

Predicción de procesos ARMA mediante modelos autorregresivos y en el espacio de estados

Segismundo S. Izquierdo Millán^{1*}, Javier Pajares Gutiérrez¹, Cesáreo Hernández Iglesias¹

¹ Dpto. de Organización de Empresas y CIM. E.T.S. Ingenieros Industriales. Pº del Cauce s/n, 47011 Valladolid. segis@eis.uva.es, pajares@eis.uva.es, cesareo@eis.uva.es

Resumen

En econometría es frecuente el uso de modelos autorregresivos (AR) para representar procesos estocásticos lineales. Los modelos AR ofrecen algunas características deseables tales como la sencillez para especificar el modelo (seleccionar el orden) o para estimar los parámetros (la estimación puede realizarse por mínimos cuadrados en un solo paso). Sin embargo, si el proceso en estudio sigue en realidad un modelo autorregresivo media móvil (ARMA), entonces no puede ser representado de forma exacta por ningún modelo AR de orden finito. Por el contrario, un modelo de orden finito en el espacio de estados (SS) sí puede representar de forma exactamente equivalente un proceso ARMA. Los métodos de subespacios ofrecen además la posibilidad de realizar una especificación y estimación rápida y sencilla de modelos SS.

Este artículo estudia la cuestión de si, al trabajar con procesos tipo ARMA, los modelos en el espacio de estados obtenidos mediante un método de subespacios proporcionan mejores o peores predicciones que los modelos autorregresivos estimados por mínimos cuadrados. Los resultados de nuestras simulaciones indican que, cuando el componente MA de un ARMA(1,1) es importante (cuando su valor es próximo a 1), los modelos en el espacio de estados tienen alta probabilidad de proporcionar mejores predicciones que los modelos AR.

Palabras clave: predicción, identificación de sistemas, subespacios

1. Introducción

Los modelos autorregresivos (AR) son una opción frecuente para el modelado y predicción de procesos estocásticos lineales. Otras opciones frecuentes son los modelos autorregresivos medias móviles (ARMA) y los modelos en el espacio de estados (SS). La formulación de un modelo AR(p) es de la forma:

$$y_t + a_1 y_{t-1} + \dots + a_p y_{t-p} = e_t \quad (1)$$

donde e_t es ruido blanco con $E(e_t) = 0$ y $E(e_t e_t) = \sigma^2$. Por su parte, un modelo ARMA(p,q) es de la forma:

$$y_t + a_1 y_{t-1} + \dots + a_p y_{t-p} = e_t + b_1 e_{t-1} + \dots + b_q e_{t-q} \quad (2)$$

* Este trabajo se deriva de la participación de sus autores en un proyecto de investigación financiado por el Ministerio de Educación y Ciencia con referencia DPI2005-05676, titulado "SIGAME".

y, finalmente, un modelo en el espacio de estados $SS(n)$ es de la forma:

$$\mathbf{z}_{t+1} = \mathbf{A} \mathbf{z}_t + \mathbf{K} \mathbf{e}_t \quad \text{Ecuación de transición de estados} \quad (3)$$

$$y_t = \mathbf{C} \mathbf{z}_t + \mathbf{e}_t \quad \text{Ecuación de observaciones} \quad (4)$$

donde \mathbf{z}_t es un vector ($n \times 1$) de variables auxiliares denominado vector de estados, y \mathbf{A} , \mathbf{K} , \mathbf{C} son matrices constantes de dimensiones apropiadas.

Los modelos AR ofrecen entre sus características deseables la facilidad para estimar los parámetros del modelo (esta estimación puede llevarse a cabo en un sólo paso, sin recurrir a técnicas numéricas iterativas) y la facilidad para estimar un orden para el modelo. Sin embargo, si el sistema en estudio sigue un proceso ARMA, no puede ser representado de forma exacta por un modelo AR finito.

Un modelo finito en el espacio de estados sí puede representar de forma exacta un proceso ARMA. Además, los métodos de subespacios permiten realizar de forma rápida la estimación de los parámetros y del orden del modelo.

En este artículo comparamos, en procesos ARMA(1,1) simulados, la capacidad predictiva de modelos autorregresivos con la capacidad predictiva de modelos en el espacio de estados.

