

Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

TAREA 4

EJERCICIOS

Esta tarea tiene por objeto familiarizarte con la forma general del modelo en espacio de estado (SSF, *state space form*) y mostrar como multitud de modelos, ya conocidos y de gran interés práctico, pueden expresarse en SSF. También completarás algunas cosas vistas en la tarea anterior.

1. Reescribe en forma de modelo en espacio de estado los siguientes modelos:

a) ARMA(3, 2):

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \theta_1 \nu_{t-1} + \theta_2 \nu_{t-2} + \nu_t.$$

b) VARMA(2, 1) en que \mathbf{y}_t es un vector bivariante y los coeficientes en la expresión que sigue son consecuentemente matrices 2×2 .

$$\mathbf{y}_t = \Phi_1 \mathbf{y}_{t-1} + \Phi_2 \mathbf{y}_{t-2} + \Theta_1 \nu_{t-1} + \nu_t.$$

c) ARIMA(1, 1, 1):

$$z_t = \phi_1 z_{t-1} + \theta_1 \nu_{t-1} + \nu_t$$

siendo $z_t = \nabla y_t$.

2. En una tarea anterior escribiste la función `kf.ll`. Modifícala de manera que haga suavizado además de filtrado y admita valores faltantes.

a) Haciendo uso, por ejemplo, de los datos de caudal del Nilo, fija algún intervalo de observaciones como perdidas (NA). Comprueba empíricamente (y demuestra analíticamente) que el filtro de Kalman cuando el modelo subyacente es de nivel local proporciona una trayectoria filtrada horizontal.

b) Comprueba empíricamente (y demuestra analíticamente) que el suavizador de Kalman cuando el modelo subyacente es de nivel local proporciona una interpolación lineal.

3. Considera el modelo con tendencia local lineal dado por:

$$y_t = \mu_t + \epsilon_t \quad (1)$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t \quad (2)$$

$$\beta_t = \beta_{t-1} + \psi_t \quad (3)$$

En este caso, la serie y_t se supone fluctúa en torno a un nivel μ_t . Este nivel a su vez evoluciona en el tiempo, del modo que prescribe la ecuación (??).

- Reescribelo como un modelo en espacio de estado.
 - ¿Qué condiciones sobre las varianzas de η_t y ψ_t convertirían a este modelo en uno con tendencia lineal ordinaria (no estocástica) de pendiente α (es decir, $y_t = \alpha t + \epsilon_t$)?
 - Escribe una función que haga filtrado y suavizado de Kalman.
 - “Por cada observación y_t el modelo anterior utiliza dos nuevos parámetros, μ_t y β_{t-1} . Un modelo así, en que el número de parámetros crece a doble velocidad que el número de observaciones, *nunca* puede estimarse, y es un artificio teórico de dudosa utilidad.” Comenta.
4. El fichero `paro.rda` contiene datos sobre paro estimado (EPA, del INE), paro registrado (INEM) y un realineamiento del paro EPA (renombrado EPAC) válido sólo a los efectos de este ejercicio¹. Puedes ver sus trayectorias respectivas en la gráfica siguiente.

Observa que la serie EPAC sólo es observada cada tres meses. Tu trabajo consiste en emplear un modelo estructural simple para proporcionar una serie estimada de EPAC con datos mensuales, explotando la información en la serie del INEM².

