Boletín, Cobros Judiciales, Pecuarios, Multas, Eventuales Generales, Juegos y Tasas Policiales* (los últimos 10, por su poca significación individual en los valores totales, aparecen generalmente agrupados en un solo ítem como *Otros*).

² No se consideran los ingresos tributarios provenientes de fondos coparticipables.

3. Forma matemática de la propagación del error

Como se mencionó en la sección anterior, cada uno de los impuestos que se recauda en la provincia, tiene un porcentaje de participación en el total, diferente. Dichos porcentajes se calcularon mensualmente desde enero a julio de 2004, debido a que en ese período, en particular, se adoptaron diversas medidas para aumentar la recaudación. La figura 1 representa la participación porcentual de cada impuesto en el total recaudado por mes para el primer semestre de 2004.

Se observa que el impuesto a los Ingresos Brutos es el de mayor porcentaje de participación (entre 48% y 69% del total), con marcada diferencia con respecto a los otros. Similar comportamiento se observa en la figura 2 para el segundo semestre de 2005, donde los porcentajes de participación oscilan entre 52% y 75% del total.

De acuerdo a esta situación, y para el cálculo del error, se propone ponderar el error de proyección correspondiente a cada una de las series que componen el total, de acuerdo a su correspondiente porcentaje de participación.

Para ello se definen los siguientes parámetros: p_i , con $i = 1, \dots, n$; la participación con respecto al total, n el número de series que componen el total; e_i el error de proyección de cada una de éstas; p_T la relación de la suma de las proyecciones de cada uno de las series que componen el total sobre la proyección del total y e_T su correspondiente error de proyección.

La relación matemática propuesta para observar la propagación de los errores obedece a la siguiente expresión:

$$\sum_{i=1}^n p_i^2 e_i^2 \cong p_T^2 e_T^2 \quad (4)$$

El proceso de estimación y proyección se realizó considerando las series con los valores observados hasta el mes de septiembre de 2005 inclusive, y con el modelo ajustado se proyectaron tres meses en adelante (octubre, noviembre y diciembre), con sus respectivos

errores. El mismo procedimiento se realizó con la serie observada hasta octubre inclusive y luego incluyendo el mes de noviembre. Se ajustaron diferentes modelos para proyectar noviembre y diciembre de 2005, enero y febrero de 2006, respectivamente. En cada caso, se fueron comparando los valores observados con los proyectados para corroborar la bondad del método estadístico empleado.

Por otro lado, también se hizo lo propio con la serie de la recaudación total y se consideró cuánto difería ésta de la suma de las otras, desde el punto de vista de las proyecciones.

4. Resultados de las proyecciones y sus correspondientes errores.

Las tablas 1,2 y 3 muestran los resultados de aplicar el procedimiento expuesto en la sección anterior, para las series procesadas, ajustadas y proyectadas en pesos corrientes.

De acuerdo a los valores que resultaron de los cálculos, se puede observar que la propagación de los errores correspondiente al total obtenido de la suma y al total proyectado por sí mismo, difieren a partir de la tercera cifra decimal (el error es del orden 10^{-2}) y a medida que se agrega información a la serie el modelo ajusta mejor en cuanto a las proyecciones. Además, en la primera proyección de cada caso la diferencia es menor que en las subsiguientes, lo cual se debe fundamentales, a que a medida que se alarga el plazo de proyección crece su varianza (error cuadrático medio).

Este mismo análisis se realizó ajustando y proyectando las series en pesos constantes, actualizados con (el IPC³ de Tucumán). Los valores proyectados luego fueron transformados a valores corrientes aplicando el procedimiento en forma inversa, sin que este proceso afecte los errores ni los porcentajes de participación.

³ IPC: Índices de Precios al Consumidor (Nivel General, diciembre de 2001)

5. Conclusiones.

El tratamiento estadístico de análisis de series cronológicas aplicado a los datos considera un amplio rango de irregularidades en las mismas y demuestra que todos los problemas estadísticos asociados pueden ser manejados mediante el uso del enfoque de espacio de estado. Las series de tiempo estructurales tienen intuitivamente una representación de espacio de estado y el filtro y el suavizador producen valores que tienen interpretación práctica, aun cuando los datos presentan diversas irregularidades.

El hecho que la tendencia cambie en el tiempo, permite capturar los cambios estructurales que se produjeron en la economía durante el período de observación. Las otras variables, como la estacionalidad y las intervenciones, se incluyen a fin de dar mayor generalidad al modelo de modo que en él intervengan las estructuras usuales de los modelos econométricos observados en el tiempo.

En el estudio de los errores, específicamente tratado en este trabajo, es conveniente destacar la confiabilidad del métodos para la obtención de los errores propios de la proyección y su importancia en una combinación lineal.

Observando los resultados, se concluye que la influencia de los errores de proyección de cada una de las series que componen el total, a través de la combinación lineal planteada, no difiere mucho del error obtenido al proyectar la serie total. Además, a medida que se va agregando información a la serie las diferencias se achican.