2. Método

La comparación de modelos no es directa, porque existen distintas técnicas de identificación de sistemas (obtención del modelo: especificación y estimación de parámetros) que pueden usarse, y las distintas técnicas pueden conducir a distintos modelos. Para la identificación de modelos AR elegimos una estimación que minimice el error de predicción en la muestra (PEM, Ljung 1999), que es equivalente a una estimación por mínimos cuadrados (LS), y una especificación conforme al criterio de información de Akaike (AIC). Para la identificación de modelos en el espacio de estados elegimos el algoritmo de Análisis de Correlaciones Canónicas (CCA, Larimore 1983), el cual, dentro de la familia de métodos de subespacios, presenta algunas propiedades óptimas para la identificación de procesos estocásticos estacionarios (Bauer and Ljung 2002).

Para distintas réplicas de procesos ARMA(1,1) se identificaron modelos AR y SS (utilizando muestras de trabajo) y se midió el error cuadrático medio de predicción a horizonte uno en muestras de predicción. Los resultados fueron analizados utilizando varios de los estadísticos propuestos por Diebold y Mariano (2002).

En particular, consideraremos un proceso generador de datos ARMA(1,1) cuyas propiedades estadísticas dependen de un vector de parámetros $(\varphi, \theta, \sigma)$. Para distintas combinaciones de valores de los parámetros, el proceso seguido es:

- Generar 1000 muestras de tamaño T (para diferentes valores de T), más 10 observaciones adicionales en cada muestra, que usaremos para calcular el error de predicción (muestra de predicción).
- Para cada muestra, obtener un modelo SS y un modelo AR (usando T observaciones).

- Para cada $i = 1, \dots, 9, 10$, medir el error de predicción a horizonte 1 para la observación $T+i$ (muestra de predicción), recalculando los modelos SS y AR a medida que crece i usando todas las observaciones anteriores.
- Obtener el error cuadrático medio de predicción a horizonte uno en la muestra de predicción, tanto el valor correspondiente a los modelos SS ($MSPE_{SS}$) como el correspondiente a los modelos AR ($MSPE_{AR}$).

De este modo, para cada tamaño muestral T y para cada combinación de valores de los parámetros del modelo ARMA(1,1) obtenemos 1000 pares de valores ($MSPE_{SS}$, $MSPE_{AR}$) que podemos comparar.

3. Resultados

Consideramos el proceso univariante estocástico ARMA(1,1) definido por la ecuación:

$$y_t - \varphi y_{t-1} = e_t + \theta e_{t-1}$$

donde e_t es ruido blanco $N(0, \sigma^2)$. Una formulación equivalente en el espacio de estados es

$$\begin{aligned} z_{t+1} &= \varphi z_t + e_t \\ y_t &= (\varphi + \theta) z_t + e_t \end{aligned}$$

Buscando procesos con distintas propiedades estadísticas dentro de este modelo ARMA(1,1) hemos considerado el espacio paramétrico $\varphi = [0, .5, .9, 1]$ y $\theta = [0, .5, .9]$, y los tamaños muestrales $T = [50, 100, 200, 500]$. Dado que el proceso es totalmente estocástico, no hay pérdida de generalidad por asumir $E(e_t e_t) = \sigma^2 = 1$, dado que simular una serie usando una varianza $E(e_t e_t) = \lambda^2$ es equivalente a simular la serie usando una varianza $E(e_t e_t) = 1$ y multiplicar los valores por un factor λ (equivalente a cambiar la escala). Los estimadores proporcionados por los métodos comparados (CCA, LS) son invariantes frente a cambios de escala.

3.1. Recuperación de los parámetros del modelo original

Antes de comenzar las comparaciones “calibraremos” la capacidad del algoritmo CCA para recuperar los valores de los parámetros del modelo φ , θ y σ .

Un modelo SS(1) formulado como en (3) y (4):

$$\begin{aligned} z_{t+1} &= A z_t + K e_t \\ y_t &= C z_t + e_t \end{aligned}$$

admite una representación ARMA(1,1) equivalente expresada como

$$y_t - A y_{t-1} = e_t + (C \cdot K - A) e_{t-1} \tag{5}$$

a la que se puede llegar por eliminación directa del estado en las ecuaciones ($z_t = C^{-1}(y_t - e_t)$) o bien aplicando un procedimiento de conversión SS-ARMA como el propuesto por Aoki y Havenner (1991).

Así pues, una vez tenemos un modelo SS(1) estimado, los parámetros de una representación ARMA(1,1) equivalente pueden obtenerse como $\hat{\phi} = \hat{A}$, $\hat{\theta} = \hat{C} \cdot \hat{K} - \hat{A}$, $\hat{\sigma}^2 = \hat{R}$.