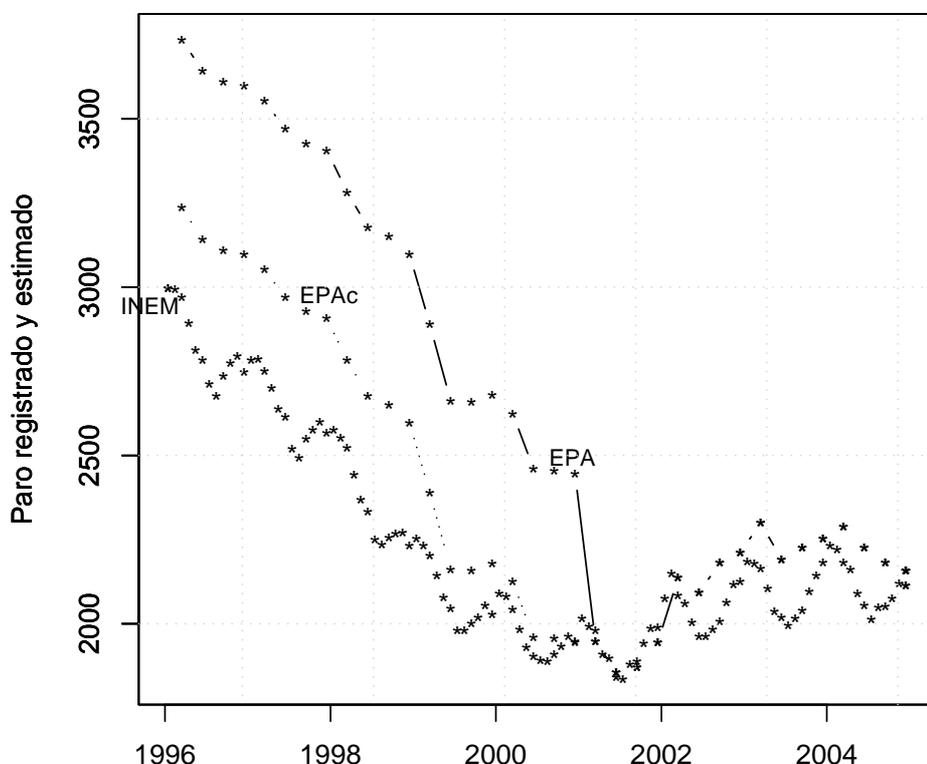
AYUDAS, SUGERENCIAS Y COMPLEMENTOS

- Puedes ver una de las varias formas de reexpresar modelos $ARMA(p, q)$ y $ARIMA(p, d, q)$ generales en [?], sección 3.3.
- Un tratamiento pormenorizado de modelos estructurales (nivel aleatorio, tendencia local, tendencia + ciclo, etc.) puede encontrarse en [?], sección 2.3.6.
- La frase entrecomillada en ?? más arriba y que has de comentar quiere hacerte reflexionar sobre este tipo de modelos en términos de conceptos que has estudiado en Econometría (como el de identificación). *En principio* son necesarias *más* observaciones que parámetros para poder estimar un modelo, pero ¿es siempre esto así? ¿Quizá la imposición de alguna estructura sobre los parámetros del modelo puede hacer que, después de todo, éste sea estimable?

¹A partir de 2001 se produjo un cambio en el modo en que los parados se computaban. Ello originó una ruptura en la serie EPA, que hemos corregido aquí de una manera chapucera, al objeto de no distraer del objetivo central del ejercicio. Podrías dar entrada a dicho cambio –y estimar su efecto– mediante un modelo en espacio de estado un poco más complejo, tal y como se hace en el análisis de intervención a la *Box-Jenkins*.

²Este ejercicio es algo abierto en la respuesta que quieras darle, que depende del tiempo y energía que puedas consagrarle. Tiene un poco más de reto que los precedentes, de directa aplicación de la teoría. Al mismo tiempo te mostrará, dentro de su simplicidad, la potencia y versatilidad de las herramientas que manejas.

Evolución del paro 1996--2005 en miles de personas



4. Observa: si en el modelo de tendencia lineal local la varianza de ψ_t fuera cero, $\beta_t = \beta_{t-1} = \beta$ para cualquier t , y $\mu_t = \mu_{t-1} + \beta + \eta_t$; μ_t fluctuaría en torno a una tendencia lineal βt de pendiente β . En el modelo de tendencia lineal local, β —y por tanto la pendiente de μ_t — se permite que varíe; de ahí el nombre de tendencia lineal *local*.
5. Sobre el empleo de series con diferente frecuencia de observación, estimación de las menos frecuentemente observadas (o “caras” de observar) con ayuda de las más “baratas”, etc. puedes leer [?] y [?] (este último es un ejercicio similar al que se te propone en ??).
6. Los datos pueden leerse mediante un `load("paro.rda")`. Ello pondrá en tu espacio de trabajo dos objetos, `paro` y `paro.its`, que es el que te recomiendo utilizar: contiene series temporales en un formato especial, fechado, propio del paquete `its`, que debes obtener y cargar previamente (con un `library(its)`). Puedes convertir las columnas que te interesen a vectores o matrices sin problemas (con `as.vector()` o `as.matrix()`)³.
7. Observa: tendría sentido definir un “paro subyacente” α_t , del que tanto EPAC como INEM son observaciones. Este “paro subyacente” evolucionaría mes a mes, con una dinámica que podrías

³El motivo por el que hago uso de este formato es que facilita mucho la elaboración de gráficos como el que figura en esta misma tarea.

suponer de paseo aleatorio. Las series que observas, INEM y EPAC, dependen de dicho paro subyacente: INEM podrías imaginarla como $\delta\alpha_t$ con un $\delta < 1$ a estimar. EPAC probablemente tendría sentido imaginarla dependiendo de tres valores de α_t consecutivos —dado que cuando se observa EPAC, proporciona estimaciones “promedio” correspondientes a un trimestre—.

8. Observa: la serie INEM posee una periodicidad evidente —que la referida “promediación” trimestral realizada por la EPA (y por tanto EPAC)— hace mucho menos aparente. Podrías imaginar que el “paro subyacente” experimenta una evolución similar, y dotarle de una dinámica más compleja que la de mero paseo aleatorio.
9. Puedes estimar modelos en espacio de estado con casi completa generalidad utilizando las funciones en el paquete `sspir` de R. Recuerda que entre las cosas que devuelve la función `kfilter` que examinamos en clase, está la verosimilitud. Ello abre la puerta a que estimes lo que necesites con ayuda de la función `optim` a la que pases como argumento una función que calcule dicha verosimilitud (invocando a su vez a `kfilter`). Verás ejemplos en clase.
10. Los datos de la EPA proceden de INEbase (búscalo en la página de entrada a la Web del INE, <http://www.ine.es>). Los datos de paro registrado provienen del Ministerio de Trabajo y Asuntos Sociales, cuya Web puedes encontrar en <http://www-org.mtas.es>.