

4. ESTADÍSTICA OFICIAL

PREDICCIÓN DEL IPC UTILIZANDO MÉTODOS DE SUBESPACIO

Victor Gómez Enríquez¹, Félix Aparicio Pérez² y Angel Sánchez Avila¹¹ Ministerio de Economía.² Instituto Nacional de Estadística.**Resumen**

En la literatura sobre series temporales últimamente se está observando interés en los llamados métodos de subespacio. Estos métodos utilizan regresión lineal y análisis de correlaciones canónicas para estimar un modelo ARMAX expresado en forma de espacio de estados. En este trabajo utilizamos métodos de subespacio para predecir el IPC con ayuda de algunas variables relacionadas con él. Comparamos los resultados con los obtenidos empleando una función de transferencia. Las conclusiones son que ambos métodos proporcionan resultados parecidos, un poco mejores a corto plazo los de la función de transferencia y un poco mejores a medio y largo plazo los de los métodos de subespacio. Este resultado es alentador, puesto que con los métodos de subespacio se consigue un ajuste casi automático, en contraste con la función de transferencia, que precisa de considerable trabajo a medida.

1. Introducción

En este trabajo se estudian las predicciones del Índice de Precios de Consumo (IPC) español, con año base 2001 utilizando métodos de subespacio. Se comparan estas predicciones con las de una función de transferencia clásica. En ambos casos se toman como variables exógenas el Índice de Precios Industriales para Bienes de Consumo, (IPI, año base 2000), el Precio (Spot) del barril de Petróleo Brent (OIL), y el Índice de Importación de Bienes de Consumo, (IIBC, año base 2000). La variable OIL se considera como determinística y sus predicciones son, simplemente, los precios en el Mercado de Futuros que corresponden a contratos con el mismo vencimiento. En las Figuras 1 y 2 se muestran las cuatro series en el periodo 1993:1–2006:3.

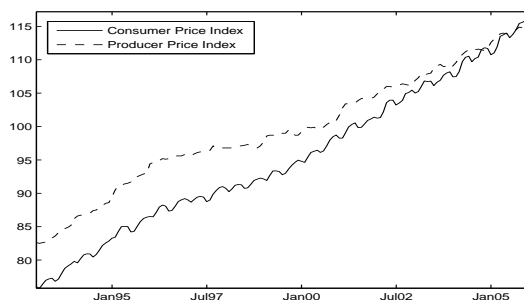


Figura 1: IPC y IPI españoles, 1993:1–2006:3.

Para comparar las predicciones se emplea un estimador móvil, que utiliza doce modelos, el primero hace predicciones basándose en la muestra 1993:1–2004:12, y cada uno de los siguientes añade un mes, así el último basa sus predicciones en la muestra 1993:1–2005:12. Cada uno de estos 12 modelos hace predicciones para los 15 meses posteriores a la muestra en que se basa.

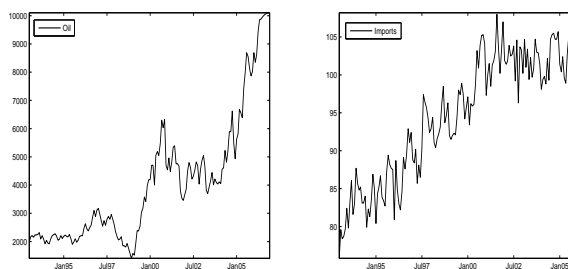


Figura 2: OIL, 1993:1–2006:3, y IIBC, 1993:1–2006:2.

2. Función de Transferencia

Para estimar las funciones de transferencia, en primer lugar ajustamos un modelo VAR a las cuatro series conjuntamente. Se concluyó que el número de retardos adecuados a considerar es 13, que no hay relación entre IIBC y las demás series y que la relación entre los inputs IPI (x_{1t}) y OIL (x_{2t}) y

el output IPC (y_t) es unidireccional. Por tanto se ajustaron modelos de función de transferencia para explicar IPC en función de IPI y OIL.