Para cada combinación de valores de ϕ , θ y T generamos 1000 muestras, y para cada muestra estimamos un modelo SS(1) usando el algoritmo CCA y un modelos ARMA(1,1) usando un método PEM (Ljung 1999) equivalente a máxima verosimilitud (ML). A continuación calculamos y representamos los sesgos medios $\hat{\phi} - \phi$, $\hat{\theta} - \theta$, $\hat{\sigma}^2 - \sigma^2$ y sus desviaciones típicas.

En la Figura 1 presentamos resultados representativos para la estimación de ϕ . En general las estimaciones proporcionadas por CCA están muy próximas a las proporcionadas por ML, pero para bajos tamaños muestrales ($T = 50$) las estimaciones CCA presentan una dispersión mucho mayor que las estimaciones ML.

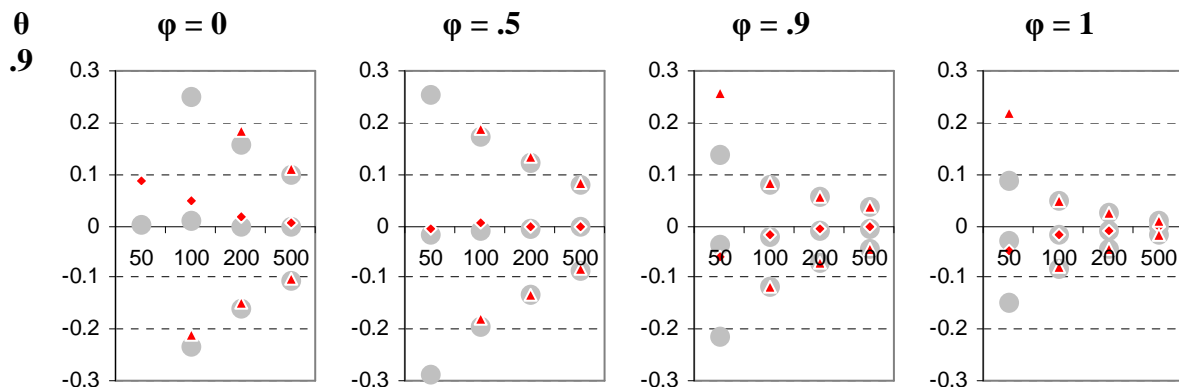


Figura 1. Sesgo medio y bandas a ± 2 errores estándar para la estimación de ϕ . Con $\theta = .9$ y $T = [50, 100, 200, 500]$. Los puntos grises corresponden a las estimaciones ARMA(1,1)-ML y los pequeños rombos rojos (sesgo medio) así como los triángulos (bandas de error) corresponden a las estimaciones SS(1)-CCA. Los valores fuera del rango de las figuras no se han representado.

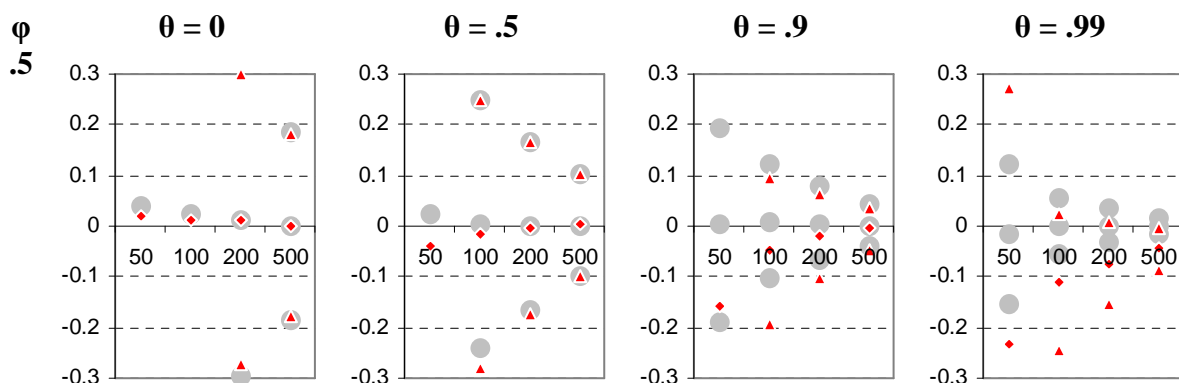


Figura 2. Sesgo medio y bandas a ± 2 errores estándar para la estimación de θ . Con $\phi = .5$ y $T = [50, 100, 200, 500]$. Los puntos grises corresponden a las estimaciones ARMA(1,1)-ML y los pequeños rombos rojos (sesgo medio) así como los triángulos (bandas de error) corresponden a las estimaciones SS(1)-CCA. Los valores fuera del rango de las figuras no se han representado.