La expresión utilizada para la propagación del error es una aplicación de la propiedad que tiene la varianza de una combinación lineal de variables.

En todos los casos la diferencia absoluta al calcular ambos miembros de la expresión (4) es menor que 10^{-2} .

Bibliografía

- Ø ABRIL, J. C. (1997). *Series de tiempo irregulares: un enfoque unificado*. Conferencia invitada pronunciada durante el XXV Coloquio Argentino de Estadística. Sociedad Argentina de Estadística. Noviembre 1997.
- Ø ABRIL, J.C.: *Modelos para el Análisis de las Series de Tiempo*. Ediciones Cooperativas, Buenos Aires (2004)
- Ø BROCKWELL, P. J., DAVIS, R. A.: “*Times Series: Theory and Methods*”. New York, (1987)
- Ø HARVEY, A. C.: *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press. (1989)
- Ø KOOPMAN, S. J., HARVEY, A. C., DOORNIK, J. A. SHEPHARD, N.: *STAMP 5.0: Structural Time Series Analyser, Modeller and Predictor*. Chapman And Hall, (1995).
- Ø LEGUIZAMON, M. C. y PANICO, A. F.: *Un Estudio del Comportamiento de los Ingresos Tributarios de la Provincia de Tucumán, Aplicando un Análisis de Series de Tiempo*”. Documentos de Trabajo de Economistas de la Coruña, Segundo volumen, Número 9. Referencia ISSN1579-1475.
- Ø CODIGOS TRIBUTARIOS de la Provincia de Tucumán y Provincias del NOA.

Figura 1: Participación porcentual de cada impuesto en el total (Enero-Julio de 2004)

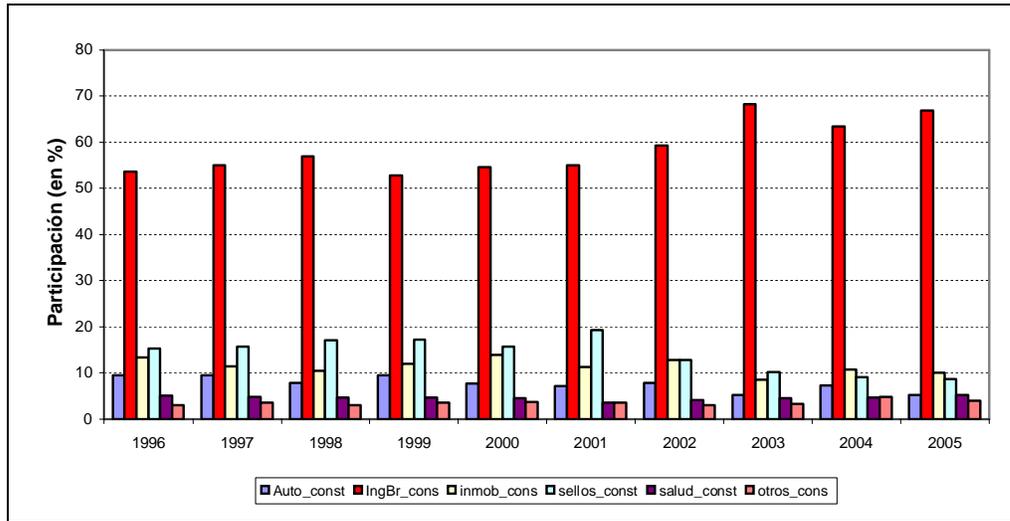


Figura 2: Participación porcentual de cada impuesto en el total (Julio a Diciembre de 2005)

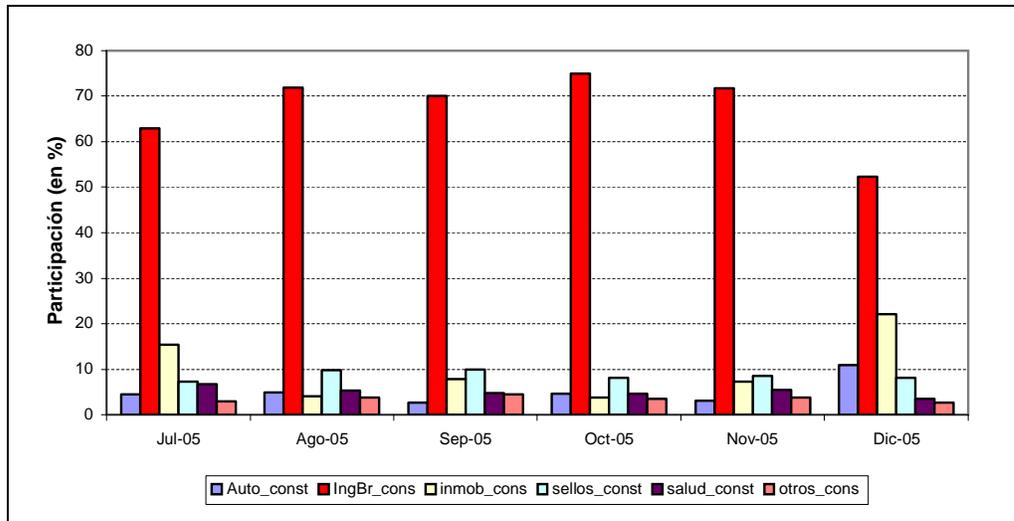


Tabla 1: Participación y propagación de los errores para las proyecciones de octubre, noviembre y diciembre, en pesos constantes