Tomando 13 retardos en los filtros se llegó al modelo (con la muestra 1993:1–2004:12)

$$\nabla\nabla_{12}y_t = \frac{\omega_{10} + \omega_{11}B}{1 + \delta_1 B + \delta_2 B^2} \nabla\nabla_{12}x_{1t} + (\omega_{20} + \omega_{21}B) \nabla\nabla_{12}x_{2t} + \frac{1}{1 + \Phi B^{12}} a_t,$$

donde $\nabla = 1 - B$ y $\nabla_{12} = 1 - B^{12}$ son los operadores de diferenciación habituales. Los tests de diagnóstico standard confirman que el ajuste es bueno. También se concluye que el modelo no cambia (excepto los valores numéricos de sus parámetros) al ir añadiendo cada observación adicional hasta el total de 12.

3. Métodos de Subespacio

Supongamos que nuestro sistema se puede expresar en forma de espacio de estados de las innovaciones:

$$x_{t+1} = Ax_t + Bu_t + Ke_t, \quad (1)$$

$$y_t = Cx_t + Du_t + e_t. \quad (2)$$

Si tenemos una estimación del estado, entonces podemos emplear la ecuación de regresión

$$\begin{bmatrix} \hat{x}_{t+1} \\ y_t \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \hat{x}_t \\ u_t \end{bmatrix} + \begin{bmatrix} w_t \\ v_t \end{bmatrix}$$

para estimar las matrices del sistema A , B , C y D . Después usamos los residuos de la regresión (\hat{w}'_t, \hat{v}'_t) para estimar $W = Var(e'_t K', e'_t)'$. Sea esta estimación

$$\widehat{W} = \begin{bmatrix} M & N \\ N' & R \end{bmatrix}.$$

Resolvemos la siguiente Ecuación Algebraica de Riccati en tiempo discreto $P = APA' - (APC' + N)(CPC' + R)^{-1}(APC' + N)' + M$ para obtener la solución de estabilidad $P \geq 0$ y la correspondiente ganancia de Kalman $K = (APC' + N)(CPC' + R)^{-1}$ que hace a $A - KC$ estable.

Para obtener las estimaciones del estado calculamos regresiones de observaciones "futuras" con respecto a observaciones "pasadas". (Hacemos regresión de trozos de longitud f de observaciones adyacentes del output "futuro" frente a trozos de longitud p de observaciones adyacentes del output e inputs "pasados", para enteros f y p adecuados).

El problema de elegir el orden del sistema (la dimensión del estado) se resuelve utilizando los resultados clásicos de Anderson sobre correlaciones canónicas parciales.

En cuanto a los enteros f y p una posibilidad es ajustar un modelo ARX a los datos y hacer p igual al número de retardos en el modelo resultante. El número de retardos del "futuro" f , debería ser al menos igual al orden del sistema n . Otra posibilidad es hacer que f crezca con el tamaño de la muestra utilizando una fórmula del tipo $f = \log^a(T)$ con $1,5 \leq a \leq 2$.

Empleamos el método de subespacio a las series diferenciadas utilizando el operador $\nabla\nabla_{12}$. Utilizamos los siguientes valores de f y p . Para tamaños muestrales entre 134 y 137, $f = 18$ y $p = 11$, para tamaños muestrales entre 138 y 141, $f = 19$ y $p = 11$, y para el resto de tamaños muestrales $f = 20$ y $p = 12$.

Para calcular las predicciones utilizamos un modelo de espacio de estados aumentado que incluye al modelo (1) y (2) junto con un modelo en forma de espacio de estados para el input IPI. El otro input, OIL, como dijimos, se considera determinístico y sus predicciones son simplemente los precios correspondientes en el Mercado de Futuros de un contrato con la misma fecha de vencimiento que el horizonte de la predicción.

El modelo para la serie IPI es un modelo de líneas aéreas con una intervención,

$$x_{1t} = \frac{1}{1 - \delta B} I_t^{37} + \frac{(1 + \theta B)(1 + \theta B^{12})}{\nabla\nabla_{12}} a_t,$$

donde $0 < \delta < 1$ y $I_t^{37} = 1$ si $t = 37$ y $I_t^{37} = 0$ en otro caso.