En la Figura 2 podemos observar valores representativos para la estimación del parámetro de media móvil θ . A medida que el valor de θ se aproxima a 1 las estimaciones provistas por CCA presentan mayor sesgo, mientras que el sesgo de las estimaciones ML se mantiene reducido. Para bajos tamaños muestrales la dispersión de las estimaciones CCA es también mayor.

Nótese que la diferente precisión en la estimación de los parámetros del ARMA no está asociada a la representación elegida (SS o ARMA) sino a los diferentes procedimientos de estimación elegidos (CCA o ML). En la **Figura 3** podemos comprobar cómo las estimaciones obtenidas usando una representación SS(1) con un procedimiento iterativo ML son prácticamente las mismas que usando la representación ARMA(1,1). Existen algunas diferencias para $T = 50$, pero estas son posiblemente debidas a que los dos métodos usan estimaciones iniciales distintas para comenzar la optimización numérica del procedimiento ML. El método ML proporciona mejores resultados que el CCA (de hecho, el método ML realiza una búsqueda por métodos numéricos de los parámetros que maximizan la función de verosimilitud muestral tomando como valores iniciales los proporcionados por CCA), pero es más costoso en términos computacionales, y presenta los problemas asociados a los métodos iterativos.

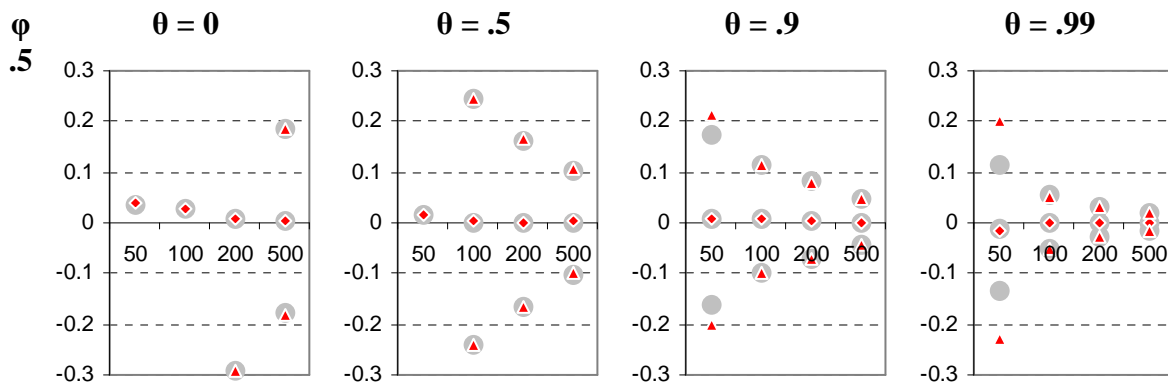


Figura 3. Sesgo medio y bandas a ± 2 errores estándar para la estimación de θ . Con $\varphi = .5$ y $T = [50, 100, 200, 500]$. Los puntos grises corresponden a las estimaciones ARMA(1,1)-ML y los pequeños rombos rojos (sesgo medio) así como los triángulos (bandas de error) corresponden a las estimaciones SS(1)-ML con estimador inicial para la búsqueda numérica proporcionado por CCA. Los valores fuera del rango de las figuras no se han representado.