	Automot.	Ing. Brutos	Inmobil.	Sellos	Sal. Pública	Otros	Total (suma)	Total (proyec.)
Oct-05	1.035.800	17.240.000	1.127.200	2.513.200	1.257.100	1.005.200	24.178.500	24.636.000
Particip.	0,043	0,713	0,047	0,104	0,052	0,042	1,000	1,019
Error	0,331	0,040	0,291	0,101	0,064	0,235		0,057
$p_i^2 e_i^2$	0,0002011	0,00081	0,00018	0,00011	0,00001	0,00010	0,00142	0,003
Nov-05	691.290	17.749.000	1.807.400	2.198.800	1.240.600	980.440	24.667.530	24.657.000
Particip.	0,03	0,72	0,07	0,09	0,05	0,04	1,00	1,00
Error	0,346	0,042	0,191	0,129	0,073	0,245		0,065
$p_i^2 e_i^2$	0,000094	0,000913	0,000196	0,000132	0,000013	0,000095	0,001	0,0042
Dic-05	1.640.100	17.749.000	1.266.300	2.358.200	1.183.100	891.190	25.087.890	25.662.000
Particip.	0,065	0,707	0,050	0,094	0,047	0,036	1,000	1,023
Error	0,363	0,045	0,287	0,131	0,083	0,254		0,069
$p_i^2 e_i^2$	0,000563	0,001014	0,000210	0,000152	0,000015	0,000081	0,002	0,0050

Tabla 2: Participación y propagación de los errores para las proyecciones de noviembre, diciembre y enero de 2006, en pesos constantes.

	Automot.	Ing. Brutos	Inmobil.	Sellos	Sal. Pública	Otros	Total (suma)	Total (proyec.)
Nov-05	834.370	18.212.000	1.693.800	1.958.000	1.143.200	985.620	24.826.990	24.424.000
Particip.	0,034	0,734	0,068	0,079	0,046	0,040	1,000	0,984
Error	0,337	0,036	0,251	0,108	0,083	0,254		0,058
$p_i^2 e_i^2$	0,0001283	0,00070	0,00029	0,00007	0,00001	0,00010	0,00131	0,0033
Dic-05	2.841.600	17.872.000	1.013.100	2.108.600	1.109.400	901.030	25.845.730	25.421.000
Particip.	0,11	0,69	0,04	0,08	0,04	0,03	1,00	0,98
Error	0,338	0,039	0,290	0,108	0,097	0,343		0,063
$p_i^2 e_i^2$	0,001381	0,000727	0,000129	0,000078	0,000017	0,000143	0,00248	0,0038
Ene-06	1.037.200	18.777.000	2.351.200	2.127.000	1.363.200	644.470	26.300.070	27.695.000
Particip.	0,039	0,714	0,089	0,081	0,052	0,025	1,000	1,053
Error	0,340	0,040	0,211	0,114	0,087	0,191		0,064
$p_i^2 e_i^2$	0,000180	0,000816	0,000356	0,000085	0,000020	0,000022	0,00148	0,0045

Tabla 3: Participación y propagación de los errores para las proyecciones de diciembre, enero y febrero, en pesos constantes.

	Automot.	Ing. Brutos	Inmobil.	Sellos	Sal. Pública	Otros	Total (suma)	Total (proyec.)
Dic-05	1.560.400	18.153.000	1.458.300	2.146.400	1.190.400	875.820	25.384.320	25.346.000
Particip.	0,061	0,715	0,057	0,085	0,047	0,035	1,000	0,998
Error	0,179	0,037	0,389	0,104	0,079	0,285		0,060
$p_i^2 e_i^2$	0,0001211	0,00070	0,00050	0,00008	0,00001	0,00010	0,00151	0,0036
Ene-06	1.068.100	19.492.000	3.623.800	2.224.500	1.431.100	685.140	28.524.640	27.164.000
Particip.	0,04	0,68	0,13	0,08	0,05	0,02	1,00	0,95
Error	0,310	0,036	0,438	0,108	0,075	0,343		0,066
$p_i^2 e_i^2$	0,000135	0,000605	0,003096	0,000071	0,000014	0,000068	0,00399	0,0040
Feb-06	1.415.400	18.407.000	1.110.200	1.905.000	1.280.900	870.620	24.989.120	23.548.000
Particip.	0,057	0,737	0,044	0,076	0,051	0,035	1,000	0,942
Error	0,250	0,040	0,458	0,135	0,093	0,199		0,087
$p_i^2 e_i^2$	0,000201	0,000868	0,000414	0,000106	0,000023	0,000048	0,00166	0,0067