4. Resultados

Las Tablas 1 a 3 resumen los resultados de la predicción de los dos modelos. El RMSE de la Tabla 1 se define como

$$\sqrt{\frac{1}{12} \sum_{t=T+h}^{T+11+h} (y_t - y_{t|t-h})^2},$$

donde T es el tamaño de la muestra, h es el número de periodos adelante de la predicción y $y_{t|t-h}$ es la predicción de 1 a 15 meses hacia el futuro de cada uno de los dos modelos.

Las Tablas 2 y 3 resumen la media de los errores de predicción $y_t - y_{t|t-h}$, para los dos modelos.

Podemos ver que la función de transferencia predice mejor para horizontes de entre uno y cuatro meses hacia adelante, y peor para el resto de horizontes. Hay un sesgo sistemático en las predicciones de los dos modelos, que se debe al gran incremento en los precios de la variable OIL en el período en que se calculan las predicciones. Este incremento no lo pueden capturar ninguno de los dos modelos.

Estos resultados sugieren que los métodos de subespacio pueden ser una alternativa a los métodos clásicos empleados en análisis de series temporales, con la ventaja de que su utilización es casi automática, mientras que algunos de los métodos empleados en análisis de series temporales, como la función de transferencia, requieren bastante trabajo a medida de cada caso realizado por un experto.

Tabla 1. Cociente de RMSE's por horizonte

Horizonte							
1	2	3	4	5	6	7	8
1.10	1.27	1.16	1.04	0.91	0.79	0.69	0.69

Horizonte						
9	10	11	12	13	14	15
0.72	0.75	0.71	0.70	0.72	0.73	0.73

NOTA: Los valores son los cocientes entre los RMSE's de las predicciones por los métodos de subespacio y función de transferencia para cada horizonte.

Tabla 2. Sesgo de las predicciones (subespacio)

Horizonte							
1	2	3	4	5	6	7	8
1.41	2.27	2.76	3.66	3.65	3.11	2.86	3.13

Horizonte						
9	10	11	12	13	14	15
3.08	3.12	2.58	2.47	3.36	4.15	4.71

NOTA: Los valores son 10^3 veces las medias de las diferencias entre los valores verdaderos y las predicciones para cada horizonte.

Tabla 3. Sesgo de la predicciones (transferencia)

Horizonte							
1	2	3	4	5	6	7	8
1.25	1.74	2.67	3.40	3.83	3.91	4.25	4.02

Horizonte						
9	10	11	12	13	14	15
4.17	3.68	3.36	4.18	5.39	6.47	7.53

NOTA: Los valores son 10^3 veces las medias de las diferencias entre los valores verdaderos y las predicciones para cada horizonte

Referencias

- [1] Anderson T.W. (1951). Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions. *Annals of Mathematical Statistics*, **22**, 327-351.
- [2] Beguin J.M., Gourieroux C., and Monfort A. (1980). *Identifcation of a Mixed Autoregressive-Moving Average Process: the Corner Method*, in Time Series (Anderson, O.D., ed.), Amsterdam: North-Holland.
- [3] Box G.E.P., and Jenkins, G.M. (1976). *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco.
- [4] Larimore W.E. (1983). System identification, reduced order filters and modeling via canonical variate analysis. *Proceedings of the 1983 American Control Conference 2*, (H.S. Rao and P. Dorato, eds.), Piscataway, NJ, 445-451.
- [5] Ljung L. (1999). *System Identification: Theory for the User*, Prentice Hall, New Jersey.
- [6] Liu L.M., and Hanssens D.M. (1982). Identification of Multiple-Input Transfer Function Models. *Communications in Statistics, Theory and Methods*, **11**, 297-314.
- [7] Tsay R. S. (1985). Model Identification in Dynamic Regression (Distributed Lag) Models. *Journal of Business and Economic Statistics*, **3**, 228-237.
- [8] Tsay R.S. (1989). Parsimonious Parameterization of Vector Autoregressive Moving Average Models. *Journal of Business and Economic Statistics*, **7**, 327-341.