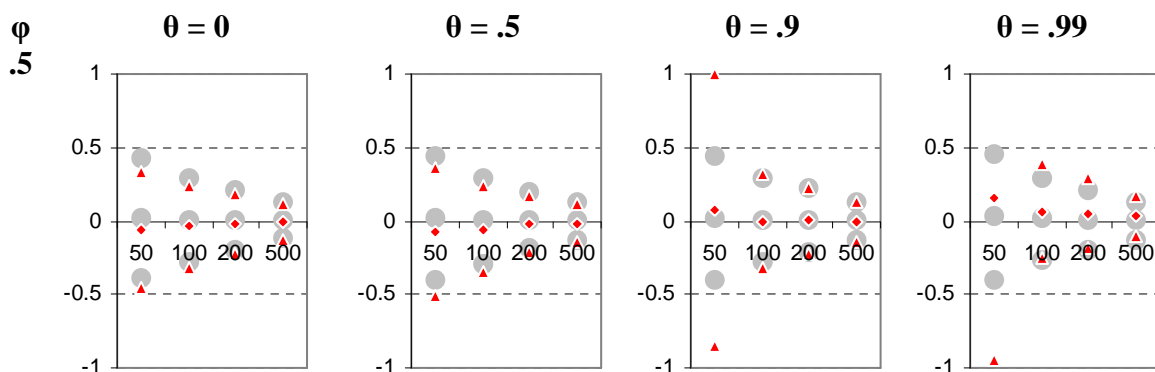


Figura 4. Sesgo medio y bandas a ± 2 errores estándar para la estimación de σ^2 . Con $\varphi = .5$ y $T = [50, 100, 200, 500]$. Los puntos grises corresponden a las estimaciones ARMA(1,1)-ML y los pequeños rombos rojos (sesgo medio) así como los triángulos (bandas de error) corresponden a las estimaciones SS(1)-CCA. Los valores fuera del rango de las figuras no se han representado.

Por ultimo, la Figura 4 y la Figura 5 muestran la precisión en la estimación de σ^2 . Las principales diferencias se dan de nuevo para $T = 50$, donde las estimaciones CCA presentan una elevada variabilidad.

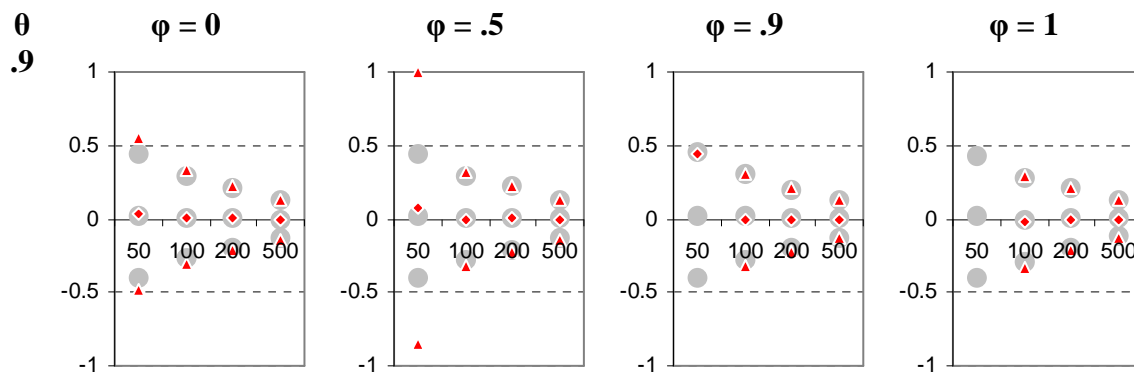


Figura 5. Sesgo medio y bandas a ± 2 errores estándar para la estimación de σ^2 . Con $\theta = .9$ y $T = [50, 100, 200, 500]$. Los puntos grises corresponden a las estimaciones ARMA(1,1)-ML y los pequeños rombos rojos (sesgo medio) así como los triángulos (bandas de error) corresponden a las estimaciones SS(1)-CCA. Los valores fuera del rango de las figuras no se han representado.

En resumen, en estos experimentos, asumiendo un tamaño muestral superior a 100, el algoritmo CCA que estamos estudiando recupera bien el parámetro autorregresivo φ del modelo ARMA(1,1), pero no tan bien el parámetro de media móvil θ , especialmente cuando su valor está próximo a 1.

Nótese que en estos experimentos hemos impuesto los órdenes de las representaciones SS(1) y ARMA(1,1) para que sean los que corresponden al proceso generador de datos subyacente. Esta condición es necesaria para que exista una formulación (especificación) equivalente común de los modelos obtenidos y poder así comparar la precisión en la estimación de los parámetros. A continuación realizaremos comparaciones de capacidad predictiva en las que no impondremos (sino que estimaremos) el orden de los modelos (incluyendo modelos AR). Dado que entonces obtendremos especificaciones no equivalentes, no será posible comparar la precisión en la estimación de parámetros.

3.2. Predicción

Compararemos la capacidad predictiva de modelos AR estimados por mínimos cuadrados (AR-LS) con la capacidad predictiva de modelos SS estimados por el algoritmo de subespacios CCA (SS-CCA).

Conforme a la metodología anteriormente indicada, para cada combinación de valores (φ , θ , T) considerada generamos 1000 réplicas de tamaño $T + 10$. Para cada réplica y para cada método de identificación de sistemas (AR-LS, SS-CCA) calculamos el indicador de calidad de predicción a horizonte uno ($MSPE_{\text{método}}$), usando las últimas 10 observaciones de la muestra para medir el error de predicción, y recalculando el modelo con todas las observaciones anteriores a medida que avanzamos en la muestra de predicción. Por último comparamos los valores $MSPE_{\text{método}}$ obtenidos para los diferentes métodos estudiados.

Cualitativamente los resultados mostraron poca dependencia de los valores de φ considerados, por lo que mostraremos resultados representativos correspondientes a $\varphi = .9$.

En la Figura 6 mostramos, para distintas combinaciones de T y θ , gráficos de los 1000 puntos ($MSPE_{AR-LS}$, $MSPE_{SS-CCA}$) obtenidos. Los puntos por debajo de la bisectriz del primer cuadrante (definida por la ecuación $MSPE_{SS-CCA} = MSPE_{AR-LS}$) corresponden a réplicas en las que $MSPE_{SS-CCA} < MSPE_{AR-LS}$, es decir, en las que los modelos SS-CCA proporcionaron predicciones con menor MSPE que los modelos AR-LS.

Hemos representado también la recta de regresión de $MSPE_{SS-CCA}$ sobre $MSPE_{AR-LS}$, y las pendientes de estas rectas se muestran en la **Tabla 1**; los valores menores de 1 están a favor de los modelos CCA.

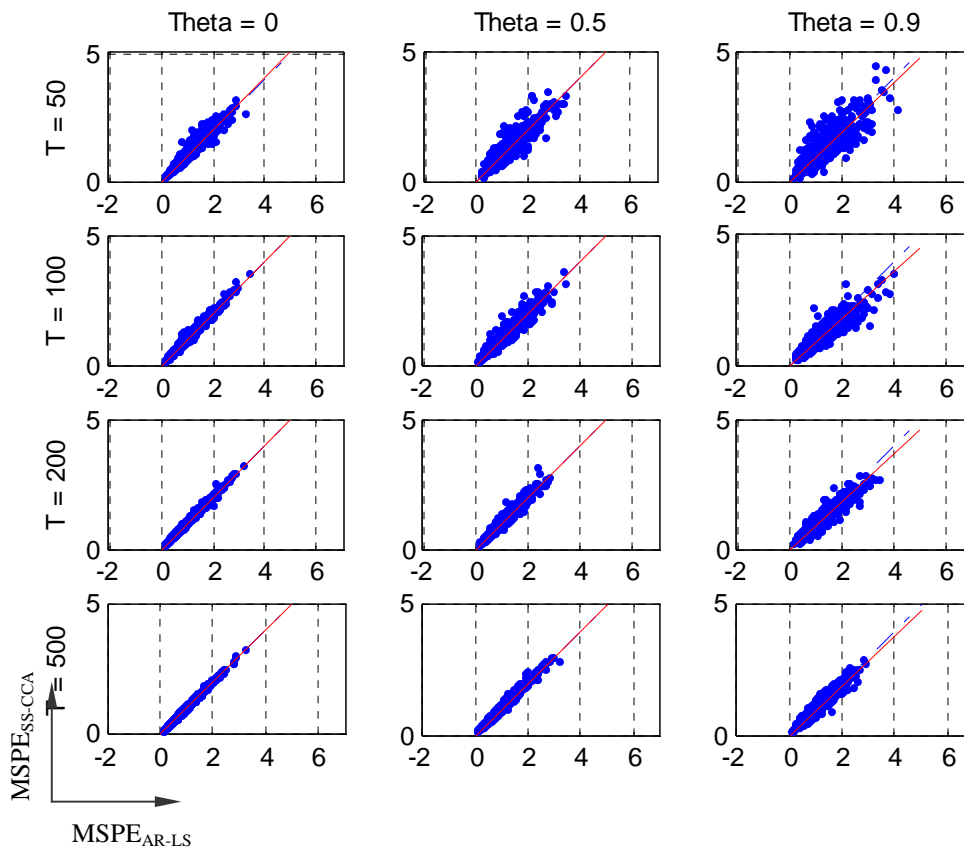


Figura 6. MSPE de las predicciones de modelos SS-CCA frente a MSPE de las predicciones de modelos AR-LS en series de 1000 réplicas de un ARMA(1,1) con $\varphi = .9$

Tabla 1. Pendientes de las rectas de regresión de $MSPE_{SS-CCA}$ sobre $MSPE_{AR-LS}$ calculadas en series de 1000 réplicas de un ARMA(1,1) con $\varphi = .9$

	θ		
T	0	0.5	0.9
50	1.02	1.01	0.94
100	1.01	0.99	0.90
200	1.00	1.00	0.92
500	1.00	1.00	0.95

En la **Tabla 2** mostramos el número de réplicas en las que los modelos SS-CCA “ganan” a los modelos AR-LS, calculado como el número de réplicas en las que $MSPE_{SS-CCA} < MSPE_{AR-LS}$ más la mitad del número de réplicas en las que $MSPE_{SS-CCA} = MSPE_{AR-LS}$ ¹.

Con los datos de la **Tabla 2** podemos realizar un contraste estadístico binomial (Siegel and Castellan 1988, Diebold and Mariano 2002). Bajo la hipótesis nula de que los modelos SS-CCA y AR-LS tienen a priori la misma probabilidad de proporcionar la mejor (menor MSPE) predicción, es decir, la misma probabilidad a priori de “ganar” en un “torneo de predicciones”, el número de veces que los modelos SS-CCA vencen a los modelos AR-LS en una serie de N réplicas debe seguir una distribución binomial tal que la probabilidad de x victorias es

$$P(x \text{ de entre } N) = \binom{N}{x} (1/2)^x (1 - 1/2)^{N-x}$$

Para una serie de 1000 réplicas se tiene $P(459 \leq x \leq 541) = 0.991$, y $P(448 \leq x \leq 552) = 0.999$. Luego para valores de x fuera del rango [459, 541] podemos rechazar (con error $\alpha < 0.01$) la hipótesis nula de que los dos métodos tienen la misma probabilidad de proporcionar la mejor predicción, y aceptar que uno de los métodos tiene mayor probabilidad que el otro de proporcionar las predicciones de menor MSPE.

Tabla 2. Número de replicas en las que $MSPE_{SS-CCA} < MSPE_{AR-LS}$, más la mitad del número de replicas en las que $MSPE_{SS-CCA} = MSPE_{AR-LS}$, de entre series de 1000 réplicas. En negrita los valores significativos para un contraste binomial ($H_0 : P(MSPE_{SS-CCA} < MSPE_{AR-LS}) = P(MSPE_{AR-LS} < MSPE_{SS-CCA})$; $\alpha < 0.01$, contraste de dos lados).

T	θ		
	0	0.5	0.9
50	370.5	472	575
100	420.5	514	653
200	447.5	544.5	677
500	421	512	644

En la **Tabla 2** podemos observar cómo cuando el componente de media móvil θ es 0 (cuando el proceso generador es un AR), hay diferencias significativas a favor de los modelos AR-LS, pero cuando el parámetro θ es .9, hay diferencias especialmente significativas a favor de los modelos SS-CCA.

En la **Tabla 3** podemos ver el incremento porcentual del MSPE de los modelos SS-CCA con respecto al MSPE de los modelos AR-LS. Podemos usar un contraste para la “igualdad de poder predictivo” basado en Diebold y Mariano (2002): dado que las 1000 réplicas generadas para una combinación de valores de los parámetros fija son independientes, podemos asumir que los valores $d_i = (MSPE_{SS-CCA} - MSPE_{AR-LS})_i$, con $i = 1, 2, \dots, 1000$, provienen de variables i.i.d. (en nuestras simulaciones los autocorrelogramas de los momentos centrados de d_i apoyan la hipótesis de independencia). Así pues, para contrastar la hipótesis nula $E(MSPE_{SS-CCA})_i = E(MSPE_{AR-LS})_i$ o, lo que es equivalente, $E(d_i) = 0$, podemos utilizar el estadístico

¹ La incidencia de casos $MSE_{\text{método1}} = MSE_{\text{método2}}$ resultó ser muy reducida: menos de 5 casos entre 1000, con aproximadamente un 0.1% de media. Sólo cuando los métodos fueron SS(1)-ML and ARMA(1,1)-ML se produjeron un número elevado de coincidencias, llegando ocasionalmente a superar el 10 %.

$$S = \sqrt{N} \frac{\bar{d}_i}{std(d_i)}$$

donde N es el número de replicas (1000), \bar{d}_i es la media muestral y $std(d_i)$ es la desviación típica de la muestra. Bajo la hipótesis nula la distribución del estadístico S debe aproximarse a una distribución normal. La Tabla 3 muestra valores significativos para este contraste con $\alpha = 0.01$ ($|S| > 2.58$).

Nuestros resultados muestran que, para $\theta = 0$ (cuando la especificación correcta es un AR), los modelos AR-LS tienen mayor probabilidad de proporcionar predicciones mejores (con menor MSPE) que los modelos SS-CCA (Tabla 2), si bien las diferencias en términos porcentuales son muy reducidas (1 ó 2 %. Ver Tabla 3 y Figura 6). Por el contrario, para el caso $\theta = 0.9$ (ahora la especificación del proceso generador no es un AR y el valor del componente MA está próximo a la unidad), los modelos SS-CCA tienen alta probabilidad de proporcionar mejores predicciones que los modelos AR-LS (Tabla 2), y las diferencias en MSPE son más notables (un MSPE total 7% menor en algunos casos: Figura 6 y Tabla 3, columna derecha).

Tabla 3. Izquierda: incremento en la suma acumulada de $MSPE_{SS-CCA}$ con respecto a la de $MSPE_{AR-LS}$ en series de 1000 réplicas de un ARMA(1,1) con $\varphi = .9$. Derecha: valores correspondientes del estadístico S . En negrita los valores significativos ($\alpha = 0.01$).

	θ		
T	0	0.5	0.9
50	2%	3%	-3%
100	1%	0%	-7%
200	0%	0%	-7%
500	0%	0%	-4%

	θ		
T	0	0.5	0.9
50	-6,01	-4,31	3,79
100	-4,95	-0,55	12,00
200	-1,87	-0,35	13,24
500	-2,60	1,00	10,35

Estos resultados nos hacen proponer el uso de modelos SS-CCA como alternativas o como complementos a los modelos AR para la identificación de series temporales estocásticas lineales, dado que ambos modelos pueden obtenerse de forma rápida por procedimientos no iterativos, poco costosos en términos de tiempo y computacionales, y su capacidad predictiva, como acabamos de mostrar, puede resultar significativamente diferente.

4. Conclusiones

Nuestros resultados indican que, para la predicción de modelos ARMA(1,1) con un valor elevado (próximo a 1) del componente MA, los modelos en el espacio de estados estimados por el método de subespacios CCA (Canonical Correlation Análisis) proporcionan, en general, predicciones con menor error cuadrático medio que los modelos autorregresivos estimados por mínimos cuadrados.

Por otra parte, la identificación de modelos ARMA, aunque más compleja y costosa que la identificación de modelos AR o la de modelos SS mediante métodos de subespacios, es una técnica bien desarrollada, por lo que podría ser la mejor opción en el caso univariante. Sería interesante extender la comparación al caso multivariante, donde la identificación de modelos SS presenta algunas ventajas comparativas frente a la identificación de modelos VARMA, especialmente en lo que respecta a la especificación del modelo (en el caso SS se tiene un

único hiperparámetro a estimar). El modelo en el espacio de estados ofrece también mayor flexibilidad para su descomposición en componentes dinámicos e interpretación económica.

Agradecimientos

Los autores desean agradecer al profesor D. Juan del Hoyo Bernat sus comentarios e indicaciones, que han contribuido a enriquecer notablemente este documento.

Referencias

- Aoki, M. y Havenner, A. (1991). State Space Modeling of Multiple Time Series. *Econometric Reviews*, 10(1), 1-59.
- Bauer, D. y Ljung, L. (2002). Some Facts about the Choice of the Weighting Matrices in Larimore Type of Subspace Algorithms. *Automatica*, 38(5): 763-773
- Diebold, F.X. y Mariano, R.S. (2002). "Comparing Predictive Accuracy", *Journal of Business & Economic Statistics*, 20(1): 134-44.
- Larimore, W.E. (1983). "System Identification, Reduced Order Filters and Modelling via Canonical Variate Analysis". In: Rao, H.S., Dorato, P. (Eds.), *Proceedings of the 1983 American Control Conference*, Vol. 2, IEEE Service Center, Piscataway, NJ, 445-451.
- Ljung, L. (1999). *System Identification. Theory for the User*. 2nd edition. Prentice Hall.
- Siegel, S. y Castellan, N.J. (1988). *Non Parametric Statistics for the Behavioural Sciences*. 2nd edition. McGraw-Hill.